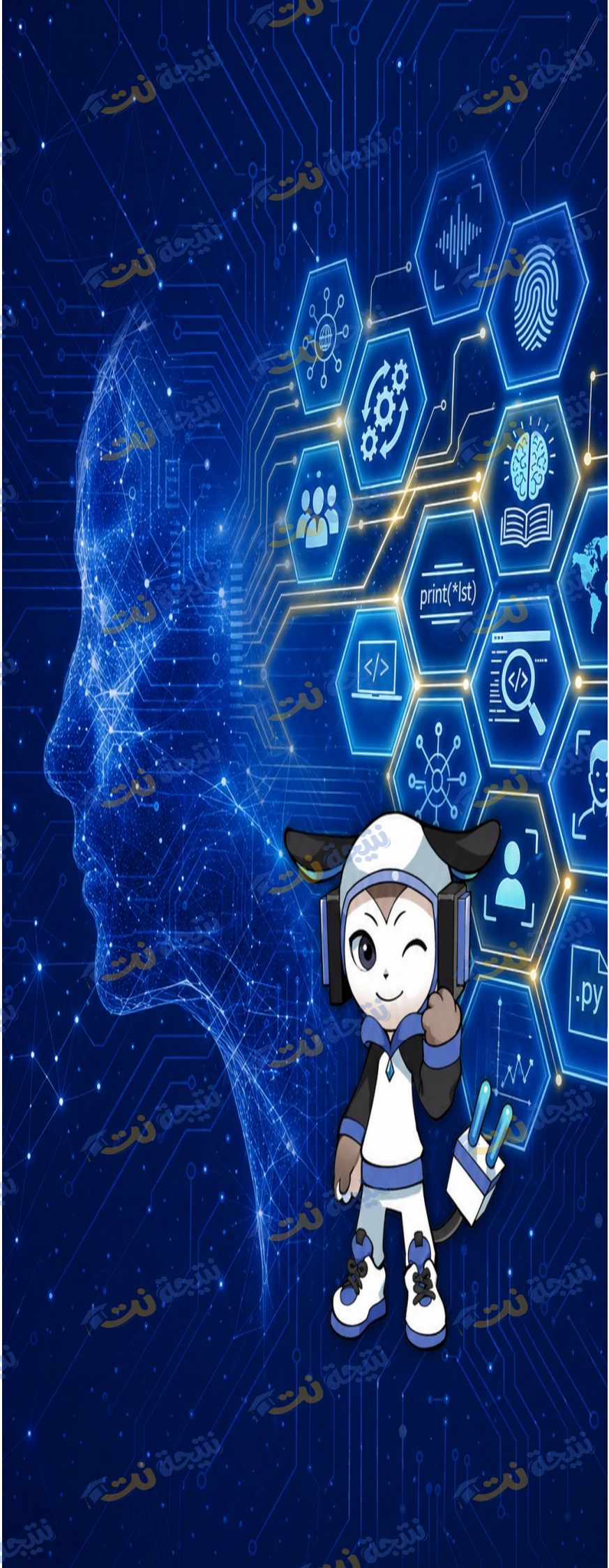


البرمجة والذكاء الاصطناعي

الصف الثاني بكالوريا

الجزء الثاني

11

nezakr.natega4dk.net

Programming and Artificial Intelligence

Egyptian Baccalaureate

2nd Year

Advised by INTERNATIONAL BACCALAUREATE®

The International Baccalaureate™

has reviewed the pedagogical and assessment approaches underpinning this textbook.

© 2026 Ministry of Education and Technical Education

All rights reserved.

Copyrights and accountability for the content of this textbook remain with the Ministry of Education and Technical Education.

No part of this publication may be reproduced, stored in a retrieval system,

or transmitted in any form or by any means - electronic, mechanical, photocopying, recording,

or otherwise - without the prior written permission of the copyright holders.

Ministry of Education and Technical Education

New Administrative Capital

Cairo, Egypt

nezakr.natega4dk.net

مقدمة

في ضوء **رؤية مصر 2030**، واستمرار جهود الدولة لتطوير التعليم، تولي وزارة التربية والتعليم والتعليم الفني اهتمامًا كبيرًا بتحديث المناهج الدراسية، ولا سيما مجالات البرمجة والذكاء الاصطناعي وتكنولوجيا المعلومات والاتصالات؛ لما لها من دور أساسي في تنمية التفكير النقدي، والقدرة على حل المشكلات، وفهم العالم الرقمي المتغير.

نُظّم محتوى هذا الكتاب بصورة متدرجة ومنهجية تساعد الطالب على الانتقال بسلاسة من المفاهيم الأساسية إلى التطبيقات الأكثر تقدمًا. ويربط الكتاب المعرفة التقنية بمواقف من الحياة اليومية، ويعتمد أنشطة عملية، ومناقشات، ومهامًا تعاونية تجعل التعلم أكثر فاعلية ومتعة.

وقد روعي في إعداد المحتوى الإفادة من الخبرات والممارسات التعليمية الدولية الناجحة، مع الحفاظ على الهوية المصرية، وتنمية الاستخدام المسؤول والآمن للتكنولوجيا. لذلك يجمع الكتاب بين الفهم العلمي، والتطبيق العملي، والتحقق من المعلومات، واحترام الخصوصية والأمان الرقمي.

يمثل هذا الكتاب خطوة ضمن خطة متكاملة لتطوير مناهج تكنولوجيا المعلومات والاتصالات عبر المراحل الدراسية، بما يهيئ الطلاب للتعامل الواعي مع البرمجة والبيانات والذكاء الاصطناعي، ويمنحهم أساسًا يمكن البناء عليه في الدراسة والعمل.

وتؤمن وزارة التربية والتعليم والتعليم الفني بأن تطوير التعليم استثمار حقيقي في مستقبل الوطن، وأن تحديث المناهج وسيلة لبناء جيل واعٍ ومبدع، قادر على مواكبة التغيرات العلمية والتكنولوجية والمشاركة الإيجابية في المجتمع الرقمي.

nezakr.natega4dk.net

المحتويات

5 جمع البيانات وتنقيتها 4

5-1 طرق جمع البيانات 4

5-2 تنقية البيانات وتحويلها 11

5-3 البيانات المفتوحة وواجهات برمجة التطبيقات 19

6 التحليل والتواصل 26

6-1 الاستدلال الإحصائي 26

6-2 استخدام تحليل الانحدار وتقييمه 34

6-3 تمثيل البيانات المرئي والتواصل 42

7 التعلم الآلي والذكاء الاصطناعي 50

7-1 أساسيات التعلم الآلي 50

7-2 الشبكات العصبية والتعلم العميق 59

7-3 نماذج اللغة الكبيرة والذكاء الاصطناعي التوليدي 67

nezakr.natega4dk.net



الدرس 5-1

طرق جمع البيانات

أهداف التعلم

- التمييز بين البيانات الأولية والثانوية، وتوضيح الحالات التي تكون فيها كلٌّ منها مناسبة.
- شرح آلية أخذ العينات، ووصف كيف يمكن أن يؤدي التحيز في العينة إلى استنتاجات خاطئة رغم أنها تبدو واثقة.

nezakr.natega4dk.net

يبدأ التحليل الجيد بجمع بيانات مناسبة للسؤال. فالبيانات الأولية تُجمع مباشرة لغرض الدراسة الحالية، أما البيانات الثانوية فهي بيانات موجودة مسبقًا يعاد استخدامها لغرض جديد. وعندما يكون فحص المجتمع الإحصائي كله صعبًا أو مكلفًا، يمكن دراسة عينة منه. وكلما كانت طريقة الاختيار أفضل وأقل تحيزًا، زادت الثقة في تعميم النتائج، مع مراعاة حجم العينة وتصميم الدراسة وعدم اليقين.

؟ السؤال الرئيس: كيف تجمع بيانات يمكنك الوثوق بها، وكيف يمكن لعينة صغيرة أن تمثل مجتمعًا إحصائيًا كاملًا بعدالة؟

📌 ملحوظة مهمّة

1. مرحلة جمع البيانات

(1) يتقدم تحليل البيانات عبر تسلسل "الجمع ← التنظيم ← التحليل"، بعد توضيح الغرض. وفي **(جمع البيانات)**، المرحلة الأولى، من الضروري التخطيط لما سيُجمع من بيانات وكيف، بما يتماشى مع غرض التحليل.

(2) إذا كانت البيانات غير مناسبة للسؤال أو منخفضة الجودة، فقد تكون النتائج مضللة مهما كان التنظيم والتحليل دقيقين. لذلك تعد طريقة الجمع مرحلة مهمة تؤثر بقوة في جودة التحليل.

2. البيانات الأولية والبيانات الثانوية



البيانات الأولية تُجمع مباشرة للسؤال الحالي؛ والثانوية بيانات موجودة مسبقًا يعاد استخدامها.

(1) **البيانات الأولية**: بيانات يجمعها الباحث أو الفريق مباشرة لغرض الدراسة أو التحليل الحالي، مثل بيانات الاستبيانات والمقابلات والملاحظات والتجارب. ويمكن تصميم طريقة الجمع لتناسب السؤال، لكن الجودة تعتمد على التصميم والتنفيذ، وقد يتطلب الجمع وقتًا وجهدًا وتكلفة.

(2) **البيانات الثانوية**: بيانات موجودة مسبقًا، جُمعت لغرض سابق ثم يعاد استخدامها، سواء جمعها آخرون أو الباحث نفسه سابقًا. قد توفر وقتًا وتكلفة، لكن يجب فحص طريقة جمعها وتعريفاتها ونقصها وتحيزاتها ومدى مناسبتها للسؤال الحالي.

(3) قد تكون البيانات الأولية مناسبة جدًا لسؤال محدد، لكن موثوقيتها لا تكون عالية تلقائيًا؛ فهي تعتمد على جودة التصميم والتنفيذ. وقد توفر البيانات الثانوية معلومات جاهزة لا تحتاج إلى جمعها مرة أخرى، لكنها قد تناسب السؤال أو لا تناسبه. ويمكن الجمع بين النوعين بحسب الغرض والقيود.

⭐ خريطة الدرس

الفكرة الأساسية:

يبدأ التحليل الجيد بالجمع الجيد: البيانات الأولية تُجمع حديثًا لغرض محدد؛ والبيانات الثانوية موجودة مسبقًا؛ وأخذ العينات يدرس جزءًا من المجتمع الإحصائي لتقدير الكل – ولا يدعم استنتاجات صحيحة سوى العينة غير المتحيزة.

المفاهيم الأساسية:

- جمع البيانات
- البيانات الأولية / الثانوية
- المجتمع الإحصائي والعينة
- أخذ العينات (العشوائي البسيط، الطبقي)
- تحيز أخذ العينات



الجمع السليم للبيانات هو أساس البيانات الموثوقة.

مصطلحات أساسية

جمع البيانات – المرحلة الأولى: التخطيط لما سيُجمع وكيف.

البيانات الأولية – تُجمع حديثًا لغرض محدد.

البيانات الثانوية – سبق أن جمعها آخرون.

nezakr.natega4dk.net

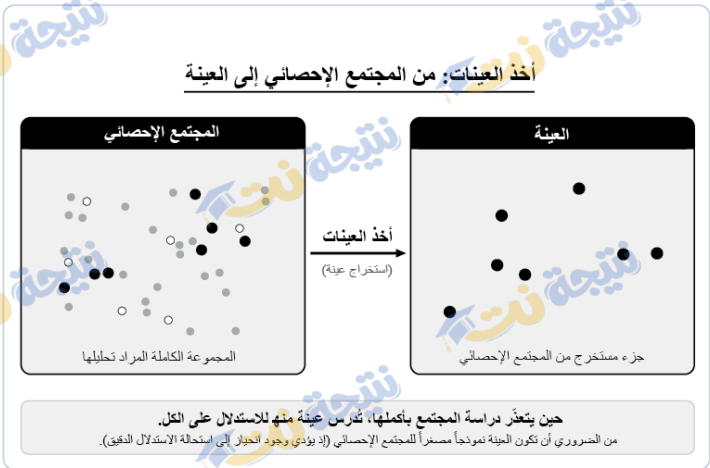
3. أخذ العينات



نحاول اختيار عينة تمثل المجتمع قدر الإمكان، ونتجنب طرق الاختيار التي تزيد التحيز.

(1) **المجتمع الإحصائي**: جميع الأفراد أو العناصر التي نريد أن تنطبق عليها الدراسة. و**العينة**: جزء منه ندرسه ونريد أن يمثله قدر الإمكان. وتسمى دراسة جميع الوحدات حصرًا شاملًا أو تعدادًا في بعض السياقات، وتسمى دراسة عينة مختارة مسح عينة.

(2) **(أخذ العينات)**... اختيار عينة من المجتمع الإحصائي. تختلف موثوقية الاستنتاجات المُتحصَّل عليها اختلافًا كبيرًا حسب طريقة أخذ العينات.



الشكل 5.1.1 – أخذ العينات: من المجتمع الإحصائي إلى العينة

(3) قد يقلل الاختيار العشوائي بعض أنواع تحيز الاختيار إذا طُبق بطريقة سليمة، وقد يدعم تعميم النتائج مع مراعاة حجم العينة وإطار المعاينة وعدم الاستجابة وأخطاء القياس.

الطريقة	الوصف
(المعاينة العشوائية البسيطة)	اختيار وحدات عشوائيًا من إطار المجتمع، بحيث تكون لكل وحدة فرصة معروفة ومتساوية في النموذج البسيط
(المعاينة الطبقية)	تقسيم المجتمع إلى فئات مهمة لسؤال الدراسة، ثم اختيار عينة من كل فئة وفق تصميم محدد

(4) المعاينة العشوائية البسيطة طريقة أساسية وسهلة الفهم، لكنها تحتاج إلى إطار مناسب يمكن الاختيار منه عشوائيًا. وقد تساعد المعاينة الطبقية على تمثيل فئات مهمة وتحسين الدقة إذا اختيرت الطبقات والنسب بطريقة مناسبة.

4. تحيز أخذ العينات

(1) **تحيز المعاينة**: انحراف يحدث عندما تجعل طريقة اختيار العينة بعض أفراد المجتمع أكثر أو أقل احتمالًا للتمثيل، فتكون النتائج أقل تمثيلًا للمجتمع وقد يصبح تعميمها غير موثوق.

(2) أمثلة على مصادر مختلفة للانحراف في الاستطلاع

توقّف وفكّر

التعداد السكاني الحكومي الذي تُنزّله – هل هو بيانات أولية أم ثانوية بالنسبة لك؟

nezakr.natega4dk.net

نوع التحيز	الوصف	مثال
(تحيز الاختيار الذاتي)	تحيز ناتج عن استجابة من يرغبون في الاستجابة فقط	في استطلاع عبر الويب، يستجيب فقط من لديهم اهتمام عالٍ
(تحيز الاستجابة)	تحيز ناتج عن اختيار المستجيبين إجابات تبدو مرغوبة بدلًا من استجابة دقيقة	الإبلاغ عن "وقت دراسة" أطول مما كان عليه فعليًا
(تحيز الوقت/ الموقع)	تحيز في التغطية أو الاختيار بسبب قصر وقت الاستطلاع أو مكانه على فئات معينة	إجراء الاستطلاع أمام محطة في بعد ظهيرة أيام الأسبوع، مما يستبعد الموظفين والطلاب

(3) يمكن تقليل تحيز المعاينة بتحسين إطار المعاينة وطريقة اختيار المشاركين وتوقيت الوصول إليهم. أما صياغة الأسئلة وطريقة الإجابة فتساعدان على تقليل أخطاء القياس وتحيز الاستجابة؛ وهما مشكلتان مختلفتان عن تحيز المعاينة.

مصطلحات أساسية

المجتمع الإحصائي – المجموعة الكاملة التي يشملها التحليل.

العينة – الجزء الذي يُمسح فعليًا.

تحيز أخذ العينات – عينة غير عادلة لا تمثل الكل تمثيلًا صحيحًا.

الفكرة الرئيسة

يعتمد التحليل الموثوق على الجمع، لا على الحساب وحده. اختر البيانات الأولية أو الثانوية بما يناسب الغرض، وادرس عينة عندما يكون المجتمع الإحصائي كبيرًا جدًا، واستخدم أخذ العينات العشوائي (البسيط أو الطبقي) بحيث تعكس العينة الكل، وانتبه لتحيز أخذ العينات – فالعينة غير المتوازنة تعطي استنتاجًا واثقًا لكنه خاطئ.

توقّف وفكّر

استطلاع عبر الويب يمكن لأي شخص المشاركة فيه – أصوات من تغيب، وأي نوع من التحيز هذا؟

مثال محلول

(1) ضع علامة ○ أو × أمام كل عبارة (أ - د) حسب صحتها.

أ تشير البيانات الأولية إلى بيانات سبق جمعها ونشرها من قِبل آخرين.

ب يُسمى اختيار عينة من مجتمع إحصائي أخذ العينات.

ج المعاينة الطبقية طريقة يُقسَّم فيها المجتمع الإحصائي إلى طبقات، وتُختار عينة عشوائيًا من كل طبقة.

د عند وجود تحيز في أخذ العينات، قد لا تكون الاستنتاجات المُتحصَّل عليها قابلة للتطبيق على المجتمع الإحصائي ككل.

(2) طابق كل وصف (1، 2) لطريقة أخذ عينات بالخيار الأنسب من الخيارات أدناه (أ، ب).

1 اختيار عينة من المجتمع الإحصائي بالصدفة، دون تحيز.

2 تقسيم المجتمع الإحصائي إلى طبقات مثل الجنس أو الفئة العمرية، والاختيار عشوائيًا من كل طبقة.

[الخيارات] أ المعاينة الطبقية ب المعاينة العشوائية البسيطة

الحلّ

(1)

أ البيانات الأولية هي بيانات تُجمع حديثًا لتناسب غرضًا محددًا. أما البيانات التي سبق جمعها ونشرها من قِبل آخرين فهي بيانات ثانوية. لذلك، ×.

ب تعريف صحيح لأخذ العينات. لذلك، ○.

ج تعريف صحيح لالمعاينة الطبقية. لذلك، ○.

د وصف صحيح لأثر تحيز أخذ العينات. لذلك، ○.

(2)

1: الاختيار دون تحيز بالصدفة هو المعاينة العشوائية البسيطة. **ب**

2: التقسيم إلى طبقات والاختيار عشوائيًا من كل منها هو المعاينة الطبقية. **أ**

تدرّب

1 أجب عن الأسئلة التالية.

1. ما اسم البيانات التي تُجمع مباشرة لغرض الدراسة الحالية؟
2. ما اسم البيانات الموجودة مسبقًا التي نعيد استخدامها لغرض جديد؟
3. ما مصطلح اختيار عينة من مجتمع إحصائي؟
4. ما مصطلح كون طريقة اختيار العينة متحيزة ولا تعكس بشكل صحيح الحالة الفعلية للمجتمع الإحصائي؟

2 أجب عن الأسئلة التالية.

(1) من الخيارات التالية (أ - د)، اختر الأنسب كطريقة لجمع بيانات أولية.

أ استخدام بيانات إحصائية نشرتها جهة حكومية.

ب إنشاء استبيان يناسب غرض المرء الخاص وجمع الردود.

ج الاستشهاد بأرقام من تقرير صادر عن جمعية صناعية.

د استخدام بيانات استطلاع من دراسة أجراها شخص آخر سابقًا كما هي.

(2) لكل موقف استطلاع (1، 2)، اختر طريقة أخذ العينات الأنسب من الخيارات أدناه (أ، ب).

1 عند دراسة وقت الدراسة لدى جميع طلاب مدرسة، تريد اختيار العينة مع مراعاة الفروق في الاتجاهات بين الصفوف.

2 تريد استيعاب آراء جميع الناخبين المؤهلين في مدينة دون تحيز.

[الخيارات] أ المعاينة العشوائية البسيطة ب المعاينة الطبقية

(3) من الخيارات التالية (أ - د)، اختر الأنسب كمثال حدث فيه تحيز في أخذ العينات.

أ مسح 100 طالب اختيروا عشوائيًا من جميع طلاب المدرسة بالقرعة.

ب تقسيم جميع السكان إلى طبقات عمرية ومسح عينة عشوائية من كل طبقة.

ج مقابلة المارة أمام المحطة في بعد ظهيرة أيام الأسبوع والإبلاغ عن النتائج باعتبارها "آراء جميع المواطنين".

د تصميم صياغة سؤال الاستطلاع بعناية بحيث يستطيع المستجيبون الإجابة بصدق.

📝 سؤال على نمط الامتحان

اشرح لماذا يمكن لعينة متحيزة أن تجعل التحليل عديم القيمة حتى عندما تكون الحسابات صحيحة. **[6]**

راجع: العينة بوصفها نموذجًا مصغرًا للمجتمع الإحصائي.

💡 توقّف وفكّر

أي خيار يجمع بيانات جديدة صمّمتها *أنت* – وأيها يعيد استخدام بيانات الآخرين فقط؟

nezakr.natega4dk.net

⚙️ فكّر كمهندس: ابحث ثم قرر

تريد مدرستك معرفة المدة التي يقضيها الطلاب عادةً في التنقل إلى المدرسة يوميًا، لكن لا يمكنك مسح الجميع. أنت تصمم كيفية اختيار عينة وسؤالها.

1 · خطط للعينة. حدّد المجتمع الإحصائي وطريقة اختيار عينة تقلل التحيز وتمثل الفئات المهمة. ويمكنك جمع بيانات تجريبية من 8 إلى 10 زملاء لتجربة السؤال، مع التنبيه إلى أن هذه المجموعة ليست بالضرورة عينة ممثلة للمدرسة.

2 · حلّل مصدر الانحراف. حدّد مجموعة قد تستبعدها طريقتك، واشرح أثر غيابها. ثم حدّد هل المشكلة في اختيار العينة أو تغطيتها، أم في طريقة إجابة المشاركين، ولا تجمع الأنواع المختلفة تحت اسم واحد.

3 · قرّر. اختر المعاينة العشوائية البسيطة أو الطبقية لهذا الاستطلاع وبرّر اختيارك من واقع التحدي الذي وجدته في الخطوة 2 – فكّر فيما إذا كان التنقل يختلف فعلًا بين فئات فرعية يمكن تمييزها هنا. **في أزواج،** قارنوا الطرق واتفقوا على التغيير الوحيد الذي سيقلل التحيز أكثر من غيره.

☐ **سؤال من يسهل الوصول إليه** ☐ **التخطيط لعينة عادلة وممثلة**

قدّم دليلًا على إجابتك.

هل تواجه صعوبة؟ ابدأ من من قد يُستبعد. اسأل نفسك: هل قد يجيب جزء من المدرسة بشكل مختلف تمامًا عن البقية – مجموعة كاملة، أو طلاب لا يرون استطلاعك أبدًا؟ إذا كان من المحتمل أن تختلف فئة فرعية بشكل منهجي، فالسؤال الحقيقي هو كيف تمنح كل فئة فرعية فرصة عادلة *قبل* اختيار طريقتك. حدد الفئات الفرعية المهمة هنا، ثم دع ذلك يوجّه اختيارك.

مصطلحات أساسية

المعاينة العشوائية البسيطة – يُختار بمحض الصدفة.

المعاينة الطبقية – يُقسَّم إلى طبقات، ثم يُختار من كل منها.

🌍 طبّق ما تعلمته

مركز تسوق يستطلع فقط زوار بعد ظهر يوم السبت ويُبلغ عن النتائج باعتبارها "ما يفكر فيه جميع عملائنا." اذكر نوع التحيز، واشرح من يُستبعد، وصف تغييرًا واحدًا من شأنه أن يجعل العينة أكثر عدالة.

❓ إجابة السؤال الرئيس

أجاب الدرس عن السؤال الرئيس: كيف نجمع بيانات مناسبة ونختار عينة؟ تجمع البيانات الأولية مباشرة لغرض الدراسة الحالية، بينما يعاد استخدام البيانات الثانوية الموجودة مسبقًا. ولا يضمن أي نوع الجودة تلقائيًا؛ بل يجب فحص التصميم والتنفيذ ومدى ملاءمة البيانات للسؤال. وعندما يصعب دراسة المجتمع كله نختار عينة. قد يقلل الاختيار العشوائي تحيز الاختيار، وتساعد المعاينة الطبقية على تمثيل فئات مهمة، لكن التعميم يعتمد أيضًا على إطار المعاينة والحجم وعدم الاستجابة والقياس. ويجب التمييز بين تحيز المعاينة الناتج عن الاختيار وتحيز الاستجابة الناتج عن طريقة الإجابة.

nezakr.natega4dk.net

تمارين

1 اختبر فهمك: أجب عن الأسئلة الآتية دون الرجوع إلى قسم «ملحوظة مهمة»، ثم راجع إجاباتك.

2 اقرأ الفقرة التالية وأجب عن كل سؤال.

في تحليل البيانات، من المهم التخطيط لما سيُجمع وفق السؤال. يمكن استخدام (1) التي تُجمع مباشرة لغرض الدراسة الحالية، أو (2) الموجودة مسبقًا التي يعاد استخدامها. وعندما يصعب دراسة المجتمع كله، تُستخدم (3) لاختيار جزء منه. وإذا جعلت طريقة الاختيار بعض فئات المجتمع أكثر أو أقل احتمالًا للتمثيل، يحدث (4) وقد تصبح النتائج أقل قابلية للتعميم.

1. املأ الفراغات (1) - (4).

3 أجب عن الأسئلة التالية.

(1) من الخيارات التالية (أ - د)، اختر الأنسب كمثال على بيانات ثانوية.

أ ردود استبيان جُمعت من طلاب في صف المرء الخاص.

ب إحصاءات سكانية نشرتها الحكومة.

ج حركة مرور على طريق تنقل مدرسي لاحظها المرء وسجّلها بنفسه.

د سجلات مقابلة صمّمها المرء وأجراها بنفسه.

(2) من الخيارات التالية (أ - د)، اختر الخيار الذي يصف طرق أخذ العينات بأفضل شكل.

أ المعاينة العشوائية البسيطة طريقة يختار فيها القائم بالاستطلاع تفضيليًا المشاركين الأسهل جمعًا.

ب المعاينة الطبقية طريقة يُقسَّم فيها المجتمع الإحصائي إلى طبقات، وتُختار عينة عشوائيًا من كل طبقة.

ج المعاينة العشوائية البسيطة طريقة يُقسَّم فيها المجتمع الإحصائي إلى طبقات، ويُختار ممثل من كل طبقة.

د المسح الشامل طريقة يُمسح فيها جزء فقط من المجتمع الإحصائي.

(3) من الخيارات التالية (أ - د)، اختر ما لا يُعد طريقة مناسبة لتقليل تحيز أخذ العينات.

أ اختيار المشاركين دون تحيز من المجتمع الإحصائي بأكمله.

ب تقييد توقيت وموقع الاستطلاع، وجمع الردود فقط من أشخاص في ظروف محددة.

ج اختيار العينة مع مراعاة طبقات المجتمع الإحصائي (الفئة العمرية، الجنس، إلخ).

د تصميم صياغة أسئلة الاستطلاع بحيث يستطيع المستجيبون الإجابة بصدق.

⭐ الخلاصة

تذكّر:
اجمع البيانات الصحيحة (الأولية مقابل الثانوية)، وادرس عينة عادلة من المجتمع الإحصائي، واستخدم أخذ العينات العشوائي (البسيط أو الطبقي) بحيث تعكس العينة الكل، واحترس من تحيز أخذ العينات.

⭐ فكّر وتحدَّ نفسك

تأمل: فكّر في استطلاع رأي عبر الإنترنت شاهدته. من كانت أصواتهم غائبة على الأرجح عنه، ولماذا؟

تحدٍّ: يمكنك استطلاع رأي 40 طالبًا فقط من أصل 800 طالب في مدرستك حول جدول جديد. صف بالضبط كيف ستختار هؤلاء الـ40 لتجنب التحيز. واذكر المجموعة الواحدة التي ستحرص على تمثيلها.

nezakr.natega4dk.net



الدرس 5-2

تنظيف البيانات وتحويلها

أهداف التعلم

- شرح كيفية التعامل مع القيم المفقودة، والقيم الشاذة (المتطرفة)، والبيانات المكررة عند تنظيف مجموعة البيانات.
- وصف عمليات تحويل أنواع البيانات (الربط والتجميع).

nezakr.natega4dk.net

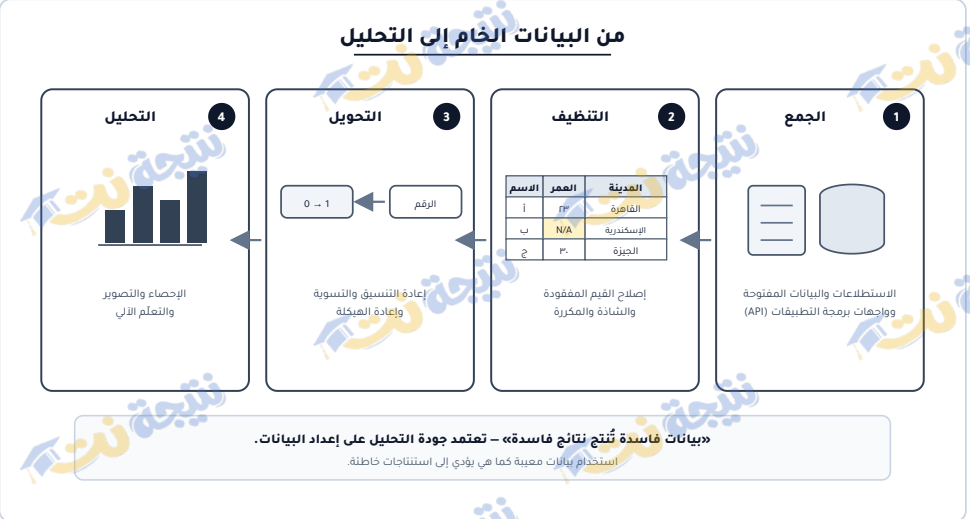
تصل البيانات الحقيقية فوضوية: خلايا فارغة، قيمة أكبر بعشر مرات من اللازم، نفس الصف مُدخَل مرتين، تواريخ مكتوبة بثلاث طرق مختلفة. قبل أن يُوثَق بأي تحليل، يجب تنظيف البيانات وتحويلها لتناسب السؤال المطروح. هناك قاعدة في علوم الحاسب – "مدخلات خاطئة – مخرجات خاطئة" – تعني أن الحساب الدقيق على بيانات معيبة لا يزال يعطي إجابة معيبة. هذا الدرس يتناول كيفية تحويل البيانات الخام إلى بيانات يمكن الاعتماد عليها فعليًا.

؟ السؤال الرئيس: لماذا يجب تنظيف البيانات وتحويلها قبل التحليل، وما الخيارات التي ينطوي عليها كل إصلاح؟

ملحوظة مهمّة

1. مراحل تنظيف البيانات وتحويلها

(1) قد تحتاج البيانات المجمعة إلى تنظيف أو تحويل قبل التحليل. **تنظيف البيانات** هو اكتشاف مشكلاتها وتصحيحها أو التعامل معها بطريقة مناسبة، أما **تحويل البيانات** فيعد الصيغ أو البنية لتناسب السؤال وطريقة التحليل.



الشكل 5.2.1 – من البيانات الخام إلى التحليل: الجمع، التنظيف، التحويل، التحليل

(2) تؤثر جودة التحضير بقوة في جودة النتائج، لكن لا توجد سلسلة إصلاح واحدة تناسب كل البيانات؛ فالقرارات تعتمد على نوع المشكلة وسؤال التحليل.

2. التعامل مع القيم المفقودة



عالج القيم الفارغة وقيّم القيم الشاذة قبل التحليل.

خريطة الدرس

الفكرة الأساسية:
نادرًا ما تكون البيانات الخام جاهزة للاستخدام مباشرة. يعالج التنظيف القيم المفقودة والقيم الشاذة والتكرارات؛ ويحوّل التحويل الأنواع، ويعيد ضبط المقياس، ويعيد تشكيل البيانات لتناسب الغرض. "مدخلات خاطئة – مخرجات خاطئة" – تعتمد جودة أي تحليل على هذا التحضير.

المفاهيم الأساسية:
- تنظيف البيانات / تحويل البيانات
- القيم المفقودة، القيم الشاذة، التكرارات
- تحويل نوع البيانات
- التحويل الكمي / التوحيد القياسي
- إعادة التشكيل (المحور، الربط، التجميع)

مصطلحات أساسية
تنظيف البيانات – إصلاح القيم المفقودة والقيم الشاذة والتكرارات.
تحويل البيانات – إعادة تشكيل/ضبط مقياس البيانات لتناسب الغرض.

توقّف وفكّر
خلية فارغة في استبيان – هل نحذف الصف، أم نملؤها، أم نضع عليها علامة؟ ما الذي يحدد ذلك؟

nezakr.natega4dk.net

(1) قد تسبب القيم المفقودة مشكلات في بعض الحسابات أو تغير النتائج بحسب مقدارها وسببها ونمطها. ومن الطرق الشائعة الحذف أو التعويض أو الاحتفاظ بمؤشر يدل على أن القيمة مفقودة، وقد توجد طرق أخرى بحسب التحليل.

الاستجابة	الوصف	متى تُطبَّق
الحذف	إزالة سجل أو متغير إذا كان ذلك مناسبًا ولا يفقد معلومات مهمة أو يسبب تحيزًا مهمًا	قد يناسب عندما يكون الفقد قليلًا وفقد السجلات لا يسبب تحيزًا أو خسارة مهمة
التعويض	تقدير قيمة بديلة بطريقة مناسبة، وقد تكون بسيطة مثل الوسيط أو أكثر تقدمًا	يُختار بعد فهم سبب الفقد ونمطه وطريقة التحليل، مع فحص أثر التعويض في التوزيع والنتائج
وضع علامة	إنشاء مؤشر يوضح أن القيمة كانت مفقودة، أو استخدام فئة مناسبة إذا كان الغياب نفسه يحمل معنى	عندما يكون للغياب نفسه معنى (مثل حقيقة أن شخصًا ما رفض الإجابة)

(2) يعتمد الاختيار على مقدار الفقد وسببه ونمطه وغرض التحليل. وقد يغير التعويض البسيط بالمتوسط توزيع البيانات ويقلل التباين، خاصةً إذا كان الفقد كبيرًا؛ لذلك يجب فحص أثر الطريقة في النتائج.

3. التعامل مع القيم الشاذة والبيانات المكررة

(1) **القيم الشاذة** بعيدة عن معظم القيم، لكنها ليست أخطاء بالضرورة؛ فقد تنتج عن خطأ إدخال أو قياس، وقد تكون قيمة حقيقية نادرة. لذلك يجب التحقق من مصدر القيمة وتعريف المتغير قبل تعديلها.

(2) تبدأ معالجة القيم الشاذة بتحديد السبب.

① عندما يكون السبب خطأ في الإدخال أو القياس... صحّح القيمة أو احذفها

② إذا كانت القيمة صحيحة، فلا تحذفها لمجرد أنها بعيدة عن الباقي؛ قرر وفق سؤال التحليل وبسبب تحليلي واضح.

(3) قد تغير السجلات المكررة الحسابات، لكن لا يُحذف سجل مكرر ظاهريًا إلا بعد التأكد أنه تكرار غير مقصود لنفس الملاحظة أو الحدث؛ فقد تمثل السجلات المتشابهة أحداثًا حقيقية منفصلة.

4. تحويل نوع البيانات

(1) يجب الفصل بين **تغيير نوع البيانات**، مثل تحويل نص رقمي إلى رقم، وبين **توحيد الصيغ أو الترميز**، مثل توحيد كتابة التاريخ أو القيم الفئوية. وقد تحتاج البيانات إلى أحد هذه التحويلات أو أكثر.

نوع التحويل	الوصف	مثال
التحويل من نص إلى رقم	تحويل الأرقام المخزَّنة كنصوص إلى قيم رقمية يمكن استخدامها في الحسابات	"١٢٣" ← 123
توحيد صيغة التاريخ	مواءمة التواريخ ذات الصيغ المختلفة في صيغة واحدة	"2026/4/1"، "2026-04-01"، "01 Apr 2026" ← 2026-04-01
توحيد طريقة الكتابة	دمج القيم ذات المعنى الواحد في صيغة كتابة واحدة	مزيج من "ذكر/أنثى" و"M/F" ← موحَّد إلى "ذكر/أنثى"

(2) إذا اختلفت الصيغ أو الترميزات، فقد يعامل البرنامج القيم المتشابهة على أنها مختلفة. وإذا خُزنت الأرقام كنصوص، فقد تفشل بعض الحسابات أو تتعامل الأداة معها بوصفها نصوصًا. لذلك تُفحص الأنواع والصيغ قبل التحليل بحسب الأداة.

مصطلحات أساسية

القيمة المفقودة – قيمة غائبة.

القيمة الشاذة – قيمة بعيدة عن باقي القيم.

القيمة غير الطبيعية – قيمة شاذة ناتجة عن خطأ.



احذف السجلات المكررة فعليًا فقط.

توقّف وفكّر

الطول بالسنتيمتر والوزن بالكيلوغرام – لماذا قد تهيمن الأرقام الأكبر على التحليل؟

nezakr.natega4dk.net

5. التحويل الكمي والتوحيد القياسي



التحجيم يغيّر مقياس المتغيرات عندما تكون الخوارزمية حساسة للمقياس.

(1) في بعض الخوارزميات الحساسة للمقياس، قد يكون للمتغير ذي النطاق الأكبر تأثير أكبر لمجرد مقياسه. عندئذ يمكن استخدام طريقة تحجيم مناسبة، مثل تحجيم Min-Max أو التوحيد القياسي.

الطريقة	الوصف	الغرض الرئيسي
تحجيم Min-Max	تحويل البيانات إلى نطاق من 0 إلى 1	وضع المتغيرات في نطاقات متقاربة عندما تحتاج الخوارزمية إلى ذلك
التوحيد القياسي	تحويل البيانات إلى صيغة متوسطها 0 وانحرافها المعياري 1	التعبير عن بُعد القيمة عن المتوسط بوحدات من الانحراف المعياري

(2) يمكن لتحجيم Min-Max تحويل كل متغير إلى نطاق من 0 إلى 1 باستخدام حديه في البيانات. وقد يساعد ذلك خوارزمية حساسة للمقياس على التعامل مع متغيرات مختلفة دون أن يسيطر أحدها بسبب نطاقه، لكنه ليس مطلوبًا لكل تحليل.

(3) يطرح التوحيد القياسي المتوسط من كل قيمة ثم يقسم على الانحراف المعياري، فيصبح متوسط البيانات المستخدمة في الحساب 0 وانحرافها المعياري 1. وتمثل القيمة الجديدة عدد الانحرافات المعيارية عن المتوسط، مع اعتماد التفسير على شكل التوزيع.

6. إعادة تشكيل البيانات حسب الغرض

(1) **إعادة تشكيل البيانات**: تغيير بنية الجدول أو دمج الجداول أو تلخيصها بما يناسب السؤال. ومن عملياتها التحويل المحوري (Pivot) والربط والتجميع.

العملية	الوصف	مثال
التحويل المحوري (Pivot)	تحويل البيانات بين شكل طويل وعريض وفق صفوف وأعمدة محددة	إعادة تشكيل بيانات التاريخ والمنتج والمبيعات المرتبة عموديًا إلى جدول "التاريخ × المنتج"
(الربط)	دمج بيانات من جداول متعددة باستخدام مفتاح أو شرط مطابق	ربط سجل الطلاب وورقة الدرجات عبر "رقم الطالب" في جدول واحد
(التجميع)	حساب المجاميع والمتوسطات وما إلى ذلك لكل مجموعة وتلخيصها	حساب الإجمالي الشهري من المبيعات اليومية

nezakr.natega4dk.net



الشكل 5.2.2 – تحويل المحور: إعادة تشكيل البيانات ذات الصيغة الطويلة إلى جدول متقاطع

(2) تعتمد العملية التي يجب إجراؤها على غرض التحليل. على سبيل المثال، إذا أردت رؤية "اتجاهات المبيعات الشهرية"، فأنت بحاجة إلى تجميع بيانات المبيعات اليومية حسب الشهر. من المهم توضيح غرض التحليل أولًا، ثم التخطيط لكيفية إعادة تشكيل البيانات.

الفكرة الرئيسة

التحضير يحدد الجودة. أصلح القيم المفقودة (احذف أو عوّض أو ضع علامة)، واحكم على القيم الشاذة حسب السبب، وأزل التكرارات الحقيقية، وحوّل أنواع البيانات ووحّد الصيغ، وأعد ضبط المقياس بالتحويل الكمي أو التوحيد القياسي حتى لا يهيمن أي متغير بحكم حجمه، وأعد التشكيل (المحور، الربط، التجميع) ليناسب السؤال – مدخلات خاطئة – مخرجات خاطئة.

✎ مثال محلول

(1) ضع علامة ○ إذا كانت العبارة صحيحة أو × إذا كانت خاطئة أمام كل عبارة (أ - د).

أ تنظيف البيانات هو عمل يُجرى قبل جمع البيانات.

ب استجابات القيم المفقودة هي الحذف والتعويض ووضع العلامة – ثلاث استجابات في المجموع.

ج جميع القيم الشاذة ناتجة عن أخطاء في الإدخال أو القياس، لذا يجب حذفها دائمًا.

د التوحيد القياسي هو معالجة تحوّل البيانات إلى صيغة متوسطها 0 وانحرافها المعياري 1.

(2) طابق كل وصف (1 - 3) مع المعالجة الأنسب من الخيارات أدناه (أ - ج).

1 تحويل البيانات إلى نطاق من 0 إلى 1

2 معالجة تربط مجموعات بيانات مختلفة عبر عنصر مشترك

3 معالجة تحسب المجاميع والمتوسطات وما إلى ذلك لكل مجموعة وتلخّصها

[الخيارات] أ التجميع ب الربط ج التحويل الكمي

💡 توقّف وفكّر

هل يوم مبيعات ضخم في متجر ما دائمًا خطأ – أم قد يكون حقيقيًا؟

nezakr.natega4dk.net

✓ الحلّ

(1)

أ تنظيف البيانات هو عمل ترتيب البيانات المجمَّعة، ويُجرى بعد الجمع. لذلك، x.

ب صحيح كاستجابات للقيم المفقودة. لذلك، ○.

جـ تشمل القيم الشاذة كلًا من القيم غير الطبيعية الناتجة عن أخطاء الإدخال والقيم التي تعكس الواقع. لا تُحذف دائمًا. لذلك، x.

د تعريف صحيح للتوحيد القياسي. لذلك، ○.

(2)

1: التحويل إلى 0-1 هو التحويل الكمي. **جـ**

2: الربط عبر عنصر مشترك هو الربط. **ب**

3: التلخيص حسب المجموعة هو التجميع. **أ**

📝 سؤال على نمط الامتحان

اشرح، باستخدام "مدخلات خاطئة – مخرجات خاطئة"، لماذا يهم تحضير البيانات بقدر أهمية التحليل نفسه. **[6]**

استعن بـ: إصلاح واحد للتنظيف وتحويل واحد.

تدرّب

1 أجب عن الأسئلة التالية.

1. ما المصطلح الذي يشير إلى المعالجة التي تملأ الأجزاء المفقودة من البيانات بقيم مثل المتوسط أو الوسيط؟
2. ما المصطلح الذي يشير إلى المعالجة التي تحوّل البيانات إلى نطاق من 0 إلى 1؟
3. ما المصطلح الذي يشير إلى المعالجة التي تحوّل البيانات إلى صيغة متوسطها 0 وانحرافها المعياري 1؟
4. ما المصطلح الذي يشير إلى العملية التي تربط مجموعات بيانات متعددة عبر عنصر مشترك؟

2 أجب عن الأسئلة التالية.

(1) من بين الخيارات التالية (أ - د)، اختر الخيار الذي لا يُعد استجابة مناسبة للقيم المفقودة.

أ عندما تكون القيم المفقودة قليلة والتأثير على الكل صغيرًا، احذف الصفوف التي تحتوي على قيم مفقودة.

ب عندما تكون القيم المفقودة موزعة عشوائيًا، املأ الأجزاء المفقودة بقيم مثل المتوسط.

جـ عندما يكون للغياب نفسه معنى، اترك علامة تشير إلى أن القيمة كانت مفقودة وحلّل البيانات.

د عوّض بالمتوسط وحلّل البيانات في العناصر التي يكون معظم بياناتها مفقودًا.

(2) لكل من الموقفين (1، 2)، اختر المعالجة الأنسب للبيانات من الخيارات أدناه (أ، ب).

1 تريد مواءمة بيانات التاريخ ذات الصيغ المختلطة مثل "2026/4/1" و"01-04-2026" و"01 Apr 2026".

2 تريد مقارنة متغيرات ذات وحدات ومقاييس مختلفة معًا، مثل الطول (بالسنتيمتر) والوزن (بالكيلوغرام).

[الخيارات] أ تحويل نوع البيانات (توحيد صيغة التاريخ) ب التحويل الكمي

💡 توقّف وفكّر

إذا كان معظم العمود فارغًا، فهل ملؤه بالمتوسط يساعد – أم يخترع بيانات؟

(3) من بين الخيارات التالية (أ - د)، اختر الخيار الذي يصف التعامل مع القيم الشاذة بأكبر قدر من الدقة.

أ جميع القيم الشاذة أخطاء في البيانات، لذا احذفها كلها فور اكتشافها.

ب بالنسبة للقيم الشاذة، حدّد السبب أولًا: صحّحها أو احذفها إذا كانت ناتجة عن أخطاء إدخال، واحتفظ بها من حيث المبدأ إذا كانت تعكس الواقع.

جـ القيم الشاذة لا تؤثر على التحليل، لذا لا حاجة لمعالجة خاصة.

د يمكن دائمًا حل القيم الشاذة عن طريق استبدالها بالمتوسط.

nezakr.natega4dk.net

مصطلحات أساسية

المحور – صفوف طويلة ← جدول صفوف × أعمدة.

الربط – ربط مجموعات البيانات عبر عنصر مشترك.

التجميع – التلخيص حسب المجموعة.

⚙️ فكّر كمهندس: ابحث ثم قرر

جمع صفك استبيانًا عن وقت الشاشة اليومي، لكن الجدول الخام فوضوي: بعض الخلايا فارغة، وإدخال واحد يقرأ "9999" ساعة، والتواريخ مكتوبة بثلاث طرق مختلفة، وصفان يبدوان متطابقين. تقوم بتحضير البيانات للتحليل.

1 · **افحص البيانات.** اذكر كل مشكلة يمكنك رؤيتها في الجدول أعلاه؛ ولكل مشكلة، حدّد فئتها (قيمة مفقودة، قيمة شاذة، تكرار، أو مشكلة صيغة) واذكر بعبارة واحدة كيف ستشوّه التحليل إذا تُركت دون إصلاح.

2 · قرّر إصلاحًا وخطره. اختر طريقة التعامل مع القيم المفقودة بعد النظر إلى مقدار الفقد وسببه ونمطه، لا النسبة وحدها. وبالنسبة إلى «9999 ساعة»، تحقّق من مصدرها وتعريف المتغير قبل تقرير تصحيحها أو حذفها، فقد تكون خطأً أو وحدة مختلفة أو رمزًا ذا معنى.

3 · خطط للتحويل. حدّد تحويل الصيغة المطلوب لتجميع التواريخ، ثم اسأل: هل يحتاج هذا التحليل أصلًا إلى تحجيم عمود وقت الشاشة؟ وإذا نعم، فما الطريقة المناسبة؟ قارن الخطة مع زميل وحددا الإصلاح الأكثر احتمالًا لتشويه النتائج.

☐ **تحليل الجدول الخام كما هو** ☐ **تنظيفه وتحويله أولًا**

قدّم دليلًا يدعم إجابتك.

هل تواجه صعوبة؟ لا يوجد ترتيب إلزامي واحد لكل حالات التنظيف. افهم المشكلة أولًا، ووحّد الصيغ اللازمة، وتحقق من القيم غير المعتادة، وعالج الفقد، ولا تحذف إلا التكرارات غير المقصودة بعد التحقق منها.

🌍 طبّق ما تعلمته

يسرد جدول مبيعات متجر مبيعات كل يوم للمنتجات A Bg Cg في صفوف منفصلة، لكن المدير يريد صفًا واحدًا لكل يوم مع عمود لكل منتج. سمِّ عملية إعادة التشكيل المطلوبة، وسمِّ العملية الأخرى المطلوبة للحصول على الإجمالي الشهري لكل منتج.

❓ إجابة السؤال الرئيس

أجاب الدرس عن السؤال الرئيس: لماذا ننظف البيانات ونحوّلها؟ يكتشف التنظيف مشكلات مثل القيم المفقودة وغير المعتادة والتكرارات، ويعالجها بحسب سببها ونمطها وسؤال التحليل؛ فلا يُحذف سجل مكرر ظاهريًا أو قيمة بعيدة قبل التحقق. ويجب الفصل بين تغيير نوع البيانات وتوحيد الصيغ والترميز. وإذا كانت الخوارزمية حساسة للمقياس، يمكن استخدام Min-Max أو التوحيد القياسي، ولا يلزم التحجيم لكل تحليل. كما تعيد عمليات Pivot والربط والتجميع تشكيل الجداول لتناسب السؤال. كل قرار يحتاج إلى تبرير وفحص لأثره في النتائج.

nezakr.natega4dk.net

تمارين

1 اختبر فهمك: أجب عن الأسئلة الآتية دون الرجوع إلى قسم «ملحوظة مهمة»، ثم راجع إجاباتك.

2 اقرأ الفقرة التالية وأجب عن كل سؤال.

قد تحتاج البيانات إلى (1) لاكتشاف مشكلاتها والتعامل معها، و(2) لتغيير الصيغ أو البنية بما يناسب التحليل. ومن طرق التعامل مع الفقد (3) إلى جانب الحذف أو الاحتفاظ بمؤشر للفقد. وإذا كانت الخوارزمية حساسة للمقياس، يمكن استخدام (4) لتحويل القيم إلى نطاق 0-1، أو (5) لتحويلها إلى عدد الانحرافات المعيارية عن المتوسط.

1. املأ الفراغات (1) - (5).

3 أجب عن الأسئلة التالية.

(1) من بين الخيارات التالية (أ - د)، اختر الخيار الذي يمثل تحويل نوع البيانات.

أ ملء القيم المفقودة بالمتوسط.

ب توحيد صيغ مثل "2026/4/1" و"01 Apr 2026" إلى "01-04-2026".

ج مواءمة مقاييس الطول والوزن إلى نطاق بين 0 و1.

د تجميع المبيعات اليومية في إجماليات شهرية.

(2) من بين الخيارات التالية (أ - د)، اختر الخيار الذي يصف التحويل الكمي والتوحيد القياسي بأكبر قدر من الدقة.

أ التحويل الكمي هو معالجة تحوّل البيانات إلى صيغة متوسطها 0 وانحرافها المعياري 1.

ب التوحيد القياسي هو معالجة تحوّل البيانات إلى نطاق من 0 إلى 1.

ج يُستخدم التحويل الكمي لمواءمة البيانات ذات الوحدات المختلفة في صورة قابلة للمقارنة.

د يُستخدم التوحيد القياسي لتكبير مقياس المتغيرات.

(3) لكل معالجة (1 - 3)، حدّد ما إذا كانت تقابل المحور أو الربط أو التجميع.

1 ربط سجل الطلاب وورقة الدرجات عبر "رقم الطالب" في جدول واحد.

2 حساب الإجماليات الشهرية من بيانات المبيعات اليومية.

3 إعادة تشكيل مبيعات المنتجات اليومية في صورة جدول "التاريخ × المنتج".

[الخيارات] أ المحور ب الربط ج التجميع

⭐ الخلاصة

تذكّر:
نظّف أولًا (القيم المفقودة، القيم الشاذة، التكرارات)، ثم حوّل (حوّل الأنواع، طبّق التحويل الكمي أو التوحيد القياسي، أعد التشكيل بالمحور والربط والتجميع). كل إصلاح هو حكم يحدده الغرض – مدخلات خاطئة – مخرجات خاطئة.

⭐ فكّر وتحدَّ نفسك

تأمّل: لماذا قد يؤدي "ملء" البيانات المفقودة أحيانًا إلى تضليل أكبر من تركها فارغة؟

تحدٍّ: لديك عمود استبيان واحد 70% من إجاباته فارغة. ناقش ما إذا كنت ستعوّض أو تضع علامة أو تحذف العمود، واذكر الشرط الوحيد الذي قد يغيّر اختيارك.

nezakr.natega4dk.net



الدرس 5-3

البيانات المفتوحة وواجهات برمجة التطبيقات

أهداف التعلم

- شرح مفهوم البيانات المفتوحة، مع تحديد أهم المصادر الموثوقة لها.
- وصف كيفية الحصول على البيانات عبر واجهة برمجة التطبيقات (API)، وشرح أهمية شروط الاستخدام وذكر المصدر.

nezakr.natega4dk.net

توجد مصادر كثيرة توفر بيانات مفيدة بشروط تسمح بإعادة استخدامها. لكن الإتاحة العامة أو المجانية وحدها لا تجعل البيانات مفتوحة؛ فالانفتاح يتطلب ترخيصًا أو شروطًا واضحة تسمح بالوصول وإعادة الاستخدام وإعادة التوزيع. وقد يكون الوصول بالتنزيل أو عبر API أو بطريقة أخرى. لذلك يجب تقييم المصدر وطريقة جمع البيانات وحداثتها وملاءمتها للسؤال، ثم الالتزام بالترخيص ومتطلبات الإسناد.

؟ السؤال الرئيس: من أين تأتي البيانات المفتوحة الموثوقة، وكيف تجلبها البرامج، وما القواعد التي تحكم استخدامها؟

📌 ملحوظة مهمّة

1. استخدام البيانات المفتوحة

(1) عند استخدام بيانات مفتوحة، تحقق من أربعة أمور: المصدر وجودته، وطريقة الوصول، والترخيص أو شروط الاستخدام، ومدى مناسبة البيانات لسؤال التحليل. تساعد هذه الخطوات على الاستخدام المسؤول، لكن جودة النتيجة تعتمد أيضًا على جودة البيانات والمنهج.

(2) قد تنشر الجهات الحكومية والمنظمات والباحثون والشركات بيانات مفتوحة، لكن الترخيص والجودة والتعريفات تختلف من مجموعة إلى أخرى. اختر مصدرًا يناسب السؤال، وافحص الجهة الناشرة وتاريخ التحديث وطريقة الجمع والترخيص.

2. مصادر البيانات المفتوحة الرئيسية

(1) فيما يأتي أمثلة على بوابات ومنصات يمكن أن توفر بيانات مفيدة، مع ضرورة تقييم كل مجموعة وشروطها.

المزوّد	الوصف	الاستخدامات الرئيسية
World Bank) (Open Data	إحصاءات عن الاقتصاد والسكان والتعليم والصحة لدول مختلف أنحاء العالم	يمكن استخدامها للمقارنات الدولية إذا كانت التعريفات والفترات الزمنية مناسبة للسؤال
(UN Data)	إحصاءات دولية تنشرها الأمم المتحدة والوكالات التابعة لها	موضوعات دولية متعددة، مع التحقق من مصدر المؤشر وتعريفه
(Kaggle)	منصة تُنظَّم عليها مسابقات لتحليل البيانات، وتُتاح فيها مجموعات بيانات في مجالات متنوعة للعامة؛ ويجب تقييم مصدر كل مجموعة وترخيصها	الحصول على بيانات لدراسة وبحوث التعلّم الآلي

(2) تختلف طرق الوصول من مصدر إلى آخر: فقد يكون التنزيل مباشرًا، أو عبر API، أو بأدوات أخرى يوفرها المصدر.

⭐ خريطة الدرس

الفكرة الأساسية:

تأتي البيانات المفتوحة من مزودين موثوقين (World Bank Open Data، UN Datag، Kaggle)؛ يمكن تنزيلها من متصفح أو جلبها بواسطة برنامج عبر واجهة برمجة تطبيقات (غالبًا ما تُعيد بيانات بصيغة JSON)؛ ولكل مجموعة بيانات شروط استخدام يجب اتباعها ومصدر يجب الإشارة إليه.

المفاهيم الأساسية:

- البيانات المفتوحة
- World Bank Open Data / UN (United Nations) Data / Kaggle
- واجهة برمجة التطبيقات (API) / نقطة النهاية (endpoint) / JSON
- شروط الاستخدام
- الإشارة إلى المصدر



البيانات المفتوحة عامة ومتاحة لإعادة الاستخدام مجانًا.

مصطلحات أساسية

البيانات المفتوحة – بيانات متاحة للعموم ليستخدمها أي شخص.

World Bank Open Data – إحصاءات الدول عن الاقتصاد والسكان وغيرها.

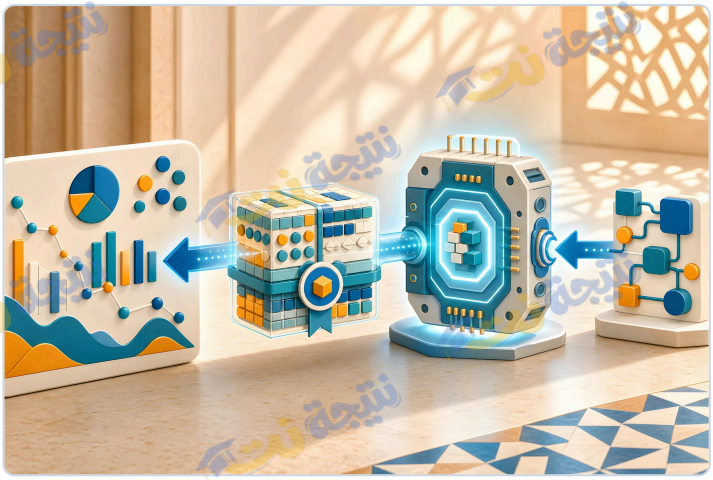
مصطلحات أساسية

واجهة برمجة التطبيقات (API) – قواعد تتفاعل بها البرمجيات، تُستخدم لجلب البيانات.

نقطة النهاية (endpoint) – عنوان URL الذي يُرسَل إليه الطلب.

JSON – صيغة شائعة لبيانات الاسم/القيمة.

3. الحصول على البيانات عبر واجهات برمجة التطبيقات



قد توفر API وصولًا برمجيًا إلى بيانات أو وظائف؛ ومدى حداثة البيانات يعتمد على الخدمة.

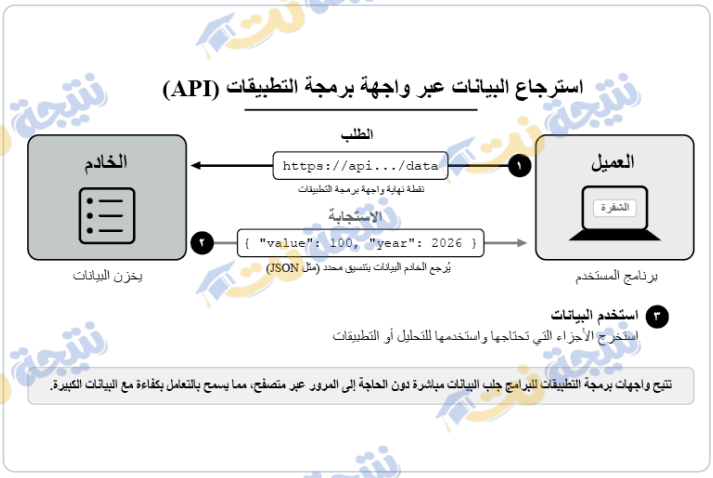
(1) **واجهة برمجة التطبيقات (API)**: مجموعة قواعد تتيح لبرنامج أن يطلب بيانات أو وظائف من برنامج أو خدمة أخرى. ويمكن للبرنامج إرسال الطلب مباشرة إلى API بدل تنفيذ الخطوات يدويًا في صفحة ويب، بما يساعد على تكرار الطلبات ومعالجة النتائج آليًا ضمن حدود الخدمة.

(2) يسير الحصول على البيانات عبر واجهة برمجة التطبيقات وفقًا للتسلسل التالي.

① يرسل العميل طلبًا إلى **نقطة نهاية (Endpoint)**، وهي عنوان أو مسار محدد في API يستقبل نوعًا معينًا من الطلبات.

② يعالج الخادم الطلب، ثم يعيد استجابة قد تحتوي على بيانات أو رسالة خطأ، بصيغة تحددها الخدمة.

③ يقرأ البرنامج الاستجابة، ثم يستخدم البيانات أو يتعامل مع الخطأ بحسب الحالة.



الشكل 5.3.1 – استرجاع البيانات عبر واجهة برمجة التطبيقات: الطلب، والاستجابة، والاستخدام

(3) **JSON (JavaScript Object Notation)** صيغة نصية شائعة لتمثيل البيانات المهيكلة باستخدام كائنات ومصفوفات وقيم مختلفة. تدعمها لغات وأدوات كثيرة، لكنها ليست الصيغة الوحيدة التي قد تستخدمها API.

(4) لا يلغي الوصول عبر API شروط الاستخدام أو الترخيص. اتبع حدود الطلبات وسياسات الحفظ والنشر، وتحقق من نطاق إعادة الاستخدام المسموح.

توقّف وفكّر

البيانات "المفتوحة" لا تزال لها قواعد – ما الأمران اللذان يجب عليك التحقق منهما قبل استخدامها؟

nezakr.natega4dk.net

4. استخدام البيانات والإشارة إلى المصدر



استخدم البيانات بمسؤولية وأشر إلى مصدرها.

(1) للبيانات المفتوحة ترخيص أو شروط تحدد ما يسمح به عند الوصول إليها وإعادة استخدامها وإعادة توزيعها. لذلك يجب قراءة الشروط الخاصة بالمجموعة قبل الاستخدام.

(2) استخدم البيانات وفق الترخيص، وانسبها إلى المصدر بالطريقة المطلوبة. وقد يتضمن الإسناد أو الاستشهاد: الجهة الناشرة، واسم مجموعة البيانات، والإصدار أو تاريخ التحديث أو الوصول، ورابط المصدر، وفق متطلبات الجهة ومعيار الاستشهاد.

(3) لا يثبت ذكر المصدر وحده موثوقية البيانات، لكنه يساعد الآخرين على العثور عليها وتقييمها ومراجعة التحليل. وتظل جودة البيانات والمنهج بحاجة إلى تقييم مستقل.

الفكرة الرئيسة

البيانات المفتوحة قوية لكنها لا تزال تخضع لقواعد. اختر مزودًا موثوقًا يناسب غرضك (Kaggle وUN Data وWorld Bank Open Data)، واجلبها عبر المتصفح أو عبر واجهة برمجة تطبيقات – حيث يرسل العميل طلبًا إلى نقطة نهاية ويُعيد الخادم البيانات، عادةً بصيغة JSON – واتبع دائمًا شروط الاستخدام وأشر إلى المصدر حتى يمكن الوثوق بتحليلك والتحقق منه.

مثال محلول

(1) ضع علامة ○ إذا كانت العبارة صحيحة أو × إذا كانت خاطئة أمام كل عبارة (أ - د).

أ تنشر World Bank Open Data إحصاءات عن الاقتصاد والسكان والتعليم وما إلى ذلك لدول العالم.

ب باستخدام واجهة برمجة التطبيقات، يمكن الحصول على البيانات اللازمة مباشرة دون المرور عبر متصفح.

جـ طالما أن البيانات مفتوحة، يمكن استخدامها بحرية دون التحقق من شروط الاستخدام.

د عند استخدام البيانات، يُطلَب الإشارة إلى المزوّد، واسم البيانات، وتاريخ الحصول عليها، وعنوان URL، وما إلى ذلك كمصدر.

(2) طابق كل وصف (1 - 3) مع مزود البيانات المفتوحة الأنسب من الخيارات أدناه (أ - جـ).

1 ينشر إحصاءات عن الاقتصاد والسكان والتعليم وما إلى ذلك لدول العالم.

2 يوفر إحصاءات دولية تنشرها الأمم المتحدة والوكالات التابعة لها.

3 منصة تُجري مسابقات تحليل البيانات.

[الخيارات] أ UN Data ب World Bank Open Data جـ Kaggle

توقّف وفكّر

في استدعاء واجهة برمجة التطبيقات، من الذي يرسل الطلب ومن الذي يُعيد البيانات؟

nezakr.natega4dk.net

✓ الحلّ

(1)

أ وصف صحيح لـWorld Bank Open Data. لذلك، ○.

ب وصف صحيح لواجهة برمجة التطبيقات. لذلك، ○.

جـ حتى مع البيانات المفتوحة، تختلف شروط الاستخدام لكل مجموعة بيانات، لذا يجب دائمًا التحقق منها قبل الاستخدام. لذلك، ×.

د محتوى صحيح للإشارة إلى المصدر. لذلك، ○.

(2)

1: إحصاءات عن دول العالم هي World Bank Open Data. **ب**

2: بيانات من الأمم المتحدة والوكالات التابعة لها هي UN Data. **أ**

3: المنصة التي تُجري مسابقات تحليل البيانات هي Kaggle. **جـ**

تدرّب

1 أجب عن الأسئلة التالية.

1. ما اسم بوابة البيانات المفتوحة التابعة للبنك الدولي التي تنشر إحصاءات عن الاقتصاد والسكان والتعليم وما إلى ذلك لدول العالم؟
2. ما المصطلح الذي يشير إلى قواعد تفاعل أجزاء البرمجيات مع بعضها البعض، والذي يُستخدم أيضًا على نطاق واسع في خدمات الويب للحصول على البيانات؟
3. اذكر صيغة نصية شائعة قد تستخدمها API لتمثيل البيانات المهيكلة.
4. ما المصطلح الذي يعني ذكر معلومات المصدر بالطريقة التي يطلبها الترخيص أو معيار الاستشهاد؟

2 أجب عن الأسئلة التالية.

(1) من بين الخيارات التالية (أ - د)، اختر الخيار الذي يصف تسلسل الحصول على البيانات عبر واجهة برمجة التطبيقات بأكبر قدر من الدقة.

أ يرسل العميل طلبًا إلى الخادم، ويُعيد الخادم البيانات بصيغة مثل JSON.

ب يستمر الخادم من جانب واحد في إرسال البيانات إلى العميل.

جـ يتبادل العملاء البيانات مباشرة مع بعضهم البعض.

د الخادم يخزّن البيانات فقط؛ ولا يمكن للعملاء استرجاعها.

(2) لكل من الغرضين (1، 2)، اختر مزود البيانات المفتوحة الأنسب من الخيارات أدناه (أ - جـ).

1 تريد مقارنة الناتج المحلي الإجمالي بين الدول.

2 تريد العثور على مجموعات بيانات من مجالات متنوعة لدراسة التعلّم الآلي.

[الخيارات] أ UN Data **ب** World Bank Open Data **جـ** Kaggle

(3) من بين الخيارات التالية (أ - د)، اختر الخيار الذي يصف استخدام البيانات المفتوحة بأكبر قدر من الدقة.

أ يمكن استخدام البيانات المفتوحة بحرية، لذا لا حاجة للتحقق من شروط الاستخدام.

ب الإشارة إلى المصدر مطلوبة فقط للبيانات المنشورة مقابل رسوم.

جـ حتى عند استخدام البيانات المفتوحة، يُطلَب الإشارة إلى المزوّد، واسم البيانات، وتاريخ الحصول عليها، وعنوان URL، وما إلى ذلك كمصدر.

د بالنسبة للبيانات التي يتم الحصول عليها عبر واجهة برمجة تطبيقات، لا حاجة لاتباع شروط الاستخدام.

📝 سؤال على نمط الامتحان

اشرح لماذا يجب على المحلل التحقق من شروط الاستخدام والإشارة إلى المصدر حتى عندما تكون البيانات مجانية ومفتوحة. **[6]**

استعن بـ: حقك في استخدام البيانات، والموثوقية، والسماح للآخرين بالتحقق.

💡 توقّف وفكّر

أي مزوّد يناسب أي غرض – وهل ستستخدم متصفحًا أم واجهة برمجة تطبيقات؟

nezakr.natega4dk.net

مصطلحات أساسية

شروط الاستخدام – القواعد المحددة لكل مجموعة بيانات.

الإشارة إلى المصدر – المزوّد، الاسم، التاريخ، عنوان URL.

⚙️ فكّر كمهندس: ابحث ثم قرر

من أجل مشروع مدرسي، تريد إظهار كيف تغيّر الوصول إلى الإنترنت في مصر خلال السنوات العشر الماضية مقارنة بدولتين مجاورتين. يجب عليك اختيار مصدر بيانات مفتوحة وطريقة للحصول عليها.

1 · **افحص مصدرين.** ابحث عن مزودَين مرشحين لهذا السؤال بالتحديد (على سبيل المثال World Bank Open Data وUN Datag). لكل منهما، دوّن ما إذا كان يحتوي فعليًا على أرقام الوصول إلى الإنترنت حسب الدولة والسنة، ومدى حداثة آخر سنة متاحة.

2 · **حلّل الملاءمة والقواعد.** قارن بين المصدرين من ناحية لا يذكرها جدول الدرس – مثل إلى أي مدى تمتد البيانات في الماضي، أو ما إذا كانت شروط استخدامه تسمح لك بإعادة نشر رسم بياني في تقرير عام. اذكر خطرًا واحدًا لاختيار المصدر الخاطئ هنا.

3 · قرّر. اختر مصدرًا وطريقة وصول وبرّر كليهما. ثم اكتب بيانات المصدر وفق متطلبات الترخيص وطريقة الاستشهاد المعتمدة في الكتاب. قارن اختيارك مع زميل وحددا أي مصدر أنسب للسؤال ولماذا.

☐ **خذ أول مجموعة بيانات تجدها** ☐ **اختر مصدرًا موثوقًا واتبع قواعده**

قدّم دليلًا يدعم إجابتك.

هل تواجه صعوبة؟ ابدأ من السؤال، لا من المصدر: اذكر بالضبط الحقول التي تحتاجها (الدولة، السنة، معدل الوصول إلى الإنترنت)، ثم احتفظ فقط بالمزودين الذين يحتوون فعليًا على الحقول الثلاثة جميعها – وتحقق من شروط كل منهم قبل الالتزام.

🌍 طبّق ما تعلمته

يعرض تطبيق طقس توقعات اليوم من خدمة عامة. صف الطلب والاستجابة بين التطبيق والخادم، واذكر صيغة بيانات محتملة مثل JSON، مع التنبيه إلى أن الصيغة تعتمد على الخدمة.

❓ إجابة السؤال الرئيس

أجاب الدرس عن السؤال الرئيس: كيف نختار بيانات مفتوحة ونصل إليها ونستخدمها؟ البيانات المفتوحة متاحة وفق ترخيص أو شروط تسمح بإعادة الاستخدام، وليست كل البيانات المجانية مفتوحة. ويجب تقييم الجهة الناشرة وطريقة الجمع والتعريفات والحداثة والملاءمة للسؤال. قد يوفر المصدر تنزيلًا مباشرًا أو API؛ فيرسل العميل طلبًا إلى نقطة نهاية ويعيد الخادم استجابة قد تحتوي على JSON أو صيغة أخرى أو رسالة خطأ. ولا يلغي الوصول عبر API الترخيص أو حدود الاستخدام. اتبع الشروط وانسب البيانات بالطريقة المطلوبة، مع تقييم جودة البيانات والمنهج بصورة مستقلة.

nezakr.natega4dk.net

تمارين

1 اختبر فهمك: أجب عن الأسئلة الآتية دون الرجوع إلى قسم «ملحوظة مهمة»، ثم راجع إجاباتك.

2 اقرأ الفقرة التالية وأجب عن كل سؤال.

لاستخدام البيانات المفتوحة يجب معرفة المصدر وطريقة الوصول وشروط الاستخدام. يمكن العثور على مؤشرات دولية في بوابات مثل (1)، مع فحص تعريف المؤشر ومصدره.

وقد يوفر المصدر (2) ليتلقى البرنامج استجابة بصيغة مثل (3) بحسب الخدمة. ويجب دائمًا التحقق من (4) الخاصة بالمجموعة، ثم تقديم (5) إلى المصدر وفق الترخيص ومعيار الاستشهاد.

1. املأ الفراغات (1) - (5).

3 أجب عن الأسئلة التالية.

(1) من بين الخيارات التالية (أ - د)، اختر الخيار الذي يصف Kaggle بأكبر قدر من الدقة.

أ بوابة يديرها البنك الدولي تنشر إحصاءات لدول العالم.

ب موقع للإحصاءات الدولية تديره الأمم المتحدة.

ج منصة تُجري مسابقات تحليل البيانات، تُتاح فيها مجموعات بيانات في مجالات متنوعة للعموم.

د نظام يسجّل تلقائيًا سجلات تصفح المواقع الإلكترونية.

(2) من بين الخيارات التالية (أ - د)، اختر الخيار الذي لا يُعد وصفًا مناسبًا للإشارة إلى المصدر.

أ الإشارة إلى المصدر لا تُثبت وحدها موثوقية البيانات، لكنها تتيح للآخرين العثور عليها وتقييمها.

ب الإشارة إلى المصدر مهمة للسماح للآخرين بالتحقق من التحليل.

ج تشمل المصادر مزود البيانات، واسم البيانات، وتاريخ الحصول عليها، وعنوان URL، وما إلى ذلك.

د عند استخدام البيانات المفتوحة، لا حاجة للإشارة إلى المصدر.

(3) من بين الخيارات التالية (أ - د)، اختر الخيار الذي لا يُعد نقطة تنبيه مناسبة عند استخدام واجهة برمجة تطبيقات.

أ استخدم واجهة برمجة التطبيقات وفقًا لشروط الاستخدام التي يحددها المزوّد.

ب تجنّب إرسال طلبات مفرطة في وقت قصير.

ج تحقق مسبقًا من نطاق الاستخدام المسموح به للبيانات التي تم الحصول عليها.

د بسبب طريقة عمل واجهات برمجة التطبيقات، يمكن استخدامها دون التحقق من شروط الاستخدام.

⭐ الخلاصة

تذكّر:
اختر مصدر بيانات مفتوحة موثوقًا يناسب غرضك (World Bank Open Data، UN Data، Kaggle)، واجلبها عبر المتصفح أو واجهة برمجة التطبيقات (يطلب العميل نقطة نهاية؛ ويُعيد الخادم JSON)، واتبع دائمًا شروط الاستخدام وأشر إلى المصدر.

⭐ فكّر وتحدَّ نفسك

تأمّل: فكّر في رسم بياني رأيته على الإنترنت بدون مصدر. هل وثقت به – وما السطر الواحد الذي كان سيغيّر ذلك؟

تحدٍّ: يُسمح لك بذكر سطر مصدر واحد فقط لرسم بياني مبني من مجموعة بيانات للبنك الدولي قمت بتنزيلها اليوم. اكتب سطر الإشارة إلى المصدر هذا، وقل أي عنصر واحد لن تحذفه أبدًا ولماذا.

nezakr.natega4dk.net



الدرس 6-1
الاستدلال الإحصائي

أهداف التعلم

- شرح الفرق بين التقدير بنقطة والتقدير بفترة، بما في ذلك معنى مستوى الثقة.
- وصف إجراءات اختبار الفرضيات، موضحًا المحاذير الواجب مراعاتها عند تفسير نتائجه.

nezakr.natega4dk.net

قد يصعب دراسة كل فرد في المجتمع، لذا ندرس عينة منه ونستخدم نتائجها لتقدير خصائص المجتمع أو اختبار ادعاءات عنه؛ وهذا هو الاستدلال الإحصائي. يجيب التقدير عن سؤال مثل «ما القيمة التقريبية؟»، بينما يسأل اختبار الفرضيات: «هل توفر البيانات دليلًا كافيًا ضد فرضية العدم؟». وبما أن العينة جزء من المجتمع، تحمل النتائج قدرًا من عدم اليقين ينبغي قياسه والإفصاح عنه. كما أن الدلالة الإحصائية لا تعني بالضرورة الأهمية العملية.

؟ السؤال الرئيس: كيف يمكننا الاستنتاج من عينة إلى مجتمع كامل، وكيف نقيس عدم اليقين في هذا الاستنتاج؟

ملحوظة مهمّة

1. مرحلة الاستدلال الإحصائي

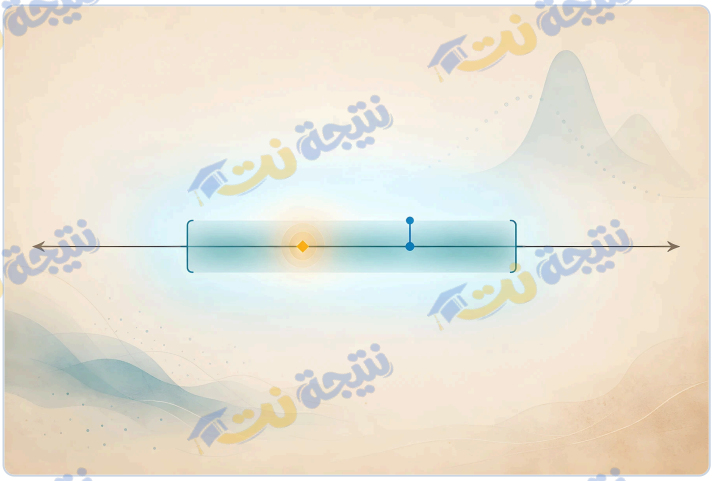
(1) **الاستدلال الإحصائي**: استخدام بيانات عينة لتقدير خصائص المجتمع أو اختبار ادعاءات عنه، مع مراعاة عدم اليقين وشروط تصميم الدراسة.

(2) يمكن تقسيم الاستدلال الإحصائي بشكل عام إلى النهجين التاليين.

النهج	الوصف	مثال
(التقدير)	استخدام بيانات العينة لتقدير خصائص المجتمع بقيمة واحدة أو بنطاق من القيم	استنتاج متوسط طول المجتمع من متوسط طول العينة
(اختبار الفرضيات)	تقييم مدى تعارض البيانات مع فرضية العدم في ظل نموذج وافتراضات محددة	اختبار ما إذا كانت البيانات تقدم دليلًا على اختلاف الأداء بين طريقتين

(3) في كلتا الحالتين، بما أن العينة ليست سوى جزء من المجتمع، فإن الاستدلال يتضمن دائمًا عدم يقين. في الاستدلال الإحصائي، يُتخذ الحكم مع مراعاة درجة ذلك عدم اليقين.

2. التقدير النقطي والتقدير الفاصل



التقدير النقطي قيمة واحدة؛ أما الفترة فهي مدى.

(1) **التقدير النقطي**: تقدير خاصية للمجتمع بقيمة واحدة تحسب من العينة. فإذا كان متوسط طول العينة 168 سم، نستخدم 168 سم تقديرًا نقطيًا لمتوسط طول المجتمع.

(2) **(التقدير الفاصل)**...تقدير أن خصائص المجتمع تقع ضمن مدى معين (فاصل). *على سبيل المثال*، يُقدَّر متوسط طول المجتمع بأنه يقع في المدى من 166 سم إلى 170 سم.

(3) يسمى الفاصل **فترة ثقة**. ويصف **مستوى الثقة** أداء طريقة بناء الفترات عند تكرار أخذ العينات، لا احتمالًا لاحقًا لاحتواء فترة بعينها. فعند مستوى ثقة 95%، إذا كررنا أخذ عينات وبناء فترات بالطريقة نفسها مرات كثيرة، فستحتوي نحو 95% من الفترات على القيمة الحقيقية. ويُختار المستوى قبل التحليل وفق الغرض.

⭐ خريطة الدرس

الفكرة الأساسية:

الاستدلال الإحصائي يستنتج من العينة إلى المجتمع. التقدير يعطي قيمة (نقطية) أو مدى (فاصل، مع مستوى ثقة)؛ واختبار الفرضيات يوازن بين فرضية العدم والفرضية البديلة باستخدام القيمة الاحتمالية ومستوى الدلالة. كل ذلك يحمل قدرًا من عدم اليقين – والدلالة الإحصائية ليست هي نفسها الأهمية العملية.

المفاهيم الأساسية:

- الاستدلال الإحصائي
- التقدير النقطي / الفاصل
- فترة الثقة ومستوى الثقة
- فرضية العدم / الفرضية البديلة
- القيمة الاحتمالية ومستوى الدلالة
- أخطاء النوع الأول / الثاني

مصطلحات أساسية

الاستدلال الإحصائي – الاستنتاج من العينة إلى المجتمع.

التقدير – استنتاج خصائص المجتمع بالأرقام.

اختبار الفرضيات – الحكم على ادعاء بخصوص المجتمع.

💡 توقّف وفكّر

قيمة واحدة أم مدى – أي تقدير يُظهر لك مدى عدم اليقين؟

nezakr.natega4dk.net

(4) التقدير النقطي سهل الفهم لأنه يُعرض كقيمة واحدة، ولكن عدم يقين التقدير يكون غير مرئي. أما التقدير الفاصل فله عرض، لذا يلزم الحذر في تفسير النتيجة، لكنه يتميز بإظهار درجة عدم اليقين أيضًا.

3. فكرة اختبار الفرضيات



يبدأ اختبار الفرضيات بفرضية العدم، ثم يقيس مدى توافق البيانات معها.

(1) **اختبار الفرضيات**: إجراء يستخدم البيانات لتقييم الدليل ضد فرضية العدم، ولا يثبت صحة فرضية أو خطأها يقينًا.

(2) في اختبار الفرضيات، تُوضع الفرضيتان التاليتان.

الفرضية	الوصف
(فرضية العدم) (H0)	فرضية تُستخدم كنقطة بداية للاختبار، مثل "لا فرق" أو "لا علاقة"
(الفرضية البديلة) (H1)	فرضية تصف نوع الفرق أو العلاقة التي يهتم بها السؤال، ونبحث عن دليل إحصائي يدعمها عندما لا تتوافق البيانات بما يكفي مع H_0

(3) نحدد فرضية عدم مرجعية ونستخدم اختبارًا مناسبًا لقياس مدى ندرة النتيجة التي رأيناها، أو نتيجة أشد تطرفًا، إذا كانت H_0 صحيحة. وكلما كانت النتيجة أقل توافقًا مع H_0 وفق نموذج الاختبار، زاد الدليل الإحصائي ضدها. وهذا لا يساوي احتمال أن H_0 صحيحة أو خاطئة.

(4) بناءً على التقييم، يُحكم على فرضية العدم إما بأنها **(مرفوضة (منفية))** أو **(غير مرفوضة)**.

4. إجراء اختبار الفرضيات

(1) يتبع اختبار الفرضيات هذا التسلسل.

① حدّد فرضية العدم H_0 والفرضية البديلة H_1 بما يناسب السؤال قبل النظر إلى النتيجة

② حدّد مستوى الدلالة α مسبقًا؛ ومن القيم الشائعة 0.05 و0.01

③ اختر الاختبار المناسب، وتحقق من افتراضاته، ثم احسب القيمة الاحتمالية (p-value) من البيانات

④ قارن القيمة الاحتمالية بمستوى الدلالة واتخذ حكمًا

(2) **القيمة الاحتمالية (p-value)**: احتمال الحصول على نتيجة مساوية في التطرف للنتيجة المرصودة أو أشد منها، بافتراض صحة H_0 وبحسب نموذج الاختبار. وكلما صغرت، كانت النتيجة أقل توافقًا مع H_0؛ لكنها ليست احتمال صحة H_0.

(3) **مستوى الدلالة (α)**: القيمة الحدية التي يحددها الباحث مسبقًا لقرار رفض H_0، وترتبط باحتمال خطأ النوع الأول ضمن شروط الاختبار.

(4) معايير الحكم

مصطلحات أساسية

فرضية العدم (H0) – نقطة البداية القائلة بـ"لا فرق".

الفرضية البديلة (H1) – تُتبنى إذا رُفضت H0.

القيمة الاحتمالية – احتمال الحصول على بيانات بهذا التطرف أو أشد إذا كانت H0 صحيحة.

توقّف وفكّر

إذا كانت القيمة الاحتمالية 0.02 ومستوى الدلالة 0.05، هل ترفض فرضية العدم أم لا؟

nezakr.natega4dk.net

الشرط	الحكم
$p \le \alpha$	نرفض H_0، وتكون النتيجة ذات دلالة إحصائية وفق α المحدد، من دون إثبات H_1 يقينًا
$p > \alpha$	عدم رفض فرضية العدم (لا يمكن القول بوجود دلالة إحصائية)



شكل 6.1.1 – إجراء اختبار الفرضيات موضحًا من خلال مثال

5. طرق اختبار الفرضيات الرئيسية

(1) تختلف طريقة الاختبار المستخدمة باختلاف غرض التحليل ونوع البيانات. من أبرز الأمثلة ما يلي.

الاختبار	الاستخدام الرئيسي	مثال
(اختبار t)	عائلة اختبارات لمقارنة المتوسطات في حالات محددة: يعرض الدرس مثال مقارنة متوسط مجموعتين	فحص ما إذا كان متوسط درجات اختبار الطلاب يختلف بين طريقة تعلم معينة والطريقة التقليدية
(اختبار مربع كاي)	اختبار كاي-تربيع للاستقلال: فحص وجود ارتباط بين متغيرين فئويين في جدول تقاطعي	فحص، من نتائج المسح، ما إذا كانت هناك علاقة بين الجنس وتفضيل المادة الدراسية

توقّف وفكّر

أي اختبار يقارن متوسطي مجموعتين، وأيهما يفحص علاقة بين فئات؟

6. تنبيهات بخصوص الاستدلال

(1) في الاستدلال الإحصائي، هناك دائمًا احتمال لأخطاء الحكم. هناك نوعان من الأخطاء.

نوع الخطأ	الوصف
(خطأ من النوع الأول)	رفض فرضية العدم بينما تكون في الواقع صحيحة
(خطأ من النوع الثاني)	عدم رفض فرضية العدم بينما تكون في الواقع خاطئة

مصطلحات أساسية

خطأ من النوع الأول – رفض فرضية عدم صحيحة.

خطأ من النوع الثاني – الإبقاء على فرضية عدم خاطئة.

المفاضلة – خفض أحد الخطأين يرفع الآخر.

(2) عند ثبات حجم العينة وتصميم الدراسة وبقية العوامل، يقلل خفض α احتمال خطأ النوع الأول، لكنه قد يزيد احتمال خطأ النوع الثاني. ويمكن تحسين التصميم أو زيادة حجم العينة أحيانًا لتقليل كليهما.

(3) يلزم الحذر في تفسير القيمة الاحتمالية. يجب تجنب سوء الفهم التالي.

① القيمة الاحتمالية ليست "احتمال صحة فرضية العدم".

② القيمة الاحتمالية الصغيرة لا تعني "أن التأثير كبير".

nezakr.natega4dk.net

③ قرار p مقابل α قاعدة إجرائية في الاختبار، لكنه لا يثبت صحة النتيجة أو أهميتها العملية، ولا يحدد الحقيقة العلمية وحده.

(4) كون النتيجة ذات دلالة إحصائية وكونها مهمة عمليًا أمران مختلفان. *على سبيل المثال*، لنفترض أن تجربة تقارن بين تصميمين لموقع إلكتروني، جديد وقديم، ببيانات معدل النقر من 100,000 شخص. إذا كان معدل النقر للتصميم الجديد أعلى بمقدار 0.1 نقطة مئوية فقط من التصميم القديم، فبسبب كبر حجم مجموعة البيانات، من المرجح أن يكون هذا الفرق ذا دلالة إحصائية. ومع ذلك، فإن ما إذا كان هذا الفرق البالغ 0.1 نقطة يؤدي إلى ربح فعلي أو تحسين لتجربة المستخدم فهو أمر يحتاج إلى الحكم عليه من منظور منفصل.

(5) الاستدلال الإحصائي وسيلة قوية لاستخلاص استنتاجات من البيانات، ولكن من المهم فهم أنه أداة للحكم تتضمن عدم يقين، ويجب تفسير النتائج بشكل مناسب.



خطأ النوع الأول هو إنذار كاذب؛ وخطأ النوع الثاني هو تفويت للاكتشاف.

الفكرة الرئيسة

الاستدلال يستنتج من عينة إلى مجتمع، ودائمًا مع وجود عدم يقين. التقدير يُبلغ عن قيمة (نقطية) أو مدى (فاصل) مع مستوى ثقة؛ واختبار الفرضيات يضع فرضية عدم وفرضية بديلة، ثم يقارن قيمة احتمالية بمستوى دلالة لرفضها أو عدم رفضها. احترس من أخطاء النوع الأول والثاني (وهي علاقة مفاضلة)، ولا تقرأ القيمة الاحتمالية أبدًا على أنها احتمال صحة فرضية العدم، وتذكر أن الدلالة الإحصائية ليست هي نفسها الأهمية العملية.

توقّف وفكّر

مع 100,000 شخص، يمكن أن يكون فرق ضئيل "ذا دلالة". هل هذا يجعله مهمًا؟

مثال محلول

(1) ضع علامة أمام كل عبارة (أ - د) بـ○ إذا كانت صحيحة أو × إذا كانت خاطئة.

أ استنتاج خصائص المجتمع من العينة يُسمى الاستدلال الإحصائي.

ب التقدير النقطي طريقة تقدر خصائص المجتمع ضمن مدى معين.

ج في اختبار الفرضيات، يُقيَّم احتمال أن يؤدي جمع البيانات إلى الحصول على البيانات المتاحة بافتراض صحة فرضية العدم.

د إذا كانت نتيجة ما ذات دلالة إحصائية، فإن الفرق يكون بالضرورة مهمًا عمليًا أيضًا.

(2) طابق كل وصف (1 - 3) مع المصطلح الأنسب من الخيارات أدناه (أ - ج).

1 فرضية تُستخدم كنقطة بداية للاختبار، مثل "لا فرق" أو "لا علاقة"

2 احتمال الحصول على نتيجة تساوي في تطرفها البيانات المتاحة أو أكثر تطرفًا منها، بافتراض صحة فرضية العدم

3 القيمة الحدية لتحديد ما إذا كان يجب رفض فرضية العدم

[الخيارات] أ القيمة الاحتمالية ب مستوى الدلالة ج فرضية العدم

الحلّ

(1)

أ تعريف صحيح للاستدلال الإحصائي. لذلك، ○.

ب التقدير ضمن مدى هو التقدير الفاصل. التقدير النقطي يُقدِّر قيمة واحدة. لذلك، ×.

ج وصف صحيح لفكرة اختبار الفرضيات. لذلك، ○.

د الدلالة الإحصائية والأهمية العملية أمران مختلفان؛ مع كمية كبيرة من البيانات، حتى الفرق الضئيل يمكن أن يصبح ذا دلالة. لذلك، ×.

(2)

1: الفرضية المستخدمة كنقطة بداية للاختبار هي فرضية العدم. ج

2: احتمال الحصول على البيانات في ظل فرضية العدم هو القيمة الاحتمالية. أ

3: القيمة الحدية للرفض هي مستوى الدلالة. ب

nezakr.natega4dk.net

تدرّب

1 أجب عن الأسئلة التالية.

1. ما المصطلح الخاص باستنتاج خصائص المجتمع من العينة؟
2. ما المصطلح الخاص بطريقة تقدير أن خصائص المجتمع تقع ضمن مدى معين؟
3. ما المصطلح الخاص بالفاصل المستخدم في التقدير الفاصل؟
4. ما المصطلح الخاص بطريقة اختبار فرضيات نموذجية تفحص ما إذا كانت هناك علاقة بين متغيرين فئويين؟

2 أجب عن الأسئلة التالية.

(1) من بين الخيارات التالية (أ - د)، اختر الخيار الذي يصف بشكل أنسب إجراء اختبار الفرضيات.

أ ضع فرضية العدم والفرضية البديلة ← حدد مستوى الدلالة ← احسب القيمة الاحتمالية ← اتخذ حكمًا

ب احسب القيمة الاحتمالية ← ضع فرضية العدم ← حدد مستوى الدلالة ← اتخذ حكمًا

جـ حدد مستوى الدلالة ← تبنَّ الفرضية البديلة ← احسب القيمة الاحتمالية ← ضع فرضية العدم

د اتخذ حكمًا ← ضع فرضية العدم ← حدد مستوى الدلالة ← احسب القيمة الاحتمالية

(2) لنفترض أنه في اختبار فرضيات تم فيه تحديد مستوى الدلالة عند 0.05 (5%)، كانت القيمة الاحتمالية 0.03. من بين الخيارات التالية (أ - د)، اختر الخيار الذي يصف بشكل أنسب الحكم في هذه الحالة.

أ رفض فرضية العدم وتبني الفرضية البديلة.

ب عدم رفض فرضية العدم.

جـ رفض الفرضية البديلة.

د تغيير مستوى الدلالة إلى 0.01 وإعادة الحساب.

(3) لكل من غرضي التحليل (1، 2)، اختر طريقة الاختبار الأنسب من الخيارات أدناه (أ، ب).

1 تريد فحص ما إذا كان متوسط درجات اختبار الطلاب يختلف بين طريقتي تدريس.

2 تريد فحص ما إذا كانت هناك علاقة بين الجنس وفئة المنتجات المشتراة.

[الخيارات] أ اختبار مربع كاي ب اختبار t

(4) من بين الخيارات التالية (أ - د)، اختر الخيار الذي يصف بشكل أنسب القيمة الاحتمالية.

أ القيمة الاحتمالية هي احتمال صحة فرضية العدم.

ب كلما صغرت القيمة الاحتمالية، كان التأثير أكبر.

جـ القيمة الاحتمالية هي احتمال الحصول على نتيجة تساوي في تطرفها البيانات المتاحة أو أكثر تطرفًا منها، بافتراض صحة فرضية العدم.

د القيمة الاحتمالية < 0.05 معيار مطلق للصحة.

📝 سؤال على نمط الامتحان

اشرح سبب اختبار الباحثين فرضية العدم التي يأملون في دحضها، بدلًا من إثبات التأثير مباشرة. **[6]**

استعن بـ: ما تُظهره قيمة احتمالية صغيرة.

💡 توقّف وفكّر

قيمة احتمالية 0.03 مقابل حد 0.05 – أقل أم أعلى؟ إذن ماذا تفعل؟

nezakr.natega4dk.net

⚙️ فكّر كمهندس: ابحث ثم قرر

يدّعي نادٍ مدرسي أن أداة اختباراته الإلكترونية الجديدة ترفع متوسط درجات الطلاب في الاختبارات الأسبوعية. طُلب منك الحكم على هذا الادعاء بإنصاف باستخدام اختبار فرضيات.

1 · صِغ الفرضيات. يفيد النادي بأن متوسط مجموعة أداته كان 74 في آخر اختبار مقابل 70 لمجموعة ضابطة من نفس الحجم. اكتب فرضية العدم والفرضية البديلة لهذا الادعاء بالذات بكلماتك الخاصة، واذكر ما البيانات التي ستحتاجها فعلًا لإجراء الاختبار (بخلاف هذين المتوسطين)، وحدد مستوى الدلالة الذي ستختاره (0.05 أو 0.01) ولماذا.

2 · حلل المخاطر. صف، في هذا السياق، ما الذي سيعنيه خطأ من النوع الأول وخطأ من النوع الثاني بالنسبة للنادي، وحدد أيهما تودّ تجنبه أكثر هنا – وهو حكم لا تقدمه لك جداول الدرس.

3 · قرر وفسّر. لنفترض أن الاختبار أعطى قيمة احتمالية قدرها 0.04. اذكر قرارك بناءً على مستوى الدلالة الذي اخترته في الخطوة 1 (لاحظ أنه سيختلف عند 0.05 مقابل 0.01). ثم، إذا كانت النتيجة ذات دلالة، اشرح لماذا قد لا تبرر النتيجة ذات الدلالة تحويل جميع الفصول إلى استخدام الأداة. **في أزواج،** قارنوا التفسيرات واتفقوا على الأدلة الإضافية التي ما زلتم بحاجة إليها بخلاف القيمة الاحتمالية.

☐ **الوثوق بملخص النادي الخاص** ☐ **صياغة الادعاء واختباره بشكل صحيح**

قدّم دليلًا على إجابتك.

هل تواجه صعوبة؟ حدّد H_0 وH_1 وα قبل رؤية p، واختر الاختبار المناسب وافحص افتراضاته. ثم ناقش تكلفة خطأ النوع الأول والثاني في السياق، ولا تكتفِ بقيمة p: راجع حجم الأثر وفترة الثقة وتصميم الدراسة وآثار القرار.

🌍 طبّق ما تعلمته

يختبر متجر لونين لزر الدفع باستخدام بيانات من 200,000 زائر ويجد أن معدل النقر للون الجديد أعلى بمقدار 0.05 نقطة مئوية، بقيمة احتمالية 0.01 عند مستوى دلالة 5%. حدد ما إذا كانت النتيجة ذات دلالة إحصائية، واشرح لماذا قد يقرر المتجر مع ذلك عدم تغيير الزر.

❓ إجابة السؤال الرئيس

أجاب الدرس عن السؤال الرئيس: كيف نستدل من عينة مع قياس عدم اليقين؟ يستخدم التقدير بيانات العينة لتقدير خصائص المجتمع بقيمة واحدة أو بفترة ثقة، ويصف مستوى الثقة أداء طريقة بناء الفترات عند تكرارها. ويبدأ اختبار الفرضيات بـH_0 وH_1 ومستوى α محدد مسبقًا واختبار مناسب. فإذا كانت $p \le \alpha$ نرفض H_0، وإذا كانت $p > \alpha$ لا نرفضها. ولا تمثل p احتمال صحة H_0، ولا تعني الدلالة الإحصائية أهمية عملية. لذلك نراجع أيضًا حجم الأثر وفترة الثقة والتصميم وتكلفة القرار.

مصطلحات أساسية

مستوى الدلالة – القيمة الحدية لرفض H0.

ذو دلالة إحصائية – القيمة الاحتمالية أقل من القيمة الحدية.

مهم عمليًا – كبير بما يكفي ليكون له أثر في الواقع.

nezakr.natega4dk.net

تمارين

1 اختبر فهمك: أجب عن الأسئلة الآتية دون الرجوع إلى قسم «ملحوظة مهمة»، ثم راجع إجاباتك.

2 اقرأ الفقرة التالية وأجب عن كل سؤال.

في الاستدلال الإحصائي نستخدم العينة لتقدير خصائص المجتمع أو اختبار ادعاءات عنه. يسمى التقدير بقيمة واحدة (1)، والتقدير بنطاق (2). ويسمى تقييم الدليل ضد فرضية مرجعية (3). نحدد (4) و(5) قبل النتيجة، ونحسب (6) باختبار مناسب، ثم نقارنها بـ(7) المحدد مسبقًا.

1. املأ الفراغات (1) - (7).

3 أجب عن الأسئلة التالية.

(1) من بين الخيارات التالية (أ - د)، اختر الخيار الذي يصف بشكل أنسب فترات الثقة ومستويات الثقة.

أ فترة ثقة بمستوى ثقة 95% تعني أن القيمة الحقيقية للمجتمع تقع ضمن تلك الفترة باحتمال 95%.

ب كلما ارتفع مستوى الثقة، ضاقت فترة الثقة.

ج فترة ثقة بمستوى ثقة 95% تعني أنه، لو تكررت طريقة المسح نفسها، فإن نحو 95% من فترات الثقة الناتجة ستحتوي على القيمة الحقيقية للمجتمع.

د مستوى الثقة قيمة تُستخدم كمعيار للحكم في اختبار الفرضيات.

(2) من بين الخيارات التالية (أ - د)، اختر الخيار الذي لا يُعد وصفًا مناسبًا للأخطاء في اختبار الفرضيات.

أ خطأ النوع الأول هو رفض فرضية العدم بينما تكون في الواقع صحيحة.

ب خطأ النوع الثاني هو عدم رفض فرضية العدم بينما تكون في الواقع خاطئة.

ج خفض مستوى الدلالة يمكن أن يقلل من خطأ النوع الأول وخطأ النوع الثاني في آن واحد.

د لخطأي النوع الأول والثاني علاقة مفاضلة.

(3) من بين الخيارات التالية (أ - د)، اختر الخيار الذي يصف بشكل أنسب تفسير الاستدلال الإحصائي.

أ الفرق ذو الدلالة الإحصائية هو بالضرورة فرق مهم عمليًا.

ب مع كمية كبيرة من البيانات، حتى الفرق الضئيل يمكن أن يصبح ذا دلالة إحصائية.

ج إذا انخفضت القيمة الاحتمالية عن 0.05، فإن الاستنتاج صحيح بالتأكيد.

د يمكن للاستدلال الإحصائي أن يزيل عدم اليقين تمامًا.

⭐ الخلاصة

تذكّر:
قدّر بقيمة نقطية أو بفاصل (مستوى الثقة = عدد مرات احتواء الطريقة للحقيقة)؛ واختبر بمقارنة قيمة احتمالية بمستوى دلالة لرفض فرضية العدم أو الإبقاء عليها. راقب أخطاء النوع الأول / الثاني، ولا تخلط أبدًا بين الدلالة الإحصائية والأهمية العملية.

⭐ فكّر وتحدَّ نفسك

تأمل: لماذا يكون مستوى ثقة 95% بيانًا عن الطريقة، لا عن فترة بعينها؟

تحدٍّ: تُبلِّغ دراسة على 500,000 مستخدم عن تأثير "ذي دلالة عالية" (القيمة الاحتمالية < 0.001) بحجم 0.02 نقطة مئوية. ناقش ما إذا كنت ستتصرف بناءً عليه، واذكر المعلومة الإضافية الوحيدة التي قد تغيّر رأيك.

nezakr.natega4dk.net



الدرس 6-2

استخدام وتقييم تحليل الانحدار

أهداف التعلم

- توضيح الفرق بين الانحدار البسيط والانحدار المتعدد، وما يمثله نموذج الانحدار.
- تقييم نموذج الانحدار باستخدام مقاييس R^2 وRMSE وMAE، موضحًا سبب عدم إثبات «الملاءمة الجيدة» لوجود علاقة سببية.

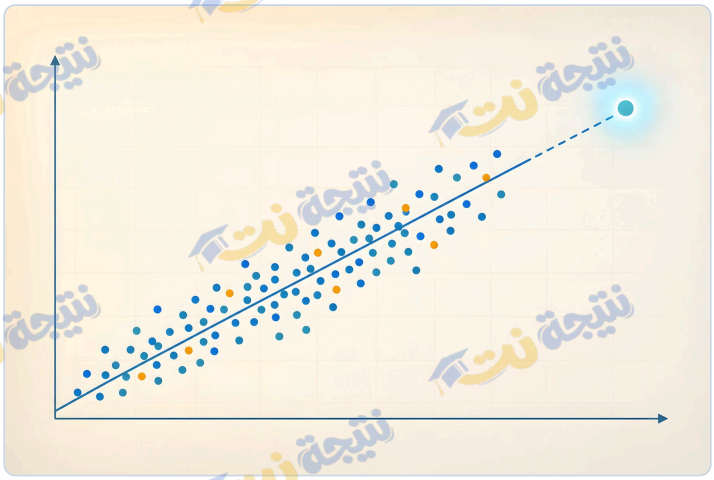
nezakr.natega4dk.net

قد نلاحظ في البيانات علاقة بين درجة الحرارة ومبيعات الآيس كريم. يبني تحليل الانحدار معادلة تصف العلاقة المقدرة بين متغير استجابة ومتغير تفسيري أو أكثر، وقد تستخدم للتفسير أو التنبؤ. لكن المتغير التفسيري لا يصبح سببًا لمجرد دخوله النموذج. ولاستخدام النموذج عمليًا، نقيّم الملاءمة وأخطاء التنبؤ وحدود البيانات بأكثر من مقياس مناسب، مثل R^2 وRMSE وMAE، ونفصل بين جودة التنبؤ والادعاء السببي.

؟ السؤال الرئيس: كيف يحوّل الانحدار البيانات إلى تنبؤ، وكيف نحكم على مدى إمكانية الوثوق بذلك التنبؤ؟

ملحوظة مهمّة

1. مرحلة تحليل الانحدار



ابن نموذجًا يلائم البيانات، ثم قيّمه قبل استخدامه للتنبؤ داخل نطاق مناسب.

(1) **تحليل الانحدار**: طريقة لبناء نموذج رياضي يصف العلاقة المقدرة بين متغير استجابة ومتغير أو أكثر من المتغيرات التفسيرية، وقد يستخدم للتفسير أو التنبؤ إذا كان النموذج مناسبًا وصالحًا للغرض.

(2) **متغير الاستجابة (Response Variable)**: المتغير الذي يحاول النموذج تفسيره أو التنبؤ بقيمته. و**المتغير التفسيري (Explanatory Variable)**: متغير نستخدمه للمساعدة على التفسير أو التنبؤ، ولا يمثل سببًا مثبتًا بمجرد دخوله النموذج.

(3) تُسمى المعادلة التي يتم الحصول عليها من البيانات في تحليل الانحدار **(النموذج)**. النموذج هو تمثيل مبسط لعلاقة حقيقية – على سبيل المثال، وضع الاتجاه "كلما ارتفعت درجة الحرارة، ارتفعت المبيعات" في صيغة يمكن حسابها.

(4) لاستخدام النموذج عمليًا نحتاج إلى تقييم ملاءمته، وخطأ التنبؤ، وافتراضاته، وأدائه على بيانات مناسبة، وحدود النطاق الذي بُني منه.

⭐ خريطة الدرس

الفكرة الأساسية:

يُعبِّر تحليل الانحدار عن العلاقة بين متغير تفسيري (مدخل) ومتغير استجابة (نتيجة) كنموذج – خط $y = ax + b$ في حالة الانحدار البسيط، أو مجموع متغيرات موزونة في حالة الانحدار المتعدد. احكم على النموذج باستخدام R^2 وRMSE وMAE، ولا تخلط أبدًا بين الملاءمة الجيدة وإثبات السببية.

المفاهيم الأساسية:

- تحليل الانحدار
- متغير الاستجابة / المتغير التفسيري
- الانحدار البسيط / المتعدد
- خط الانحدار، المربعات الصغرى
- R^2، RMSE، MAE
- الارتباط مقابل السببية

مصطلحات أساسية

تحليل الانحدار – التعبير عن علاقة كمعادلة للتنبؤ.

متغير الاستجابة – النتيجة (التابع).

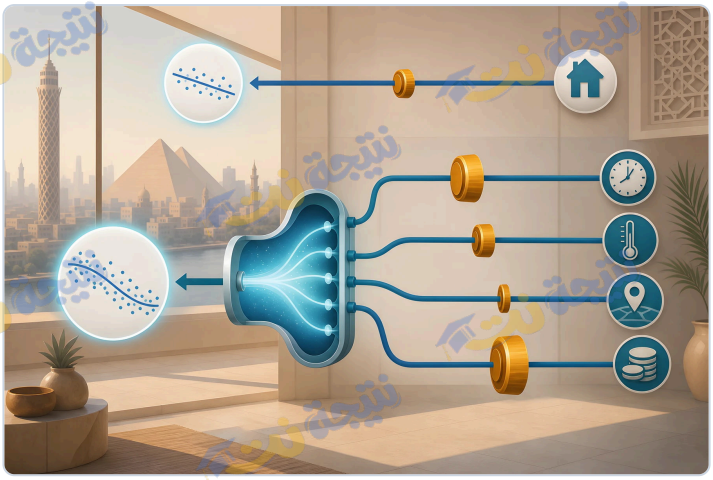
المتغير التفسيري – المتغير المستخدم للتفسير أو التنبؤ، وليس سببًا بالضرورة.

💡 توقّف وفكّر

سبب واحد أم عدة أسباب؟ أي نوع من الانحدار ستحتاجه أسعار المنازل؟

nezakr.natega4dk.net

2. الانحدار البسيط والمتعدد



الانحدار المتعدد يجمع بين عدة مدخلات.

(1) **الانحدار البسيط**: نمذجة متغير الاستجابة باستخدام متغير تفسيري واحد.

على سبيل المثال، يمكن نمذجة مبيعات الآيس كريم y بدرجة الحرارة x وفق خط $y = ax + b$. وتقدّر معاملات الخط غالبًا بطريقة المربعات الصغرى التي تصغر مجموع مربعات المتبقيات؛ و**المتبقّي** هو الفرق بين القيمة الفعلية والقيمة المتوقعة.

(2) **الانحدار المتعدد**: نمذجة متغير الاستجابة باستخدام عدة متغيرات تفسيرية. قد يسمح بإدخال معلومات أكثر، لكنه لا يضمن تنبؤًا أفضل؛ بل يجب تقييم النموذج واختيار المتغيرات المناسبة.

على سبيل المثال، يمكن استخدام حجم المنزل وعمره والمسافة من محطة النقل متغيرات تفسيرية محتملة لسعره. ولكل متغير معامل يصف مقدار تغير التنبؤ المرتبط بزيادة وحدة واحدة منه مع ثبات المتغيرات الأخرى، من دون تفسير سببي تلقائي.

(3) توجد الفروق التالية بين الانحدار البسيط والمتعدد.

الجانب	الانحدار البسيط	الانحدار المتعدد
عدد المتغيرات التفسيرية	واحد	اثنان أو أكثر
خصائص النموذج	بسيط وسهل التفسير	يمكنه التعبير عن علاقات معقدة
الحالات المناسبة	عندما تريد نمذجة متغير استجابة باستخدام متغير تفسيري واحد	عندما تريد إدخال عدة متغيرات تفسيرية في النموذج

مصطلحات أساسية

R^2 – مدى قدرة النموذج على تفسير التباين في البيانات (0-1).

RMSE – حجم الخطأ، يعطي وزنًا أكبر للأخطاء الكبيرة.

MAE – متوسط الخطأ، أقل تأثرًا بالقيم المتطرفة.

3. مؤشرات تقييم النماذج

(1) للحكم على مدى فائدة نموذج تم إنشاؤه بتحليل الانحدار، تُستخدم **(مؤشرات التقييم)**. فيما يلي ثلاثة مؤشرات نموذجية.

nezakr.natega4dk.net

المؤشر	الوصف (مفهوم الحساب)	التفسير
R^2 (معامل التحديد)	يوضح مقدار التباين الذي يفسره النموذج مقارنة بنموذج مرجعي؛ وفي السياق التعليمي المبسط يعرض عادة بين 0 و1	كلما اقتربت القيمة من 1 كان تفسير النموذج للتباين أفضل؛ لكنها لا تكفي وحدها للحكم على النموذج أو أدائه على بيانات جديدة
RMSE (الجذر التربيعي لمتوسط مربع الخطأ)	الجذر التربيعي لمتوسط مربعات الفروق بين القيم المتوقعة والفعلية	كلما صغرت القيمة زادت دقة التنبؤ. وتُقاس بوحدات متغير الاستجابة نفسها؛ وهي تتأثر أكثر بالأخطاء الكبيرة
MAE (متوسط الخطأ المطلق)	متوسط القيم المطلقة للفروق بين القيم المتوقعة والفعلية	كلما انخفضت القيمة زادت دقة التنبؤ؛ وهي أقل تأثرًا بالفروق الكبيرة للغاية من RMSE

(2) يشير كل من RMSE وMAE إلى حجم أخطاء التنبؤ، لكن RMSE يميل إلى إعطاء وزن أكبر للفروق الكبيرة. أما MAE فيقيّم الفروق في المتوسط، لذا فهو أقل تأثرًا بالقيم المتطرفة. تُستخدم الطريقتان بشكل مختلف وفقًا لغرض التحليل.

4. تنبيهات عند استخدام نماذج الانحدار

(1) تحليل الانحدار طريقة واسعة التطبيق، لكن سوء استخدامها قد يؤدي إلى استنتاجات خاطئة. انتبه بشكل خاص للنقاط التالية.

① لا تخلط بين الارتباط والسببية: الانحدار وحده لا يثبت السببية. يحتاج الادعاء السببي إلى تصميم دراسة وأدلة إضافية تسمح بالاستدلال السببي؛ ولا تكفي الملاءمة أو الأسبقية الزمنية وحدهما.

② النموذج لا يكتمل بمجرد إنشائه: تأكد من دقته باستخدام مؤشرات التقييم وتحقق من صلاحيته في ضوء الغرض.

③ لا تعتمد على R^2 وحده: افحص أخطاء التنبؤ والافتراضات والأداء على بيانات لم تستخدم في الملاءمة، ويمكن استخدام R^2 المعدل عند مقارنة نماذج مختلفة في عدد المتغيرات.

④ إضافة المزيد من المتغيرات التفسيرية لا يُحسّن النموذج بالضرورة: يجب اختيار المتغيرات المناسبة وفقًا للغرض.

(2) لا ينبغي التعامل مع نتائج تحليل الانحدار على أنها حقيقة مطلقة كما هي، بل كمنظور واحد تقدمه البيانات. من المهم تفسير النتائج بشكل مناسب واستخدامها في الحكم على أرض الواقع.

الفكرة الرئيسة

يحوّل الانحدار علاقة إلى معادلة قابلة للاستخدام: متغير استجابة يُتنبَّأ به من متغير تفسيري واحد (بسيط، $y = ax + b$) أو من عدة متغيرات (متعدد). تُوجد الملاءمة بطريقة المربعات الصغرى، وتُقيَّم الجودة باستخدام R^2 وRMSE وMAE معًا – وليس R^2 وحده أبدًا. والأهم من ذلك، أن الملاءمة الجيدة ليست دليلًا على السببية: الارتباط ليس سببية.

✎ مثال محلول

(1) ضع علامة أمام كل عبارة (أ - د) بـ○ إذا كانت صحيحة أو × إذا كانت خاطئة.

أ تحليل الانحدار المتعدد طريقة للتنبؤ بمتغير الاستجابة باستخدام متغير تفسيري واحد.

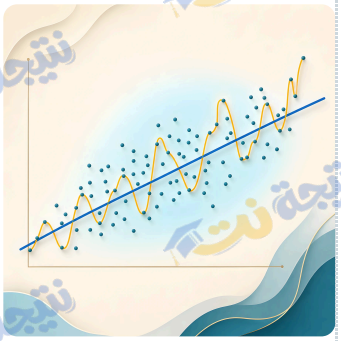
ب R^2 (معامل التحديد) مؤشر يُظهر مدى قدرة النموذج على تفسير التباين في البيانات.

جـ تُسمى المعادلة التي يتم الحصول عليها من البيانات في تحليل الانحدار النموذج، وهي تمثيل مبسط لعلاقة حقيقية.

د إذا لاءمت معادلة التنبؤ البيانات، يمكن القول بوجود علاقة سببية بين المتغير التفسيري ومتغير الاستجابة.

توقّف وفكّر

خط تنبؤ يلائم البيانات جيدًا – هل يثبت ذلك أن أحد المتغيرين يسبب الآخر؟



النموذج المفرط في الملاءمة يطارد التشويش، لا الاتجاه العام.

توقّف وفكّر

أي مقياس خطأ ستثق به أكثر عند وجود عدد قليل من القيم المتطرفة الكبيرة جدًا؟

nezakr.natega4dk.net

(2) طابق كل وصف (1 - 3) مع المؤشر الأنسب من الخيارات أدناه (أ - ج).

1 متوسط القيم المطلقة للفروق بين القيم المتوقعة والفعلية

2 الجذر التربيعي لمتوسط مربعات الفروق بين القيم المتوقعة والفعلية

3 مدى قدرة النموذج على تفسير التباين في البيانات

[الخيارات] أ RMSE ب R^2 ج MAE

✓ الحلّ

(1)

أ التنبؤ بمتغير تفسيري واحد هو تحليل الانحدار البسيط. تحليل الانحدار المتعدد يستخدم عدة متغيرات تفسيرية. لذلك، x.

ب وصف صحيح لـ R^2. لذلك، ○.

ج تعريف صحيح للنموذج. لذلك، ○.

د حتى لو لاءمت معادلة التنبؤ البيانات، فإن ذلك وحده لا يثبت السببية. لذلك، x.

(2)

1: متوسط الفروق المطلقة هو MAE. **ج**

2: الجذر التربيعي لمتوسط مربعات الفروق هو RMSE. **أ**

3: المؤشر الذي يفسر التباين في البيانات هو R^2. **ب**

تدرّب

1 أجب عن الأسئلة التالية.

1. ما المصطلح الخاص بتحليل الانحدار الذي يتنبأ بمتغير الاستجابة باستخدام عدة متغيرات تفسيرية؟
2. ما المصطلح الخاص بالمؤشر الذي يُظهر مدى قدرة النموذج على تفسير التباين في البيانات؟
3. في تحليل الانحدار، ماذا يسمى المتغير الذي نريد تفسيره، والمتغير المستخدم للمساعدة على تفسيره؟
4. ما المصطلح الخاص بـ"الفرق بين القيم الفعلية وخط الانحدار" المستخدم في تحليل الانحدار البسيط؟

2 أجب عن الأسئلة التالية.

(1) من بين الخيارات التالية (أ - د)، اختر الخيار الذي يصف بشكل أنسب الفرق بين الانحدار البسيط والمتعدد.

أ يمكن للانحدار البسيط أن يعبر عن علاقات معقدة، بينما الانحدار المتعدد بسيط.

ب يستخدم الانحدار البسيط عدة متغيرات تفسيرية، ويستخدم الانحدار المتعدد متغيرًا تفسيريًا واحدًا.

ج يستخدم الانحدار البسيط متغيرًا تفسيريًا واحدًا، ويستخدم الانحدار المتعدد عدة متغيرات تفسيرية.

د لا يوجد فرق بين الانحدار البسيط والمتعدد في عدد المتغيرات التفسيرية؛ فقط طريقة الحساب تختلف.

(2) لكل من غرضي التحليل (1، 2)، اختر طريقة تحليل الانحدار الأنسب من الخيارات أدناه (أ، ب).

1 تريد التنبؤ بمبيعات الآيس كريم من درجة الحرارة (متغير واحد).

2 تريد التنبؤ بأسعار المنازل من حجم المنزل وعمره والمسافة من محطة النقل (متغيرات متعددة).

📝 سؤال على نمط الامتحان

اشرح لماذا يمكن لنموذج انحدار ذي قيمة عالية لـ R^2 أن يظل أساسًا غير كافٍ للادعاء بأن متغيرًا ما يسبب متغيرًا آخر. **[6]**

استعن بمفهوم: الارتباط مقابل السببية.

💡 توقّف وفكّر

متغير واحد أم عدة متغيرات؟ هذا هو الفرق كله.

nezakr.natega4dk.net

[الخيارات] أ تحليل الانحدار البسيط ب تحليل الانحدار المتعدد

(3) من بين الخيارات التالية (أ - د)، اختر الخيار الذي يصف بشكل أنسب تفسير نتائج تحليل الانحدار.

أ إذا تم الحصول على معادلة تنبؤ من تحليل الانحدار، فإن المعادلة تشير بالضرورة إلى علاقة سببية.

ب يجب التعامل مع نتائج تحليل الانحدار كمنظور واحد تقدمه البيانات، ويجب تقييم ما إذا كانت السببية قد أُثبتت بعناية.

ج إذا كان R^2 مرتفعًا، فإن النموذج يكون بالضرورة نموذجًا جيدًا.

د إذا لاءمت معادلة التنبؤ البيانات، يمكن استنتاج أن المتغير التفسيري هو سبب متغير الاستجابة.

(4) من بين الخيارات التالية (أ - د)، اختر الخيار الذي لا يُعد وصفًا مناسبًا لمؤشرات تقييم نماذج الانحدار.

أ يأخذ R^2 (معامل التحديد) قيمًا تُعرض عادةً بين 0 و1 في السياق التعليمي المبسط؛ وكلما اقترب من 1، كان تفسير التباين في البيانات أفضل.

ب RMSE هو الجذر التربيعي لمتوسط مربعات الفروق بين القيم المتوقعة والفعلية، ويُعبَّر عنه بنفس وحدات متغير الاستجابة.

ج MAE هو متوسط القيم المطلقة للفروق بين القيم المتوقعة والفعلية، وهو أقل تأثرًا بالفروق الكبيرة جدًا.

د إذا كانت قيمة R^2 قريبة من 1، فإن النموذج يكون بالضرورة نموذجًا جيدًا دون الحاجة إلى فحص مؤشرات أخرى.

⚙️ فكّر كمهندس: ابحث ثم قرر

تريد مقصف مدرستك التنبؤ بمبيعاته اليومية من المشروبات الباردة حتى يتمكن من طلب الكمية المناسبة. أنت تخطط لبناء نموذج انحدار.

1 · **أنشئ النموذج.** سمِّ متغير الاستجابة، ثم اذكر المتغيرات التفسيرية التي ستجربها (مثل درجة حرارة اليوم، وعدد الطلاب الحاضرين، وما إذا كان يوم امتحان). قرر ما إذا كان هذا انحدارًا بسيطًا أم متعددًا، واذكر السبب.

2 · احكم على النموذج. اختر مؤشرًا رئيسيًا يناسب الهدف، ثم مؤشرًا آخر يكمله، واشرح ما يقيسه كل منهما. واذكر متغيرًا مربكًا قد يجعل متغيرين يتحركان معًا من دون أن يكون أحدهما سببًا للآخر.

3 · قرّر وفسّر. هل يسمح R^2 المرتفع وحده بادعاء السببية؟ وما الأدلة الإضافية المطلوبة؟ قارنوا المتغيرات التفسيرية المقترحة وحددوا أيها يستحق الاحتفاظ به، وهل يكفي متغير واحد أم أن استخدام عدة متغيرات أنسب، مع تبرير القرار.

☐ **الثقة بسبب واحد واضح** ☐ **بناء نموذج واختباره**

قدّم دليلًا على إجابتك.

هل تواجه صعوبة؟ ابدأ من النتيجة التي تريد التنبؤ بها (المبيعات)، ثم اسأل أي الأشياء القابلة للقياس تحركها بشكل معقول – ولكل منها، اسأل "هل يمكن أن يكون هناك شيء آخر يقود كليهما؟" R^2 المرتفع يخبرك أن الخط يلائم البيانات، وليس أن السبب حقيقي.

💡 توقّف وفكّر

R^2 مرتفع وحده – هل يكفي للثقة بنموذج، أم تحتاج إلى RMSE وMAE أيضًا؟

مصطلحات أساسية

المعامل – الوزن الممنوح لمتغير تفسيري.

نقطة التقاطع – الثابت b في المعادلة.

العامل المربك – سبب ثالث خفي وراء ارتباط ما.

nezakr.natega4dk.net

🌍 طبّق ما تعلمته

لاحظت إحدى المدن أن الأشهر التي تشهد مبيعات أعلى للآيس كريم تشهد أيضًا عددًا أكبر من حوادث السباحة، وأن خط الانحدار يظهر ارتباطًا وثيقًا بينهما. اشرح لماذا لا يعني هذا أن الآيس كريم هو سبب الحوادث، واذكر العامل الخفي الذي يُرجَّح أن يكون وراء كليهما.

❓ إجابة السؤال الرئيس

تناول الدرس كيفية تحويل البيانات إلى تنبؤات باستخدام الانحدار، وكيفية الحكم على مدى موثوقية هذه العملية. يُعبِّر تحليل الانحدار عن العلاقة بين المتغير التفسيري (المدخلات، يُستخدم للتفسير أو التنبؤ) ومتغير الاستجابة (النتيجة) في صورة نموذج: إما خط انحدار بالصيغة $y = ax + b$ في حالة الانحدار البسيط (متغير تفسيري واحد)، أو مجموع مرجَّح بالصيغة $y = a_1x_1 + a_2x_2 + b$ في حالة الانحدار المتعدد (عدة متغيرات)، حيث يُضبط الخط بطريقة «المربعات الصغرى» لتقليل المتبقيات (الفروق بين القيم الفعلية والمتوقعة). وتُقيَّم الموثوقية بمؤشرات مجتمعة: معامل التحديد (R^2) لقياس مدى تفسير التباين، ومقياسا RMSE وMAE لقياس حجم أخطاء التنبؤ (مع مراعاة أن RMSE يمنح وزنًا أكبر للأخطاء الكبيرة، بينما لا يتأثر MAE كثيرًا بالقيم الشاذة). ولا ينبغي أبدًا الاعتماد على R^2 وحده، ولا إضافة متغيرات لمجرد الإضافة. وفوق ذلك، فإن «الملاءمة الجيدة» للنموذج تعكس ارتباطًا إحصائيًّا فقط؛ إذ يتطلب إثبات السببية شروطًا أخرى مثل الأسبقية الزمنية، واستبعاد العوامل الأخرى، ووجود إطار نظري داعم. باختصار، يقدّم الانحدار منظورًا واحدًا مستمدًّا من البيانات، يخضع للتفسير والتحليل لا للتسليم المطلق.

⭐ الخلاصة

تذكّر:

يتنبأ الانحدار بمتغير استجابة من متغير تفسيري واحد (بسيط، $y = ax + b$) أو من عدة متغيرات (متعدد)، تُلاءم بطريقة المربعات الصغرى. احكم عليه باستخدام R^2 وRMSE وMAE معًا – ولا تقرأ أبدًا الملاءمة الجيدة على أنها دليل على السببية.

⭐ فكّر وتحدَّ نفسك

تأمل: فكر في شيئين رأيتهما يرتفعان معًا (مثل وقت الشاشة وسوء النوم). ما العامل الخفي الذي قد يقود كليهما؟

تحدٍّ: يمكنك إضافة متغير تفسيري واحد فقط إلى نموذج لديه بالفعل R^2 مرتفع. ناقش ما إذا كنت ستضيفه على الإطلاق، واذكر الفحص الوحيد الذي ستجريه قبل الوثوق بالنموذج الأكبر.

nezakr.natega4dk.net

تمارين

1 اختبر فهمك: أجب عن الأسئلة الآتية دون الرجوع إلى قسم «ملحوظة مهمة»، ثم راجع إجاباتك.

2 اقرأ الفقرة التالية وأجب عن كل سؤال.

في تحليل الانحدار يسمى المتغير الذي نريد تفسيره أو التنبؤ به (1)، ويسمى المتغير المستخدم للمساعدة على ذلك (2). ويستخدم (3) متغيرًا تفسيريًا واحدًا، بينما يستخدم (4) عدة متغيرات. ومن مقاييس التقييم (5) للتباين المفسر في السياق المحدد، و(6) و(7) لأخطاء التنبؤ. ولا يجوز الاستنتاج بوجود (8) لمجرد ملاءمة النموذج للبيانات.

1. املأ الفراغات (1) - (8).

3 أجب عن الأسئلة التالية.

(1) من بين الخيارات التالية (أ - د)، اختر الخيار الذي يصف بشكل أنسب الارتباط والسببية.

أ إذا وُجد ارتباط، فإن السببية تتحقق بالضرورة.

ب لادعاء السببية، يلزم أسبقية زمنية، واستبعاد تأثير عوامل أخرى، ودعم نظري، وما إلى ذلك.

ج إذا تم الحصول على معادلة تنبؤ من تحليل الانحدار، فإن المعادلة تُثبت السببية إثباتًا تامًا.

د إذا لم يكن هناك ارتباط، يمكن القول بعدم وجود أي علاقة على الإطلاق بين المتغيرين.

(2) من بين الخيارات التالية (أ - د)، اختر الخيار الذي يصف بشكل أنسب تفسير مؤشرات التقييم.

أ يأخذ R^2 (معامل التحديد) قيمًا تُعرض عادةً بين 0 و1 في السياق التعليمي المبسط؛ وكلما اقترب من 1، كان تفسير التباين في البيانات أفضل.

ب يستخدم RMSE وMAE طريقة الحساب نفسها، وتتطابق قيمتاهما دائمًا.

ج النموذج ذو R^2 المرتفع هو دائمًا نموذج جيد في أي حالة.

د MAE مؤشر يعطي وزنًا أكبر للفروق الكبيرة.

(3) من بين الخيارات التالية (أ - د)، اختر الخيار الذي لا يُعد نقطة حذر مناسبة عند استخدام نماذج الانحدار.

أ لا تستنتج وجود سببية لمجرد أن معادلة التنبؤ تلائم البيانات.

ب تأكد من الدقة باستخدام مؤشرات التقييم وتحقق من صلاحية النموذج في ضوء الغرض.

ج إذا كان R^2 قريبًا من 1، يمكن الحكم على جودة النموذج دون فحص مؤشرات أخرى.

د إضافة المزيد من المتغيرات التفسيرية لا يُحسّن الدقة بالضرورة. يجب اختيار المتغيرات المناسبة وفقًا للغرض.

nezakr.natega4dk.net



الدرس 6-3

تصور البيانات والتواصل

أهداف التعلم

- شرح دور تصور البيانات واختيار الرسوم البيانية المناسبة، بما في ذلك المخصصة للبيانات متعددة المتغيرات.
- توظيف أسلوب سرد القصص المستندة إلى البيانات لنقل رسالة واضحة وصادقة من واقع البيانات.

nezakr.natega4dk.net

قد يخفي جدول الأرقام الشيء ذاته الذي أنت في أمسّ الحاجة إلى رؤيته؛ فإذا حُوِّل إلى الصورة المناسبة برزت الأنماط بوضوح – كالساعات الأكثر ازدحامًا، أو أقوى الروابط، أو القيم الشاذة اللافتة. ويؤدي تصور البيانات وظيفتين: يساعدك على فهم بياناتك، ويساعدك على شرحها للآخرين. فمع متغير واحد أو متغيرين تكفي الأعمدة أو الخطوط أو المخططات المبعثرة، وعند عرض ثلاثة متغيرات أو أكثر معًا تُستخدم الخرائط الحرارية ومخططات المستطيلات – على أن يظل اختيار الرسم محكومًا بنوع البيانات والسؤال وطبيعة المقارنة، لا بعدد المتغيرات وحده. والرسم البياني وحده ليس رسالة متكاملة؛ فسرد قصة واضحة للبيانات هو ما يجعل الجمهور يستوعبها حقًّا، مع بقاء العرض أمينًا للسياق والمصدر.

؟ السؤال الرئيس: كيف نختار تصورًا يكشف نمطًا بصدق، وكيف نحوّل رسمًا بيانيًا إلى رسالة يفهمها الآخرون؟

📌 ملحوظة مهمّة

1. دور تصور البيانات



الرسم الجيد يحتاج إلى تصميم واضح وتسميات وسياق مناسب؛ والازدحام قد يخفي النمط.

(1) **(تصور البيانات)**... تمثيل البيانات كرسوم بيانية أو أشكال بحيث يمكن إدراك السمات التي يصعب قراءتها من الأرقام وحدها بصريًا.

(2) يؤدي تصور البيانات دورين بشكل عام.

الدور	الوصف	مثال
تعميق فهم المحلل نفسه	إدراك توزيع البيانات واتجاهاتها والقيم الشاذة وغيرها بصريًا واكتساب رؤى	عند عرض سجلات التدريب أو وقت الدراسة كرسم بياني خطي، تصبح التغيرات والاتجاهات التي يصعب ملاحظتها عندما تكون الأرقام مدرجة فقط في جدول أسهل للرؤية
توصيل البيانات للآخرين	عرض النتائج والاستنتاجات مع ما يدعمها، بطريقة واضحة للجمهور	عند عرض متوسط درجة الاختبار لكل فصل كرسم بياني بالأعمدة، يستطيع حتى من يرى البيانات لأول مرة مقارنة الفروق بين الفصول من نظرة واحدة

(3) يعتمد اختيار التصور على نوع البيانات والسؤال والجمهور وعدد المتغيرات وطريقة المقارنة. ويجب أن يتجنب التصميم المقاييس والاختيارات التي تمنح انطباعًا لا تدعمه البيانات.

⭐ خريطة الدرس

الفكرة الأساسية:

يحوّل التصور الأرقام إلى أنماط تستطيع العين إدراكها – لمساعدة المحلل على التفكير وللتواصل مع الآخرين على حد سواء. تناسب الرسوم البيانية الشائعة متغيرًا واحدًا أو متغيرين: أما الخرائط الحرارية ومخططات المستطيلات (الفسيفساء) فتتعامل مع أكثر من ذلك. يعتمد التواصل الجيد على سرد قصص البيانات: رسالة واضحة، وعناصر تخدم هذه الرسالة فقط، وتسميات توضيحية صادقة وداعمة.

المفاهيم الأساسية:

- تصور البيانات
- الخريطة الحرارية
- مخطط المستطيلات
- التراكب الزائد للنقاط
- سرد قصص البيانات

مصطلحات أساسية

تصور البيانات – عرض البيانات كرسوم بيانية/أشكال لإدراك سماتها بصريًا.

nezakr.natega4dk.net

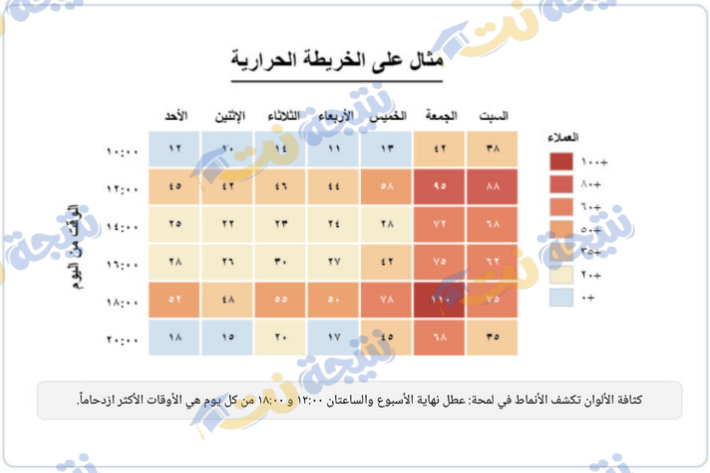
2. التصور الذي يتعامل مع متغيرات متعددة



يمكن لعرض واحد أن يجسّد عدة متغيرات في آنٍ واحد.

لا توجد قاعدة تربط نوع الرسم بعدد المتغيرات وحده. فالأعمدة تقارن فئات، والخطوط تعرض تغيرًا عبر زمن أو ترتيب، والمخطط المبعثر يعرض علاقة متغيرين كميين. أما الخريطة الحرارية ومخطط المستطيلات فلكل منهما استخدام محدد كما يأتي.

(1) **الخريطة الحرارية**: شبكة أو مصفوفة يستخدم فيها اللون لتمثيل قيمة عند كل خلية، مع مفتاح يوضح معنى الألوان. مثال: متوسط الحرارة لكل مدينة وشهر إذا كان الهدف مقارنة نمط القيم عبر الشبكة.



شكل 6.3.1 – يمثل اللون قيمة الخلية وفق المقياس الموضح في المفتاح

المخطط المبعثر مناسب أساسًا لعلاقة متغيرين كميين. وإذا تداخلت نقاط كثيرة، يمكن استخدام الشفافية أو تجميع النقاط في خلايا أو تمثيل كثافة ثنائية الأبعاد بحسب الحالة. أما الخريطة الحرارية العامة فتناسب وجود قيمة لكل خلية أو منطقة، وليست حلًا تلقائيًا لكل تراكب.

(2) **مخطط المستطيلات**: يعرض تكرارات أو نسب توليفات متغيرين فئويين باستخدام مساحات المستطيلات. وقد تكون الأعمدة المكدسة أو النسبية أوضح أحيانًا بحسب السؤال والجمهور.

مصطلحات أساسية

الخريطة الحرارية – شبكة ملونة على محورين لقيمة ثالثة.

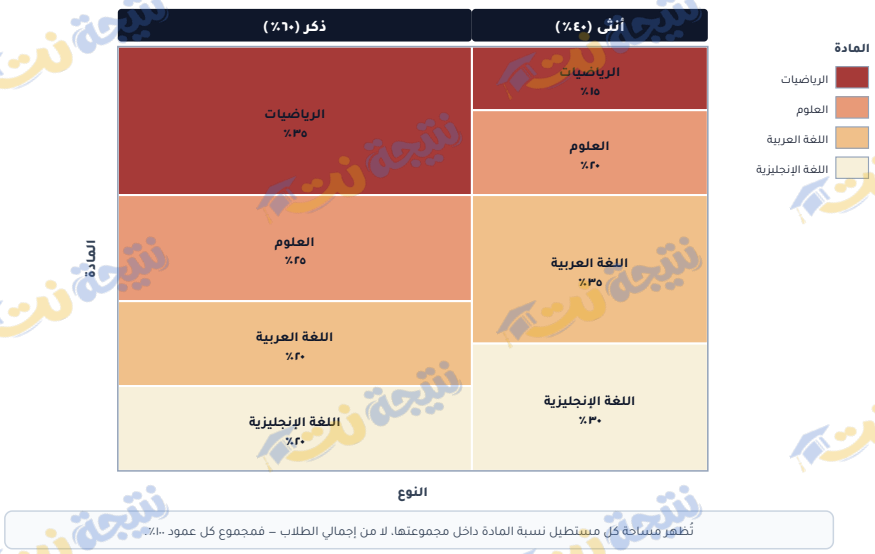
التراكب الزائد للنقاط – تداخل النقاط بحيث تختفي الكثافة.

توقّف وفكّر

فئتان تُقارَنان بالمساحة، أم قيمة ثالثة باللون – أي رسم بياني هو أيهما؟

nezakr.natega4dk.net

مثال على مخطط المستطيلات (الفسيفساء)



شكل 6.3.2 – مخطط مستطيلات: تُظهر مساحة كل مستطيل نسبة المادة داخل كل مجموعة

تمثل مساحة كل مستطيل تكرار أو نسبة إحدى توليفات الفئات بحسب إعداد الرسم. ويجب توضيح هل يعرض الرسم أعدادًا أم نسبًا وكيف حُسبت.

3. سرد قصص البيانات

(1) **سرد القصص بالبيانات**: تنظيم عرض البيانات حول سؤال أو نتيجة واضحة، مع إبقاء السياق والمعلومات التي يحتاجها القارئ لفهم البيانات وعدم إخفاء ما قد يغير التفسير.

(2) فيما يلي العناصر الأساسية الثلاثة لسرد القصص.

العنصر	الوصف	مثال
توضيح الرسالة	جعل من الممكن التعبير في جملة واحدة عما تريد نقله من خلال الرسم البياني	اجعل عنوان رسم بياني يعرض نتائج اختبار الفصل يحتوي على الادعاء، مثل "حقق الفصل درجات عالية بشكل عام في اختبار الرياضيات"
اختيار العناصر	إزالة العناصر غير المرتبطة بالرسالة والتركيز على الأجزاء التي تريد لفت الانتباه إليها	اجعل العناصر التي تريد لفت الانتباه إليها وحدها بارزة بالألوان، مع إبقاء بقية العناصر بألوان باهتة تظل واضحة، ودون إخفاء السياق أو الأدلة المخالفة
توفير معلومات داعمة	كتابة عنوان واضح، وتسميات المحاور والوحدات، والمصدر والفترة وتعريف المقياس أو الملاحظة المنهجية عند الحاجة	اجعل عنوان الرسم البياني، مثل "متوسط درجات اختبار الرياضيات حسب الفصل"، يعبّر عن المحتوى من نظرة واحدة، وحدّد بوضوح الوحدات وأسماء العناصر مثل "درجة" على المحور الرأسي و"الفصل" على المحور الأفقي

(3) يمكن أن يتغير عمق فهم الجمهور بشكل كبير بحسب طريقة نقل نفس نتيجة التحليل. بعد تصور البيانات، من المهم أن تضع نفسك في مكان الجمهور وتتحقق مما إذا كانت الرسالة قد وصلت.

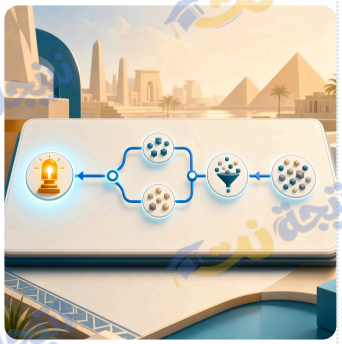
الفكرة الرئيسة

تُظهر الصور أنماطًا لا تُظهرها الأرقام الخام – للمحلل وللجمهور على حد سواء. طابِق الرسم البياني مع البيانات: الأعمدة والخطوط والمخططات المبعثرة لمتغير واحد أو متغيرين؛ الخرائط الحرارية لقيمة عبر محورين (مع تجنب التراكب الزائد)؛ مخططات المستطيلات لمتغيرين فئويين بالمساحة. ثم ارْوِ قصة بيانات: رسالة واضحة واحدة، وعناصر تخدمها فقط، وتسميات صادقة – وتحقق من أن الجمهور يستقبلها فعليًا.

مصطلحات أساسية

مخطط المستطيلات (الفسيفساء) – مساحات مستطيلة لمتغيرين فئويين.

سرد قصص البيانات – هيكلة رسالة البيانات كقصة.



السرد البصري يوجّه العين نحو رؤية واحدة محددة.

nezakr.natega4dk.net

✎ مثال محلول

(1) ضع علامة أمام كل عبارة (أ - د) بعلامة ○ إذا كانت صحيحة أو × إذا كانت خاطئة.

أ يؤدي تصور البيانات دور تعميق فهم المحلل نفسه ودور توصيل البيانات للآخرين معًا.

ب الخريطة الحرارية رسم بياني يمثل حجم قيمة ثالثة مقابلة للتوليفات على محورين بتدرج اللون أو اختلافات الصبغة.

جـ مخطط المستطيلات رسم بياني يمثل العلاقة بين متغيرين فئويين بحجم المساحات المستطيلة.

د في سرد قصص البيانات، من المهم تضمين أكبر عدد ممكن من العناصر المرتبطة بالرسم البياني.

(2) طابق كل وصف (1 - 3) مع أنسب رسم بياني من الخيارات أدناه (أ - جـ).

1 رسم بياني يمثل حجم قيمة عددية مقابلة للتوليفات على محورين بتدرج اللون أو اختلافات الصبغة

2 رسم بياني يمثل العلاقة بين متغيرين فئويين بحجم المساحات المستطيلة

3 رسم بياني يمثل العلاقة بين متغيرين كميين بموضع النقاط

[الخيارات] أ مخطط مبعثر ب خريطة حرارية جـ مخطط مستطيلات

✓ الحلّ

(1)

أ صحيح كدور من أدوار تصور البيانات. لذلك، ○.

ب تعريف صحيح للخريطة الحرارية. لذلك، ○.

جـ تعريف صحيح لمخطط المستطيلات. لذلك، ○.

د في سرد قصص البيانات، من المهم إزالة العناصر غير المرتبطة بالرسالة. لذلك، ×.

(2)

1: عرض القيم بتدرج اللون أو الصبغة هو الخريطة الحرارية. **ب**

2: عرض العلاقات بالمساحة هو مخطط المستطيلات. **جـ**

3: عرض العلاقات بالنقاط هو المخطط المبعثر. **أ**

💡 توقّف وفكّر

هل إضافة المزيد من العناصر إلى رسم بياني يجعل الرسالة أوضح، أم يخفيها؟

📝 سؤال على نمط الامتحان

اشرح لماذا يمكن أن يؤدي اختيار الرسم البياني الخطأ، أو إخفاء تسميات الرسم البياني، إلى تضليل الجمهور حتى عندما تكون البيانات صحيحة. **[6]**

راجع: مطابقة الرسم البياني مع البيانات والمعلومات الداعمة الصادقة.

تدرّب

1 أجب عن الأسئلة التالية.

1. ما هو مصطلح الرسم البياني الذي يمثل حجم قيمة ثالثة مقابلة للتوليفات على محورين بتدرج اللون أو اختلافات الصبغة؟
2. ما هو مصطلح الرسم البياني الذي يمثل العلاقة بين متغيرين فئويين بحجم المساحات المستطيلة؟
3. في تصور البيانات، من بين "تعميق فهم المحلل نفسه" و"توصيل البيانات للآخرين"، لأي غرض يُؤدَّى الأخير؟
4. ما هو مصطلح استخدام البيانات والتصور لنقل الرسالة التي تريد إيصالها، منظمة كقصة؟

nezakr.natega4dk.net

2 أجب عن الأسئلة التالية.

(1) لكل من الحالتين (1، 2)، اختر أنسب رسم بياني من الخيارات أدناه (أ، ب).

1 تريد إدراك نمط زيارات العملاء حسب يوم الأسبوع × الفترة الزمنية من نظرة واحدة.

2 تريد أن تَظهر بشكل حدسي الفروق في النسب بين الذكور والإناث × تفضيل المادة الدراسية باستخدام المساحات.

[الخيارات] أ خريطة حرارية ب مخطط مستطيلات

(2) من الخيارات التالية (أ - د)، اختر الذي يصف بأنسب طريقة موقفًا تُستخدم فيه الخريطة الحرارية.

أ موقف يريد فيه المرء ببساطة مقارنة قيمة تمثيلية لمتغير واحد.

ب موقف يمكن فيه إدراك العلاقة بين متغيرين كميين بشكل كافٍ باستخدام مخطط مبعثر لأن عدد النقاط قليل.

جـ موقف يريد فيه المرء إدراك توزيع البيانات العددية وأنماطها عبر توليفات على محورين من نظرة واحدة.

د موقف يريد فيه المرء إظهار نسبة تركيب واحدة فقط لمتغير فئوي.

(3) من الخيارات التالية (أ - د)، اختر الذي يصف بأنسب طريقة موقفًا يُستخدم فيه مخطط المستطيلات.

أ موقف يريد فيه المرء متابعة التغيرات الزمنية لمتغير كمي واحد.

ب موقف يريد فيه المرء إدراك نسب التكرارات لتوليفات متغيرين فئويين بشكل حدسي باستخدام المساحات.

جـ موقف يريد فيه المرء تمثيل الارتباط بين متغيرين كميين بالنقاط.

د موقف يريد فيه المرء إظهار نسبة تركيب الفئات في وقت معين باستخدام قطاعات دائرية.

(4) من الخيارات التالية (أ - د)، اختر الذي يصف بأنسب طريقة سرد قصص البيانات.

أ حتى لو كانت الرسالة التي تريد إيصالها غير واضحة، فستصل إلى الجمهور طالما أن الرسم البياني جميل.

ب توضيح الرسالة، واختيار العناصر، وتوفير معلومات داعمة هي العناصر الأساسية.

جـ من الأفضل تضمين أكبر عدد ممكن من العناصر ذات الصلة في الرسم البياني.

د العناوين وتسميات المحاور تنتقص من مظهر الرسم البياني ويُفضَّل حذفها.

توقّف وفكّر

محوران بقيمة باللون، أم فئتان بالمساحة؟ سمِّ الرسم البياني لكل موقف.

توقّف وفكّر

سلسلة زمنية لقيمة واحدة، أم نسب فئتين – أيهما يناسب مخطط المستطيلات؟

nezakr.natega4dk.net

مصطلحات أساسية

الجدولة المتقاطعة – جدول لعدد الحالات لتوليفات الفئات.

الرسم البياني المُضلّل – رسم بياني يشوّه تصميمه النمط الحقيقي.

⚙️ فكّر كمهندس: ابحث ثم قرر

جمع فصلك بيانات عن محافظة سكن الطلاب ورياضتهم المفضلة، وكذلك هطول الأمطار الشهري لعدة مدن مصرية على مدار العام. عليك عرض كل مجموعة بيانات بوضوح.

1 · **اختر رسمًا بيانيًا لكل حالة.** بالنسبة لبيانات المحافظة × الرياضة المفضلة، وبيانات المدينة × هطول الأمطار الشهري، سمِّ الرسم البياني الذي ستستخدمه. برر اختيارك لواحدة منهما على الأقل – هل يمكن أن يصلح النوع الآخر من الرسم البياني هنا أيضًا، ولماذا يعتبر اختيارك أفضل؟

2 · **احترس من التضليل.** سمِّ خيار تصميم واحدًا قد يجعل الرسم البياني الذي اخترته يُضلل القارئ (على سبيل المثال مقياس لوني بدون مفتاح، أو فئات مبتورة)، وقل كيف ستمنع ذلك – وهو حكم لا تمنحك أمثلة الدرس إياه.

3 · نظّم العرض. اكتب النتيجة التي تريد شرحها في جملة، وحدد ما ستبرزه ولماذا لا يخفي ذلك بقية السياق. ثم أضف العنوان والمحاور والوحدات والمصدر والملاحظات اللازمة لتفسير الرسم.

☐ **استخدم رسمًا بيانيًا بالأعمدة لكل شيء**

☐ **طابق كل مجموعة بيانات مع الرسم البياني المناسب**

قدِّم دليلًا على إجابتك.

هل تواجه صعوبة؟ إذا كانت لديك تكرارات مشتركة لفئتين فقد يناسب مخطط المستطيلات، وإذا كانت لديك قيمة عددية لكل خلية في شبكة فقد تناسب الخريطة الحرارية. لا تستخدم قاعدة العدد وحدها، واحذف الزوائد فقط مع إبقاء السياق والتسميات والمصدر الضروري.

🌍 طبّق ما تعلمته

يرسم باحث وقت الدراسة مقابل درجة الاختبار لكل طالب في مدرسة كبيرة كمخطط مبعثر، لكن تتداخل النقاط الكثيرة جدًا في كتلة صلبة بحيث لا يمكن قراءة الكثافة. سمِّ تصورًا يمكن أن يكشف أين يقع معظم الطلاب فعليًا، واشرح لماذا يعمل بشكل أفضل هنا من المخطط المبعثر المزدحم.

❓ إجابة السؤال الرئيس

أجاب الدرس عن السؤال الرئيس: كيف نختار تصورًا صادقًا وواضحًا؟ يعتمد الاختيار على نوع البيانات والسؤال والجمهور وطريقة المقارنة. تستخدم الأعمدة لمقارنة الفئات، والخطوط للتغير عبر الزمن أو الترتيب، والمبعثر لعلاقة متغيرين كميين. وتعرض الخريطة الحرارية قيمة عددية لكل خلية في شبكة، بينما يعرض مخطط المستطيلات تكرارات أو نسب توليفات فئتين. وينظم سرد البيانات العرض حول سؤال أو نتيجة مع سياق وأدلة واضحة. ويمكن إبراز المهم، لكن لا يجوز إخفاء ما يغير التفسير. ويجب توضيح المحاور والوحدات والمقياس والمصدر.

nezakr.natega4dk.net

تمارين

1 اختبر فهمك: أجب عن الأسئلة الآتية دون الرجوع إلى قسم «ملحوظة مهمة»، ثم راجع إجاباتك.

2 اقرأ الفقرة التالية وأجب عن كل سؤال.

يساعد تصور البيانات المحلل على الفهم وعلى (1) النتائج للآخرين. ويعتمد اختيار الرسم على نوع البيانات والسؤال. تسمى شبكة القيم التي يمثل لون كل خلية فيها مقدارًا (2)، ويسمى الرسم الذي تمثل مساحاته تكرارات أو نسب توليفات فئتين (3). وعند تنظيم العرض حول سؤال أو نتيجة مع إبقاء السياق الضروري نستخدم (4)، مع توفير (5) مثل العنوان والمحاور والوحدات والمصدر.

1. املأ الفراغات (1) - (5).

3 أجب عن الأسئلة التالية.

(1) من الخيارات التالية (أ - د)، اختر الذي يصف بأنسب طريقة الخريطة الحرارية.

أ رسم بياني يمثل العلاقة بين متغيرين كميين بموضع النقاط.

ب رسم بياني يمثل حجم قيمة ثالثة مقابلة للتوليفات على محورين بتدرج اللون أو اختلافات الصبغة.

جـ رسم بياني يمثل نسبة تركيب الفئات بمساحة القطاعات.

د رسم بياني يمثل التغيرات في البيانات الزمنية بالخطوط.

(2) من الخيارات التالية (أ - د)، اختر الذي ليس عنصرًا أساسيًا مناسبًا لسرد قصص البيانات.

أ جعل من الممكن التعبير في جملة واحدة عما تريد نقله من خلال الرسم البياني (توضيح الرسالة).

ب إزالة العناصر غير المرتبطة بالرسالة والتركيز على الأجزاء التي تريد لفت الانتباه إليها (اختيار العناصر).

جـ ترتيب العنوان وتسميات المحاور والوحدات والملاحظات بشكل صحيح (توفير معلومات داعمة).

د تضمين أكبر عدد ممكن من العناصر غير المرتبطة بالرسالة لزيادة كمية المعلومات.

(3) لكل من أغراض التحليل (1 - 3)، اختر أنسب رسم بياني من الخيارات أدناه (أ - جـ).

1 تريد إدراك متوسطات درجات الحرارة الشهرية لعدة مدن معًا من نظرة واحدة.

2 تريد إظهار الفروق في نسب الشراء لفئة العمر × فئة المنتج باستخدام المساحات.

3 تريد رؤية العلاقة بين متغيرين كميين في حالة يكون فيها عدد النقاط قليلًا بما يكفي لقراءة الكثافة أيضًا بشكل كافٍ.

[الخيارات] أ مخطط مبعثر ب خريطة حرارية جـ مخطط مستطيلات

⭐ الخلاصة

تذكّر:
طابِق الرسم البياني مع البيانات (خريطة حرارية لقيمة عبر محورين؛ مخطط مستطيلات لفئتين بالمساحة؛ مخطط مبعثر/أعمدة/خطوط لمتغير واحد أو متغيرين)، ثم ارو قصة بيانات: رسالة واحدة، وعناصر تخدمها فقط، وتسميات صادقة – وتحقق من وصولها إلى الجمهور.

⭐ فكّر وتحدَّ نفسك

تأمّل: تذكّر رسمًا بيانيًا أربكك. هل كان الرسم البياني الخطأ، أم عناصر كثيرة جدًا، أم تسميات مفقودة؟

تحدَّ: أُعطيت خريطة حرارية لا يبدأ مقياسها اللوني من الصفر، بحيث تبدو الفروق الصغيرة كبيرة ومثيرة. صف الانطباع الخاطئ الذي يخلقه ذلك، والتغيير الوحيد الذي من شأنه أن يجعله صادقًا دون تغيير البيانات.

nezakr.natega4dk.net



الدرس 7-1

أساسيات التعلّم الآلي

أهداف التعلم

- كيف يختلف التعلّم الآلي عن البرمجة
- التصنيف والتجميع وحدود التعلّم الآلي

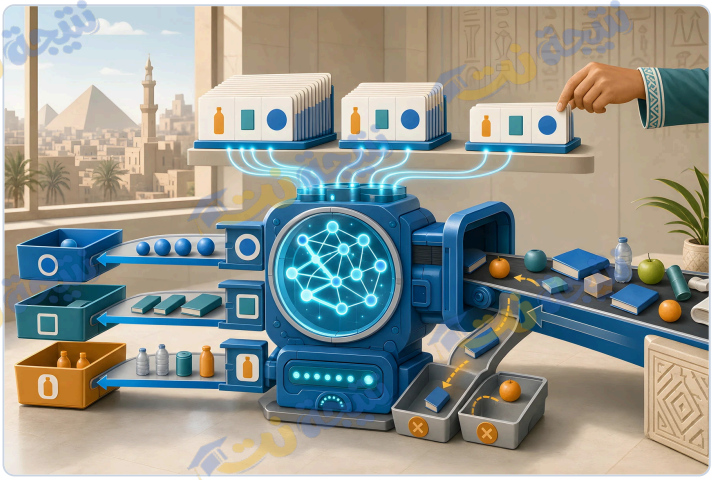
nezakr.natega4dk.net

كيف ندرّب نظامًا على التمييز بين قطة وكلب عندما يصعب كتابة كل القواعد يدويًا؟ نزوّده ببيانات تدريب، ونستخدم خوارزمية تعدّل معلمات نموذج ليتعلم أنماطًا تساعده على التنبؤ في أمثلة جديدة. هذا هو التعلم الآلي، وهو جزء مهم من كثير من أنظمة الذكاء الاصطناعي الحديثة. ويحدد البشر المهمة والبيانات ودالة الهدف وطريقة التقييم؛ لذلك لا يعمل النموذج بمعزل عن اختياراتهم. وتتأثر النتائج بالبيانات والتدريب والتقييم، ويزداد الاحتياج إلى رقابة بشرية كلما ارتفعت خطورة القرار.

؟ السؤال الرئيس: كيف يحل التعلّم الآلي المشكلات بشكل مختلف عن البرمجة العادية، وما هي أنواعه الرئيسية وحدوده؟

📌 ملحوظة مهمّة

1. فكرة التعلّم الآلي



قد يتحسن النموذج ببيانات أكثر إذا كانت مناسبة وعالية الجودة، لكن زيادة البيانات وحدها لا تضمن التحسن.

(1) **التعلم الآلي**: مجال يستخدم خوارزميات تتعلم أنماطًا من البيانات لبناء نموذج يؤدي مهمة محددة، مثل التنبؤ أو التصنيف أو دعم قرار على بيانات جديدة. وهو أحد الأساليب الرئيسة المستخدمة في بناء أنظمة الذكاء الاصطناعي.

(2) تختلف البرمجة التقليدية والتعلّم الآلي اختلافًا كبيرًا في كيفية حل المشكلات.

النهج	ما يقدمه البشر	ما يفعله الحاسوب
البرمجة التقليدية	قواعد القرار (الإجراءات) وبيانات الإدخال	ينتج الإجابات باتباع القواعد
التعلّم الآلي	كميات كبيرة من البيانات وأمثلة على الإجابات الصحيحة	يكتشف قواعد القرار من البيانات بنفسه

(3) يُستخدم التعلّم الآلي في المشكلات التي يصعب فيها على البشر التعبير عن القواعد بالكلمات. *مثال*، الحكم على ما إذا كانت صورة تُظهر كلبًا أو قطة؛ التمييز بين ما إذا كانت رسالة بريد إلكتروني رسالة مزعجة؛ تحويل الصوت إلى نص

(4) بالنسبة لمثل هذه المشكلات، تكون القواعد معقدة جدًا بحيث يصعب على البشر كتابتها بالكامل، لذا يصبح نهج ترك الحاسوب يتعلّم من الأمثلة (البيانات) فعالًا.

(5) التعلم الآلي ليس حلًا سحريًا. فجودة البيانات وتمثيلها وطريقة التدريب والتقييم كلها تؤثر في الأداء. وقد تؤدي البيانات غير الممثلة إلى فروق غير عادلة أو انخفاض الدقة لبعض الحالات. كما قد يصعب تفسير بعض النماذج. لذلك يعتمد مقدار الرقابة البشرية والضوابط على خطورة المهمة، وتزداد الحاجة إليها في القرارات المؤثرة في الناس.

⭐ خريطة الدرس

الفكرة الأساسية:

يستخرج التعلّم الآلي القواعد من البيانات بدلًا من أن تُعطى له. يأتي في ثلاثة أشكال – خاضع للإشراف (التعلّم من أمثلة موسومة)، وغير خاضع للإشراف (إيجاد مجموعات خفية)، ومعزز (التعلّم من المكافآت) – لكن نتائجه تعتمد على البيانات، لذا فإن التحيز ومشكلة الصندوق الأسود تجعلانه دعمًا للحكم البشري، لا بديلًا عنه.

المفاهيم الأساسية:

- التعلّم الآلي
- خاضع للإشراف / غير خاضع للإشراف / معزز
- التصنيف، السمة
- التجميع
- التحيز، مشكلة الصندوق الأسود

مصطلحات أساسية

التعلّم الآلي – تعلّم القواعد من البيانات للحكم على بيانات جديدة.

التحيز – بيانات منحرفة تؤدي إلى أحكام منحرفة.

مشكلة الصندوق الأسود – يصعب رؤية أساس الحكم.

nezakr.natega4dk.net

2. ثلاثة أنواع من طرق التعلّم

(1) يركز هذا الدرس على ثلاثة أنواع رئيسة من التعلم الآلي، مع وجود أساليب أخرى أيضًا.

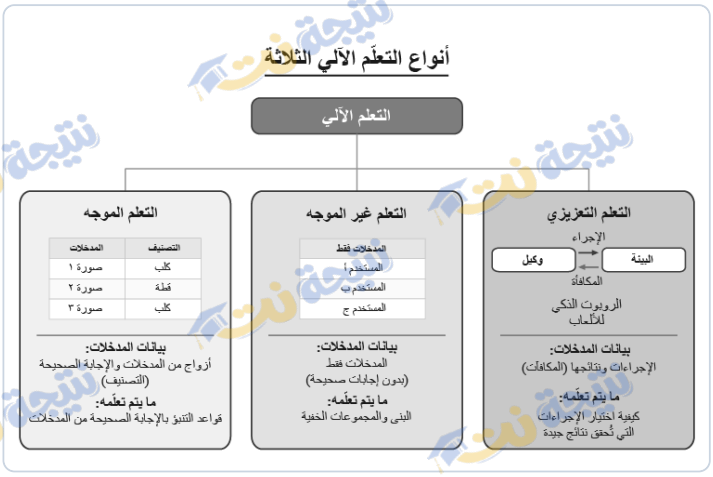
طريقة التعلّم	البيانات المُعطاة	ما يُتعلّم
(التعلّم الخاضع للإشراف)	أمثلة من المدخلات مع قيم أو فئات مستهدفة معروفة (وسوم)	نموذج يربط المدخلات بالقيمة أو الفئة المستهدفة
(التعلّم غير الخاضع للإشراف)	بيانات بلا قيم مستهدفة معروفة مسبقًا	أنماط أو بنية أو مجموعات محتملة في البيانات
(التعلّم المعزز)	حالات وأفعال ومكافآت ناتجة عن التفاعل مع البيئة	سياسة تحدد كيفية اختيار الأفعال للحصول على أفضل النتائج مع الزمن

مصطلحات أساسية

التعلّم غير الخاضع للإشراف – يجد مجموعات بدون وسوم.

التعلّم المعزز – يتعلّم الإجراءات من المكافآت.

الوسم – الإجابة الصحيحة المعروفة في بيانات التدريب.



شكل 7.1.1 – الأنواع الثلاثة للتعلّم الآلي والبيانات التي يستخدمها كل منها

(2) يتحدد اختيار طريقة التعلّم بنوع البيانات المتوفرة وطبيعة المشكلة المراد حلها.

الموقف	طريقة التعلّم المناسبة
توجد قيم أو فئات مستهدفة معروفة في بيانات التدريب، ونريد التنبؤ بقيم جديدة	التعلّم الخاضع للإشراف
تريد استكشاف بنية أو أنماط في بيانات بلا أهداف معروفة مسبقًا	التعلّم غير الخاضع للإشراف
تستطيع تعريف هدف أو مكافأة، ويمكن للوكيل التفاعل المتكرر مع بيئة مناسبة	التعلّم المعزز

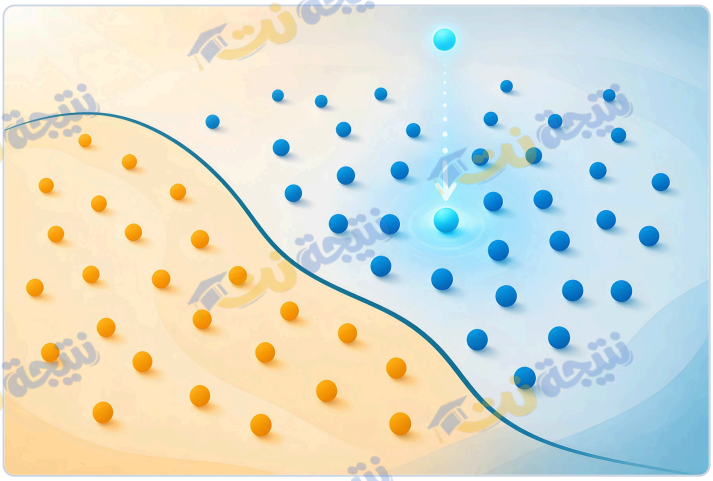
3. التصنيف (مثال تمثيلي للتعلّم الخاضع للإشراف)

مصطلحات أساسية

التعلّم الخاضع للإشراف – يتعلّم من أمثلة موسومة.

التصنيف – التنبؤ بالفئة التي ينتمي إليها الإدخال.

السمة – دليل في البيانات مفيد للتنبؤ.



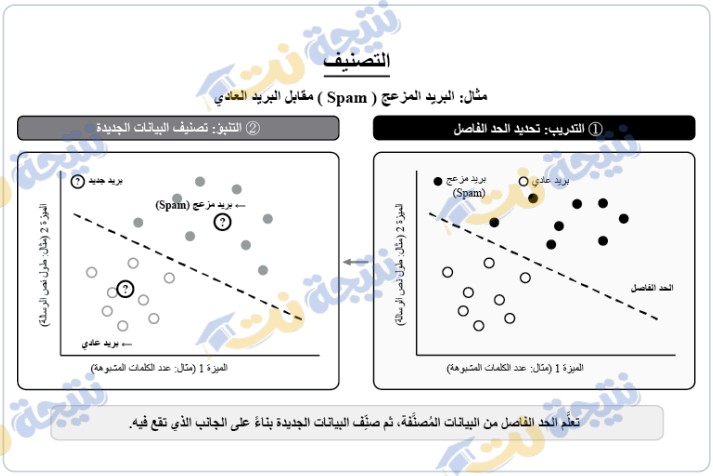
التصنيف يتعلم من أمثلة ذات فئات مستهدفة، ثم يتنبأ بفئة مثال جديد.

nezakr.natega4dk.net

(1) **(التصنيف)**... طريقة للتنبؤ بالنوع (الفئة) الذي تنتمي إليه بيانات الإدخال. إنه مثال تمثيلي للتعلّم الخاضع للإشراف. *مثال*، فرز رسائل البريد الإلكتروني إلى "مزعجة" أو "عادية"؛ فرز الصور الطبية إلى "غير طبيعية" أو "طبيعية"؛ تصنيف الحيوانات في الصور كـ"كلب" أو "قطة" أو "طائر"

(2) يستخدم التصنيف بيانات تدريب تحتوي على أمثلة موسومة بفئات مستهدفة معروفة. وقد يختلف حجم البيانات المطلوب بحسب المهمة وجودة البيانات والنموذج. وبعد التدريب يتنبأ النموذج بفئة بيانات جديدة.

المرحلة	ما يتم فعله
(التدريب)	تعلم معلمات نموذج من أمثلة المدخلات والفئات المستهدفة
(التنبؤ)	استخدام القواعد المتعلَّمة لتحديد الفئة التي تنتمي إليها بيانات جديدة



شكل 7.1.2 – التصنيف: تعلّم حد فاصل من بيانات موسومة، ثم تصنيف بيانات جديدة

(3) تُسمى المعلومات في بيانات التدريب التي تعمل كدليل مفيد للتنبؤ **(السمة)**. *مثال*، في تصنيف الرسائل المزعجة، تُعد أنواع الكلمات الواردة في الرسالة، وعنوان المرسل، وطول النص سمات.

(4) قد يؤثر اختيار أو تعلم تمثيل مناسب للسمات في أداء النموذج. بعض السمات يحددها المطور يدويًا، وبعض النماذج الحديثة تتعلم تمثيلاتها تلقائيًا، ويجب تجنب تسرب البيانات والتحيز غير الضروري.

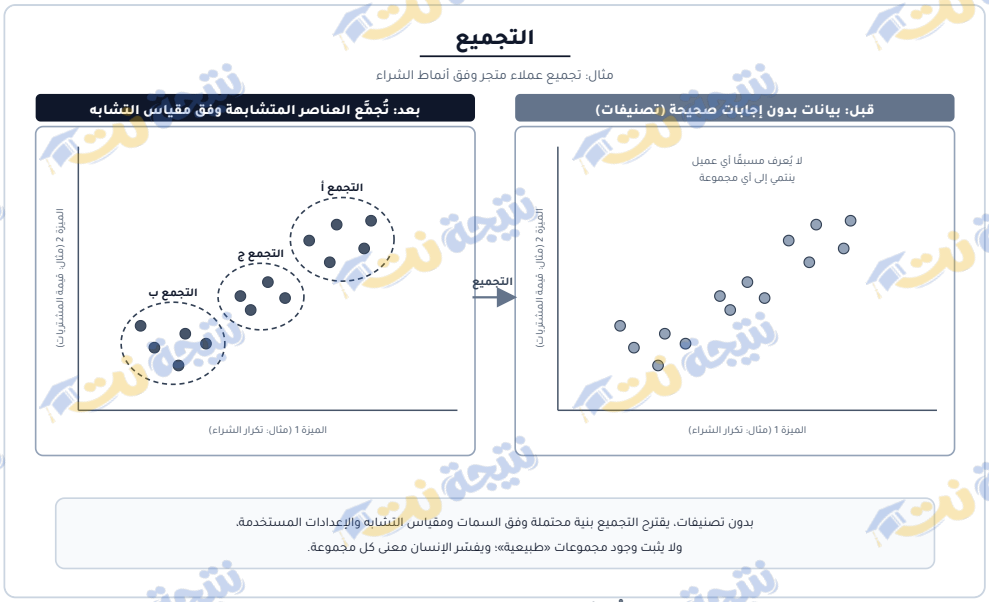
4. التجميع (مثال تمثيلي للتعلّم غير الخاضع للإشراف)

(1) **التجميع (Clustering)**: طريقة تقسم البيانات إلى عناقيد وفق السمات ومقياس التشابه والإعدادات المستخدمة. وقد تكشف بنية تجميعية محتملة، لكنها لا تثبت وجود مجموعات «طبيعية» حقيقية.

💡 توقّف وفكّر

وسوم مُعطاة أو غير مُعطاة – أي كلمة واحدة تفصل بين التصنيف والتجميع؟

nezakr.natega4dk.net



شكل 7.1.3 – التجميع: بدون وسوم، تُجمَّع العناصر المتشابهة وفق مقياس تشابه وإعدادات يحددها المصمم

(2) يكمن الفرق بين التجميع والتصنيف في ما إذا كانت الإجابات الصحيحة (الوسوم) مُعطاة.

الطريقة	الإجابات الصحيحة (الوسوم)	الغرض
التصنيف	مُعطاة مسبقًا	التنبؤ بالفئة التي تنتمي إليها بيانات جديدة
التجميع	غير مُعطاة	البحث عن بنية تجميعية محتملة وفق مقياس التشابه المستخدم

(3) تقيس الخوارزمية التشابه وفق السمات والمقياس والإعدادات التي يحددها المصمم. وقد تتطلب بعض الخوارزميات، مثل k-means، تحديد عدد العناقيد مسبقًا، بينما لا تتطلبه خوارزميات أخرى بالطريقة نفسها. وبعد التشغيل يجب تفسير النتائج وفحص معقوليتها.

(4) يُستخدم التجميع عندما يريد المرء إدراك الصورة الكاملة للبيانات أو عندما يريد النظر إلى البيانات من منظور جديد. يفسّر البشر نتائج التجميع ويستخدمونها في التسويق وتطوير المنتجات وما إلى ذلك.

5. التعلّم المعزز

(1) **التعلم المعزز**: يتفاعل وكيل مع بيئة، ويختار أفعالًا، ويتلقى إشارات مكافأة، ويتعلم سياسة تهدف إلى زيادة مجموع المكافآت المتوقعة مع الزمن. ومن أمثلته التدريب في لعبة أو محاكاة مناسبة.

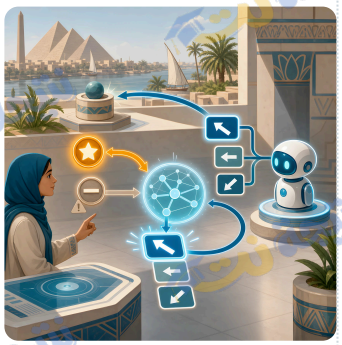
(2) في التعلّم المعزز، يتحقق التعلّم من خلال تجربة توليفات من الإجراءات والنتائج بشكل متكرر.

مصطلحات أساسية

التجميع – تجميع العناصر المتشابهة بدون وسوم.

التعلّم المعزز – تعلّم الإجراءات من المكافآت.

المكافأة – قيمة تقيس مدى جودة إجراء ما.



التعلم المعزز يتعلم من خلال المكافأة والعقاب.

nezakr.natega4dk.net

العنصر	الوصف
(الحالة)	الموقف الذي يوضع فيه الوكيل (الطرف المتعلِّم)
(الإجراء)	الخيارات التي يمكن اختيارها في ذلك الموقف
(المكافأة)	إشارة عددية يحصل عليها الوكيل من البيئة بعد فعل أو خلال التفاعل لتوجيه التعلم نحو هدف محدد

(3) ما يميز التعلّم المعزز عن نوعي التعلّم الآخرين هو أن الإجابة الصحيحة ليست ثابتة على إمكانية واحدة. على عكس التعلّم الخاضع للإشراف، حيث توجد مطابقة "هذا الإدخال له هذه الإجابة الصحيحة"، يتعلّم التعلّم المعزز تدريجيًا خيارات أفضل من نتائج الإجراءات المجرَّبة.

(4) يناسب التعلم المعزز مهامًا يمكن تعريف هدف أو مكافأة لها، ويستطيع الوكيل التفاعل مع بيئة آمنة ومناسبة. وغالبًا تستخدم المحاكيات أو الألعاب لتكرار التجربة دون مخاطر العالم الحقيقي.

الفكرة الرئيسة

يقلب التعلّم الآلي البرمجة رأسًا على عقب: بدلًا من أن يكتب البشر القواعد، يكتشف الحاسوب القواعد من البيانات. يتعلّم بثلاث طرق – خاضع للإشراف (أمثلة موسومة ← التصنيف)، وغير خاضع للإشراف (بدون وسوم ← التجميع)، ومعزز (مكافآت). ومع ذلك، فإن النتائج جيدة وعادلة بقدر جودة البيانات فقط، ويمكن أن يكون المنطق صندوقًا أسود – لذا يدعم التعلّم الآلي الحكم البشري بدلًا من استبداله.

توقّف وفكّر

التعلّم من المكافآت من خلال التجربة والخطأ – أي نوع من الأنواع الثلاثة هو ذلك؟

✎ مثال محلول

(1) ضع علامة أمام كل عبارة (أ - د) بعلامة ○ إذا كانت صحيحة أو × إذا كانت خاطئة.

أ التعلّم الآلي تقنية تتعلّم الأنماط والانتظامات من كميات كبيرة من البيانات بحيث يمكن إصدار أحكام وتنبؤات على بيانات جديدة.

ب في التعلّم الخاضع للإشراف، يتعلّم النموذج من بيانات بدون إجابات صحيحة (وسوم).

ج التجميع طريقة تجمع البيانات وفقًا لوسوم مُعطاة مسبقًا.

د التعلّم المعزز طريقة يُتعلّم فيها كيفية اختيار إجراءات أفضل بناءً على المكافأة المحصَّلة نتيجة لإجراء ما.

(2) طابق كل وصف (1 - 3) مع أنسب مصطلح من الخيارات أدناه (أ - ج).

1 طريقة تعلّم تتعلّم قواعد للتنبؤ بالإجابات الصحيحة للمدخلات من أزواج الإدخال والإجابة الصحيحة

2 طريقة تجد مجموعات من العناصر المتشابهة من بيانات بدون إجابات صحيحة

3 معلومات في بيانات الإدخال تعمل كدليل للتنبؤ في التصنيف

[الخيارات] أ سمة ب تعلّم خاضع للإشراف ج تجميع

✓ الحلّ

(1)

أ تعريف صحيح للتعلّم الآلي. لذلك، ○.

ب البيانات بدون إجابات صحيحة (وسوم) هي التعلّم غير الخاضع للإشراف. في التعلّم الخاضع للإشراف، تُعطى الإجابات الصحيحة. لذلك، ×.

ج التجميع طريقة تجد مجموعات طبيعية من بيانات بدون وسوم مُعطاة. لذلك، ×.

د وصف صحيح للتعلّم المعزز. لذلك، ○.

(2)

1: التعلّم من أزواج الإدخال والإجابة الصحيحة هو التعلّم الخاضع للإشراف. ب

nezakr.natega4dk.net

2: إيجاد مجموعات من بيانات غير موسومة هو التجميع. **ج**

3: المعلومات التي تعمل كدليل للتنبؤ هي السمة. **أ**

تدرّب

1 أجب عن الأسئلة التالية.

1. ما هو مصطلح التقنية التي تتعلّم الأنماط والانتظامات من كميات كبيرة من البيانات بحيث يمكن إصدار أحكام وتنبؤات على بيانات جديدة؟
2. في التعلّم الآلي، ما هو مصطلح المعلومات في بيانات الإدخال التي تعمل كدليل مفيد للتنبؤ؟
3. ما هو مصطلح الطريقة التمثيلية للتعلّم غير الخاضع للإشراف التي تجمع العناصر المتشابهة معًا؟
4. في التعلّم المعزز، ما هو مصطلح القيمة العددية التي تمثل التقييم المُعطى نتيجة لإجراء ما؟

2 أجب عن الأسئلة التالية.

(1) من الخيارات التالية (أ - د)، اختر الذي ليس موقفًا مناسبًا يُستخدم فيه التعلّم الآلي.

أ التمييز بين نوع الحيوان في صورة.

ب تحويل الصوت إلى نص.

ج حساب مجموع عددين صحيحين.

د التمييز بين ما إذا كانت رسالة بريد إلكتروني مزعجة.

(2) لكل من الأغراض (1 - 3)، اختر طريقة التعلّم الأنسب من الخيارات أدناه (أ - ج).

1 التنبؤ، من بيانات سجل الشراء السابقة وما إذا كان قد تم الشراء أم لا، بما إذا كان عميل جديد سيشتري المنتج.

2 إيجاد مجموعات من العملاء ذوي اتجاهات الشراء المتشابهة من بيانات شراء العملاء.

3 جعل روبوت يمشي ويتعلّم حركات تسمح له بالتقدم دون السقوط من خلال التجربة والخطأ.

[الخيارات] أ تعلّم غير خاضع للإشراف ب تعلّم خاضع للإشراف ج تعلّم معزز

(3) من الخيارات التالية (أ - د)، اختر الذي يصف بأنسب طريقة الفرق بين التصنيف والتجميع.

أ يقسّم التصنيف البيانات إلى عدد قليل من المجموعات، ويقسّم التجميع البيانات إلى مجموعات كثيرة.

ب يتعامل التصنيف فقط مع البيانات العددية، ويتعامل التجميع فقط مع البيانات النصية.

ج يتعلّم التصنيف من بيانات ذات إجابات صحيحة (وسوم)، ويتعلّم التجميع من بيانات بدون إجابات صحيحة.

د التصنيف تعلّم غير خاضع للإشراف، والتجميع تعلّم خاضع للإشراف.

(4) من الخيارات التالية (أ - د)، اختر الذي ليس نقطة تحذير مناسبة عند استخدام التعلّم الآلي.

أ إذا كانت بيانات التدريب متحيزة، فقد تكون أحكام النموذج متحيزة أيضًا.

ب قد يصعب رؤية أساس أحكام نموذج التعلّم الآلي، لذا فإن المراجعة البشرية ضرورية للأحكام المهمة.

ج التعلّم الآلي حل سحري لكل شيء ويمكنه إصدار أحكام أدق من البشر في أي مشكلة.

د من المهم استخدام التعلّم الآلي كأداة تدعم الحكم البشري.

سؤال على نمط الامتحان

اشرح لماذا يجب أن يدعم التعلّم الآلي الحكم البشري بدلًا من استبداله. **[6]**

راجع: التحيز في البيانات ومشكلة الصندوق الأسود.

توقّف وفكّر

أي مهمة لديها قاعدة بسيطة يمكن للإنسان كتابتها بالفعل – بحيث لا تحتاج إلى تعلّم؟

⚙️ فكّر كمهندس: ابحث ثم قرر

تمتلك مدرستك سنوات من البيانات وتريد مساعدة في مهمتين: (i) الإشارة مبكرًا إلى الطلاب الذين قد يحتاجون دعمًا إضافيًا، و(ii) اكتشاف مجموعات طبيعية من عادات الدراسة لم تكن معروفة من قبل. أنت تقدم المشورة حول كيفية استخدام التعلّم الآلي بمسؤولية.

1 · **حدد الطريقة.** بالنسبة للمهمة (i) والمهمة (ii)، سمِّ نوع التعلّم المناسب (خاضع للإشراف، غير خاضع للإشراف، أو معزز). لكل منهما، سمِّ عنصر بيانات واحد ستحتاج المدرسة إلى جمعه، ونوع طالب واحد قد تفوته تلك البيانات.

2 · **حلل المخاطر.** سمِّ طريقة واحدة يمكن أن تكون بها بيانات التدريب للمهمة (i) متحيزة (على سبيل المثال، إذا كانت تشمل فقط الطلاب الذين طلبوا المساعدة بالفعل)، وقل الضرر الذي يمكن أن تسببه مشكلة الصندوق الأسود إذا تصرف معلم بناءً على إشارة دون فهمها – وهو حكم لا تقدمه لك جداول الدرس.

3 · **حدد دور الإنسان.** اذكر القرارات التي ستبقي فيها إنسانًا مسؤولًا، والقرارات التي ستسمح فيها للنموذج بالمساعدة، ولماذا. **في أزواج،** قارِنوا الخطط واتفقوا على الضمانة الواحدة التي لن تُزيلوها أبدًا.

☐ **دع النموذج يقرر بمفرده** ☐ **اختر الطريقة الصحيحة، وأبقِ إنسانًا مسؤولًا**

قدِّم دليلًا على إجابتك.

هل تواجه صعوبة؟ ابدأ من البيانات المتوفرة لديك: هل توجد إجابات صحيحة يمكن التعلّم منها (خاضع للإشراف) أم لا توجد (غير خاضع للإشراف)؟ ثم اسأل، بالنسبة لكل قرار، ما الذي سيكلفه أن تكون مخطئًا – كلما ارتفعت تلك التكلفة، وجب أن يبقى الإنسان مسؤولًا أكثر.

🌍 طبّق ما تعلمته

يريد بنك برنامجًا يقرأ كل طلب قرض ويشير إلى الطلبات التي تبدو محفوفة بالمخاطر ليراجعها موظف بشري. سمِّ نوع التعلّم الذي يمثل هذا، وسمِّ سمة واحدة قد يستخدمها النموذج، واشرح لماذا يجب أن يبقى الموظف في الحلقة بدلًا من ترك النموذج يقرر بمفرده.

❓ إجابة السؤال الرئيس

أجاب الدرس عن السؤال الرئيس: كيف يختلف التعلم الآلي وما أنواعه وحدوده؟ يحدد المطورون المهمة والبيانات والهدف والتقييم، ثم تتعلم الخوارزمية معلمات نموذج. يتعلم الخاضع للإشراف من قيم أو فئات مستهدفة معروفة، ويبحث غير الخاضع عن بنية في بيانات بلا أهداف معروفة، ويتعلم المعزز سياسة من التفاعل والمكافآت. يتنبأ التصنيف بفئة، بينما يقسم التجميع البيانات وفق مقياس تشابه وإعدادات محددة. وتتأثر النتائج بالبيانات والتدريب والتقييم، لذلك نختبر التحيز والأداء، ونزيد الرقابة البشرية في القرارات عالية المخاطر.

مصطلحات أساسية

الوسم – الإجابة الصحيحة المعروفة المرفقة بالبيانات.

الإنسان في الحلقة – شخص يراجع أو يوافق على ناتج النموذج.

nezakr.natega4dk.net

تمارين

1 اختبر فهمك: أجب عن الأسئلة الآتية دون الرجوع إلى قسم «ملحوظة مهمة»، ثم راجع إجاباتك.

2 اقرأ الفقرة التالية وأجب عن كل سؤال.

يبني التعلم الآلي نماذج تتعلم أنماطًا من البيانات لتؤدي مهمة على بيانات جديدة. يركز الدرس على التعلم (1) من أمثلة ذات أهداف معروفة، و(2) للبحث عن بنية بلا أهداف معروفة، و(3) الذي يتعلم من (4) خلال التفاعل. وفي (5) نتنبأ بالفئة باستخدام (6) أو تمثيلات مناسبة، بينما يقسم (7) البيانات إلى عناقيد وفق مقياس تشابه وإعدادات محددة.

1. املأ الفراغات (1) - (7).

3 أجب عن الأسئلة التالية.

(1) من الخيارات التالية (أ - د)، اختر الذي يصف بأنسب طريقة الفرق بين البرمجة التقليدية والتعلّم الآلي.

أ في البرمجة التقليدية، يقدم البشر قواعد القرار؛ وفي التعلّم الآلي، يكتشف الحاسوب قواعد القرار من البيانات بنفسه.

ب تتطلب البرمجة التقليدية كميات كبيرة من البيانات، ويتطلب التعلّم الآلي قواعد فقط.

ج تندرج البرمجة التقليدية ضمن الذكاء الاصطناعي، لكن التعلّم الآلي لا يندرج ضمن الذكاء الاصطناعي.

د يمكن للبرمجة التقليدية التعامل مع مشكلات جديدة، لكن التعلّم الآلي يمكنه فقط التعامل مع مشكلات ثابتة.

(2) لكل من المواقف (1 - 3)، اختر طريقة التعلّم الأنسب من الخيارات أدناه (أ - ج).

1 توجد كميات كبيرة من نصوص المقالات الإخبارية وفئاتها (سياسة، اقتصاد، رياضة، إلخ)، وتريد تحديد فئة المقالات الجديدة تلقائيًا.

2 لديك سجلات للمنتجات التي اشتراها مستخدمو متجر إلكتروني وتريد إيجاد مجموعات من المستخدمين ذوي اتجاهات الشراء المتشابهة.

3 تريد أن يلعب حاسوب لعبة ويتعلّم العمليات التي تؤدي إلى درجات عالية من خلال التجربة والخطأ.

[الخيارات] أ تعلّم معزز ب تجميع ج تصنيف

(3) من الخيارات التالية (أ - د)، اختر الذي ليس وصفًا مناسبًا لخصائص التعلّم الآلي وحدوده.

أ يُستخدم التعلّم الآلي في المشكلات التي يصعب فيها على البشر التعبير عن القواعد بالكلمات.

ب تتأثر نتائج نماذج التعلّم الآلي بجودة بيانات التدريب وتحيزها.

ج أساس حكم نموذج التعلّم الآلي يُعرض دائمًا للبشر بطريقة سهلة الفهم.

د من المهم استخدام التعلّم الآلي ليس كبديل للحكم البشري، بل كأداة تدعم الحكم البشري.

⭐ الخلاصة

تذكّر:
يكتشف التعلّم الآلي القواعد من البيانات: خاضع للإشراف (موسوم ← التصنيف بالسمات)، وغير خاضع للإشراف (غير موسوم ← التجميع)، ومعزز (مكافآت). انتبه لحدوده – التحيز ومشكلة الصندوق الأسود – وأبقِه داعمًا للحكم البشري.

⭐ فكّر وتحدَّ نفسك

تأمّل: أين ربما واجهت التعلّم الآلي اليوم (التوصيات، وسم الصور، الإدخال الصوتي)؟ ما النوع الذي تعتقد أنه كان؟

تحدٍّ: تم تدريب نموذج للتعرف على الوجوه بشكل أساسي على مجموعة واحدة من الأشخاص ويعمل بشكل سيء مع مجموعات أخرى. اشرح كيف تسببت بيانات التدريب المتحيزة في ذلك، واذكر تغييرًا واحدًا في البيانات من شأنه أن يقلل التحيز.

nezakr.natega4dk.net



الدرس 7-2

الشبكات العصبية والتعلم العميق

أهداف التعلم

- بنية الشبكات العصبية وكيفية عملها، والتدريب بواسطة الأوزان.
- التعلم العميق وأثره في تطور تطبيقات الذكاء الاصطناعي.

nezakr.natega4dk.net

كيف يمكن لآلة أن تميّز كلبًا من قطة، أو أن تحول الكلام إلى نص؟ الشبكات العصبية نماذج رياضية استوحت اسمها تاريخيًا من فكرة الخلايا العصبية، لكنها ليست نسخة من الدماغ. في نموذج مبسط، تنظم وحدات حسابية في طبقات، وتحول كل طبقة تمثيل البيانات إلى تمثيل جديد. تحسب الوحدة مجموعًا موزونًا للمدخلات، وتضيف حد انحياز، ثم تطبق غالبًا دالة تنشيط. وأثناء التدريب تُعدَّل المعلمات لتقليل دالة الخسارة. ويمكن للطبقات المتتابعة أن تتعلم تمثيلات معقدة يصعب وصفها بقواعد يدوية؛ وفي بعض شبكات الصور قد تنتقل التمثيلات من أنماط بسيطة مثل الحواف إلى أشكال أكثر تركيبًا. وقد أسهم التعلم العميق في تحقيق أداء مرتفع جدًا، وأحيانًا أعلى من أداء البشر على مقاييس أو مهام محددة، من دون أن يعني ذلك فهمًا عامًا أو نجاحًا مضمونًا.

السؤال الموجِّه: كيف تعمل بنية مبسطة لشبكة عصبية، وكيف تتغير معلماتها أثناء التدريب، وماذا قد يضيف العمق؟

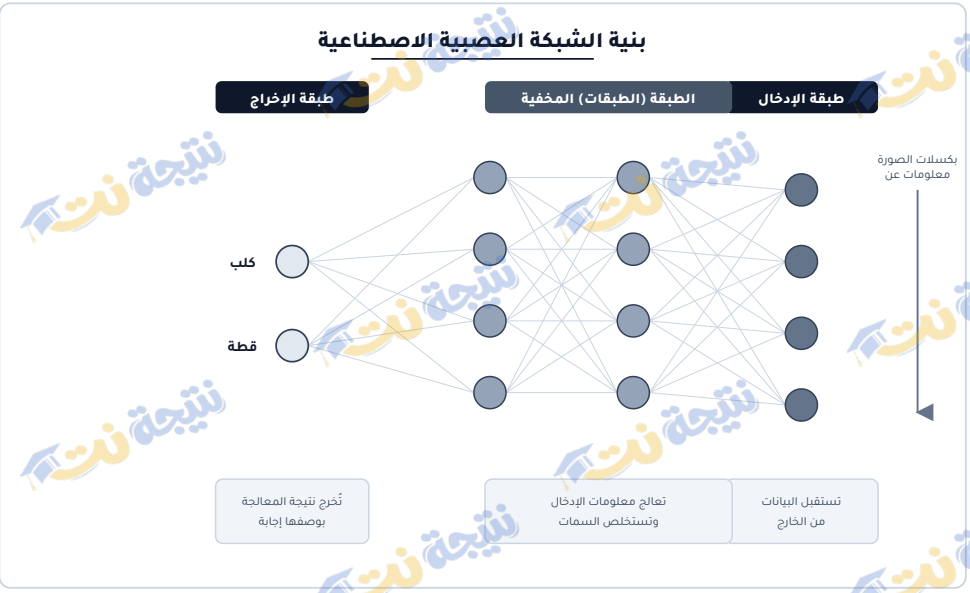
ملحوظة مهمّة

1. بنية الشبكات العصبية

(1) تتكون الشبكة العصبية من وحدات حسابية وروابط تحددها بنية النموذج. وفي المثال التعليمي المبسط، تنظم الوحدات في طبقات وتتصل الطبقات المتجاورة، لكن شكل الاتصال يختلف باختلاف نوع الشبكة.

(2) تُقسَّم الطبقات حسب الدور إلى الأنواع الثلاثة التالية.

نوع الطبقة	الدور
طبقة الإدخال	تمثل القيم التي تدخل إلى النموذج، مثل قيم بكسلات صورة
الطبقة المخفية (الطبقة الوسطى)	تحول تمثيل البيانات إلى تمثيلات جديدة تتعلمها الشبكة
طبقة الإخراج	تنتج قيمًا مناسبة للمهمة، مثل درجات احتمالية الفئات



الشكل 7.2.1 – شبكة عصبية: طبقة الإدخال، الطبقات المخفية، وطبقة الإخراج

(3) تدخل البيانات من تمثيل الإدخال، ثم تحولها الطبقات المخفية، وتنتج طبقة الإخراج قيمًا مناسبة للمهمة. ففي نموذج لتصنيف صورة حيوان، قد تمثل قيم البكسلات الإدخال، وقد تتعلم الطبقات أنماطًا مرتبطة بالحواف أو الأشكال أو تمثيلات أخرى، ثم تنتج طبقة الإخراج درجات أو احتمالات للفئتين «كلب» و«قطة» وتطبق قاعدة القرار.

⭐ خريطة الدرس

الفكرة الأساسية:

الشبكة العصبية عبارة عن طبقات من الخلايا العصبية المتصلة: طبقة إدخال، وطبقات مخفية تستخرج السمات، وطبقة إخراج. تقوم كل خلية عصبية بضرب مدخلاتها في الأوزان وجمعها؛ ويقوم التدريب بضبط هذه الأوزان اعتمادًا على البيانات. يقوم التعلم العميق بتكديس العديد من الطبقات المخفية لبناء السمات خطوة بخطوة – وهذه هي القفزة التي تقف وراء الذكاء الاصطناعي الحديث.

المفاهيم الأساسية:

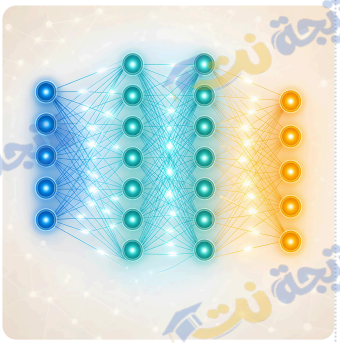
- عقدة (خلية عصبية)، طبقة
- طبقة الإدخال / الطبقة المخفية / طبقة الإخراج
- الوزن، التدريب
- التعلم العميق
- السمات الهرمية

مصطلحات أساسية

عقدة (خلية عصبية) – وحدة معالجة واحدة.

طبقة – مجموعة من الخلايا العصبية في مرحلة واحدة.

طبقة الإدخال / الطبقة المخفية / طبقة الإخراج – تستقبل، تعالج، تجيب.



ترصّ الشبكة العصبية العصبونات في طبقات مترابطة.

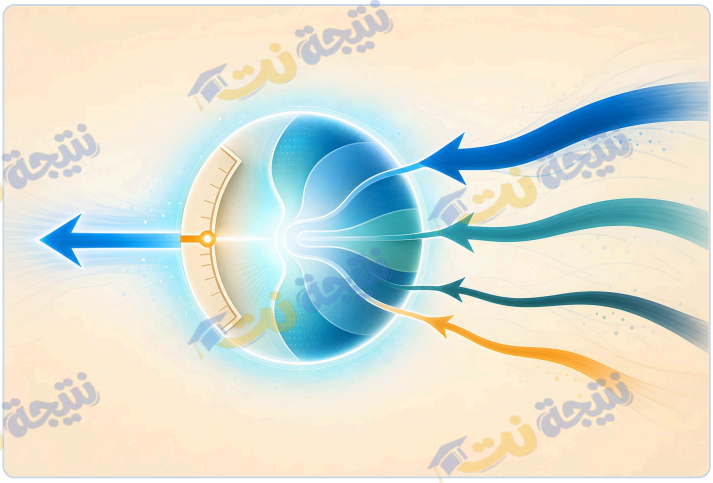
(4) يختار المصمم بنية النموذج وفق المهمة والبيانات والموارد ونتائج التجارب. وقد يفيد عمق أو عدد وحدات أكبر في بعض المسائل، لكنه يزيد الكلفة ولا يضمن تحسن الأداء أو التعميم.

مصطلحات أساسية

الوزن – مقدار التركيز على أحد المدخلات.

التدريب – ضبط الأوزان بناءً على البيانات.

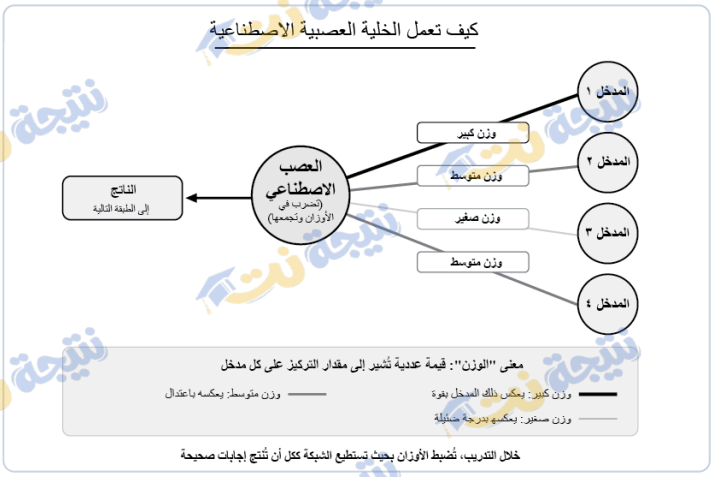
2. كيف تعمل الخلية العصبية



في النموذج الشائع، تحسب الوحدة مجموعًا موزونًا للمدخلات، وتضيف انحيازًا، ثم تطبق دالة تنشيط.

(1) في نموذج مبسط شائع، تستقبل الوحدة قيمًا متعددة، وتضربها في أوزان، وتجمع النواتج، ثم تضيف حد الانحياز وتطبق دالة تنشيط لإنتاج قيمة تنتقل إلى الجزء التالي من الشبكة.

(2) **الوزن (Weight)** معامل عددي يحدد مقدار واتجاه مساهمة قيمة ما في الحساب؛ فقد تكون مساهمته موجبة أو سالبة. ولا يعني الوزن وحده أن كلمة أو سمة بعينها ستقرر النتيجة، لأن التنبؤ ينتج من تفاعل معلمات عديدة عبر الشبكة.



الشكل 7.2.2 – تقوم الخلية العصبية بترجيح مدخلاتها، وجمعها، وتمرير النتيجة

(3) **التدريب** هو عملية تعديل معلمات الشبكة تدريجيًا لتقليل دالة خسارة على بيانات التدريب، باستخدام خوارزمية تحسين، وغالبًا يحسب الانتشار العكسي التدرجات اللازمة لهذا التعديل.

المرحلة	ما يحدث
قبل التدريب	تبدأ المعلمات غالبًا بتهيئة عشوائية أو شبه عشوائية محسوبة، فتكون التنبؤات عادة ضعيفة قبل التعلم
أثناء التدريب	تقاس الخسارة بين المخرجات والقيم المستهدفة، ثم تعدل المعلمات تدريجيًا لتقليلها
بعد التدريب	يُقاس الأداء على بيانات تحقق أو اختبار منفصلة لمعرفة مدى التعميم على أمثلة جديدة؛ ولا يضمن التدريب صحة كل تنبؤ

nezakr.natega4dk.net

(4) يهدف التدريب إلى إيجاد معلمات تقلل الخسارة وتحسن الأداء. لكن انخفاض خطأ التدريب وحده لا يكفي: فقد يحدث فرط التخصيص، لذلك نقيّم النموذج على بيانات منفصلة لم تستخدم في تعديل معلماته.

> **توقّف وفكّر**
>
> لماذا يتيح تكديس المزيد من الطبقات المخفية للشبكة التقاط سمات أكثر تعقيدًا؟

3. التعلم العميق

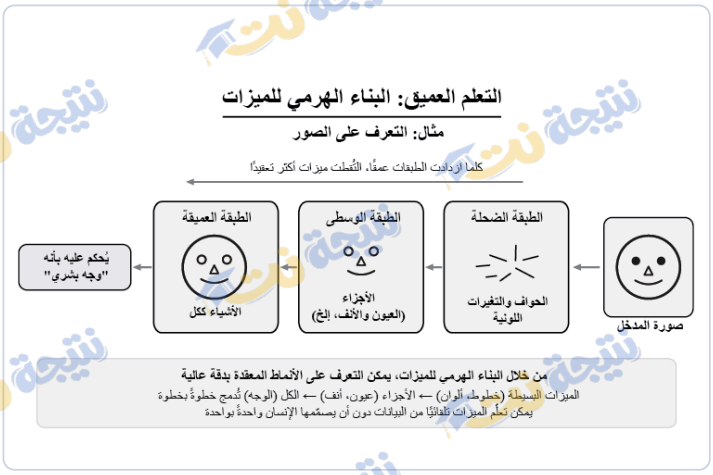


تقوم الطبقات العميقة ببناء التمثيلات بشكل تدرجي: بدءاً من الحواف إلى الكائنات الكاملة.

(1) **التعلم العميق** هو تعلم باستخدام شبكات عصبية متعددة الطبقات قادرة على تعلم تمثيلات متتابعة من البيانات. تشير كلمة «عميق» إلى عمق البنية، ولا يوجد عدد عالمي واحد من الطبقات يفصل كل نموذج عميق عن غيره.

(2) قد تتعلم الطبقات المتتابعة تمثيلات أكثر تجريدًا. فمثلًا توضح بعض شبكات الصور تعليميًا انتقالًا محتملًا من أنماط بسيطة مثل الحواف وتغير اللون، إلى تراكيب وأجزاء، ثم إلى تمثيلات مرتبطة بأجسام كاملة. هذا مثال تقريبي وليس مسارًا ثابتًا لكل شبكة.

موضع الطبقة	أمثلة على السمات المستخرجة
الطبقات الضحلة	سمات بسيطة مثل الخطوط وتغيرات اللون
الطبقات الوسطى	أشكال الأجزاء مثل العينين والأنف والأذنين
الطبقات العميقة	بنية الوجه أو الجسم بأكمله



الشكل 7.2.3 – يبني التعلم العميق السمات خطوة بخطوة، من الحواف إلى الأجسام الكاملة

(3) قد تساعد هذه التمثيلات المتتابعة على تعلم علاقات معقدة وتحقيق أداء مرتفع إذا كانت البيانات والبنية والتدريب والتقييم مناسبة، لكن زيادة العمق وحدها لا تضمن ذلك.

(4) أسهم اجتماع عوامل متعددة في انتشار التعلم العميق: تحسن المعماريات والخوارزميات ودوال التنشيط والتحسين، وتوافر بيانات وأدوات برمجية أفضل، وزيادة القدرة الحاسوبية. وتسرّع وحدات مثل GPU وTPU عمليات كثيرة، لكنها ليست شرطًا لكل استخدام أو استدلال.

nezakr.natega4dk.net

4. لماذا أحدث التعلم العميق قفزة في الذكاء الاصطناعي

(1) حقق التعلم العميق تقدمًا كبيرًا في مهام محددة مثل تصنيف الصور والتعرف على الكلام ومعالجة اللغة. وقد يقارب أداء البشر أو يتفوق عليه في اختبار محدد، لكنه قد يفشل خارج ظروف التقييم ولا يعني امتلاك ذكاء عام.

(2) تشمل المجالات التي شهدت تقدمًا كبيرًا بشكل خاص ما يلي.

المجال	أمثلة
(التعرف على الصور)	أنظمة التعرف على الوجه، الدعم التشخيصي للصور الطبية، الكشف عن الأجسام للسيارات ذاتية القيادة
(التعرف على الصوت)	التشغيل الصوتي للسماعات الذكية، تحويل الكلام إلى نص، الترجمة الآلية
(معالجة اللغة الطبيعية)	الترجمة الآلية، تلخيص النصوص، الإجابة عن الأسئلة

(3) في أساليب تقليدية كثيرة صمم البشر سمات يدويًا. أما شبكات التعلم العميق فيمكنها تعلم تمثيلات كثيرة من البيانات، مع بقاء اختيار المهمة والبيانات والبنية والهدف والتقييم مسؤولية بشرية.

(4) دعمت هذه القدرات تطور نماذج توليد النصوص والصور والصوت والبرمجيات، مع ضرورة تقييم جودة مخرجاتها ومخاطرها لكل استخدام.

الفكرة الرئيسة

تُكدّس الشبكة العصبية طبقات من الخلايا العصبية البسيطة: الإدخال، والمخفية (المستخرِجة للسمات)، والإخراج. تقوم كل خلية عصبية بترجيح مدخلاتها وجمعها، ويضبط التدريب هذه الأوزان بناءً على البيانات لتقليل الخطأ، من دون أن يضمن ذلك صحة كل تنبؤ. يبني التعلم العميق – أي العديد من الطبقات المخفية – السمات بشكل هرمي (حواف ← أجزاء ← كل)، وبمجرد أن أصبحت البيانات الضخمة ومعالجات GPU عملية، سمح ذلك بتعلم السمات تلقائيًا وبأداء مرتفع جدًا في مهام الصور والصوت واللغة، يقارب أداء البشر أو يتفوق عليه أحيانًا في اختبارات محددة، من دون أن يعني ذلك فهمًا عامًا.

توقّف وفكّر

ما هما الأمران اللذان كان يجب أن يتوفرا قبل أن يعمل التعلم العميق بشكل عملي؟

✎ مثال محلول

(1) ضع علامة على كل عبارة (أ - د) بـ○ إذا كانت صحيحة أو × إذا كانت خاطئة.

أ تحتوي الشبكة العصبية على ثلاثة أنواع من الطبقات: طبقة الإدخال، والطبقة المخفية، وطبقة الإخراج.

ب طبقة الإدخال هي الطبقة التي تُخرج نتائج المعالجة كإجابة إلى الخارج.

جـ التعلم العميق هو تقنية تستخدم شبكة عصبية بطبقة مخفية واحدة فقط.

د تدريب الشبكة العصبية هو عملية ضبط الأوزان بين الخلايا العصبية باستخدام كمية كبيرة من البيانات.

(2) طابق كل وصف (1 - 3) مع المصطلح الأنسب من الخيارات أدناه (أ - جـ).

1 في الشبكة العصبية، الطبقة التي تستقبل البيانات من الخارج

2 في الشبكة العصبية، قيمة عددية تمثل مقدار التركيز على كل مدخل

3 في الشبكة العصبية، الطبقة التي تعالج معلومات الإدخال وتستخرج السمات

[الخيارات] أ الطبقة المخفية ب الوزن جـ طبقة الإدخال

توقّف وفكّر

أي طبقة تستقبل البيانات، وأيها تعطي الإجابة النهائية؟

✓ الحلّ

(1)

أ هناك ثلاثة أنواع من الطبقات: الإدخال، والمخفية، والإخراج. لذلك، ○.

nezakr.natega4dk.net

ب طبقة الإخراج هي التي تخرج النتائج إلى الخارج. طبقة الإدخال هي الطبقة التي تستقبل البيانات من الخارج. لذلك، ×.

جـ التعلم العميق هو تقنية تستخدم شبكات عصبية ذات طبقات مخفية عديدة مكدسة بعمق. لذلك، ×.

د وصف صحيح للتدريب. لذلك، ○.

(2)

1: التي تستقبل البيانات هي طبقة الإدخال. **جـ**

2: القيمة التي تمثل مقدار التركيز على مدخل ما هي الوزن. **ب**

3: التي تعالج المعلومات وتستخرج السمات هي الطبقة المخفية. **أ**

تدرّب

1 أجب عن الأسئلة التالية.

1. ما هو مصطلح وحدة المعالجة الفردية التي تتكون منها الشبكة العصبية؟
2. في الشبكة العصبية، ما هو مصطلح الطبقة التي تعالج معلومات الإدخال وتستخرج السمات؟
3. في الشبكة العصبية، ما هو مصطلح القيمة العددية التي تمثل مقدار التركيز على كل مدخل؟
4. ما هو مصطلح التقنية التي تستخدم شبكات عصبية ذات طبقات مخفية عديدة مكدسة بعمق؟

2 أجب عن الأسئلة التالية.

(1) من بين الخيارات التالية (أ - د)، اختر الخيار الذي يصف طبقات الشبكة العصبية بأنسب شكل.

أ تتمثل وظيفة طبقة الإدخال في إخراج نتائج المعالجة إلى الخارج.

ب تتمثل وظيفة الطبقة المخفية في استقبال البيانات من الخارج.

جـ تتمثل وظيفة طبقة الإخراج في معالجة معلومات الإدخال واستخراج السمات.

د تدخل البيانات عند طبقة الإدخال، وتتم معالجتها أثناء مرورها بالطبقات المخفية، وتصل إلى طبقة الإخراج وفق هذا التدفق.

(2) من بين الخيارات التالية (أ - د)، اختر الخيار الذي يصف تدريب الشبكة العصبية بأنسب شكل.

أ قبل بدء التدريب، تُضبط الأوزان بين الخلايا العصبية على قيم صحيحة.

ب مع تقدم التدريب، يتم إيجاد تركيبة من الأوزان تتيح للشبكة ككل إنتاج إجابات صحيحة.

جـ يتم التدريب من خلال زيادة عدد الخلايا العصبية.

د بمجرد اكتمال التدريب، تصبح الشبكة غير قادرة على التعامل مع بيانات جديدة.

(3) من بين الخيارات التالية (أ - د)، اختر الخيار الذي لا يمثل خاصية مناسبة للتعلم العميق.

أ له بنية بها العديد من الطبقات المخفية مكدسة بعمق.

ب يمكنه بناء سمات البيانات خطوة بخطوة.

جـ قد تسمح زيادة العمق بتمثيل علاقات أكثر تعقيدًا، لكنها لا تضمن تحسن الأداء.

د حتى مع زيادة عدد الطبقات المخفية، لا تتغير صعوبة المشكلات التي يمكن التعامل معها.

(4) بالنسبة لشبكة تعلم عميق تقوم بالتعرف على الصور، أعد ترتيب التالي (أ - جـ) بالترتيب الذي تُستخرج به السمات من الطبقات الضحلة إلى الطبقات العميقة.

أ بنية الوجه أو الجسم بأكمله

ب سمات بسيطة مثل الخطوط وتغيرات اللون

جـ أشكال الأجزاء مثل العينين والأنف والأذنين

سؤال على نمط الامتحان

اشرح كيف يمكن لشبكة عصبية ذات خلايا عصبية بسيطة أن تُصدر أحكامًا معقدة، مع الإشارة إلى الأوزان والتدريب. **[6]**

ارجع إلى: ترجيح المدخلات وضبط الأوزان بناءً على البيانات.

توقّف وفكّر

قبل التدريب، هل الأوزان صحيحة بالفعل؟ ما الذي يغيّرها؟

توقّف وفكّر

هل مجرد إضافة المزيد من الطبقات المخفية يتيح دائمًا للشبكة حل مشكلات أصعب؟

nezakr.natega4dk.net

مصطلحات أساسية

بيانات التدريب – الأمثلة الموسومة التي تتعلم منها الشبكة.

التعميم – العمل بشكل صحيح على بيانات جديدة لم تُشاهد من قبل.

⚙️ فكّر كمهندس: ابحث ثم قرر

يريد فريق إنشاء تطبيق يقرأ الأرقام العربية المكتوبة بخط اليد من سجلات المدارس الممسوحة ضوئيًا ويحوّلها إلى أرقام مطبوعة. أنت تقدّم المشورة حول بنائه باستخدام شبكة تعلم عميق.

1 · صمّم بنية مبدئية. حدد تمثيل الإدخال المناسب لصورة الرقم، وما الذي ستنتجه طبقة الإخراج، واشرح الفائدة المحتملة للطبقات المخفية من دون افتراض عدد ثابت أو أنها تضمن النجاح.

2 · اشرح العمق. باستخدام فكرة السمات الهرمية، صف ما يمكن أن تلتقطه الطبقات الضحلة والوسطى والعميقة لرقم مكتوب بخط اليد – سلسلة يوضحها الدرس بالنسبة للوجوه لكن لا يفصّلها للأرقام، لذا عليك أن ترسمها بنفسك.

3 · قيّم المتطلبات والمخاطر. اذكر ثلاثة عوامل مهمة لنجاح النموذج، مثل بيانات متنوعة وذات جودة، وحوسبة مناسبة، وطريقة تدريب وتقييم جيدة. ثم وضح كيف قد يؤدي ضعف تمثيل بعض أنماط الكتابة في بيانات التدريب إلى أداء غير عادل، واقترح تحسينًا للبيانات وطريقة لقياس أثره على كل فئة.

☐ **كتابة القواعد لكل رقم يدويًا** ☐ **تدريب شبكة عميقة على أمثلة كثيرة**

قدّم دليلًا على إجابتك.

هل تواجه صعوبة؟ استخدم مثال الوجه في الدرس كنموذج، ثم انقله إلى رقم: اسأل ما الذي يمكن اعتباره "حافة"، و"جزءًا"، و"كلًا" بالنسبة لرقم مكتوب. بالنسبة للمخاطرة، اذكر خط يد مَن رأته الشبكة أثناء التدريب – ثم اسأل خط يد مَن لم تره أبدًا.

🌍 طبّق ما تعلمته

يحوّل مساعد صوتي الأوامر المنطوقة إلى نص ويعمل جيدًا في حالات كثيرة لم تبرمج لها قواعد منفصلة. اذكر مجال التطبيق، واشرح كيف يعمم النموذج أنماطًا تعلمها، ثم اذكر سببًا محتملًا للخطأ مثل لهجة قليلة التمثيل أو ضوضاء أو كلمة نادرة.

❓ إجابة السؤال الرئيس

شرح الدرس بنية مبسطة للشبكة العصبية: تمثيل إدخال، وطبقات تحول التمثيلات، ومخرجات مناسبة للمهمة. في الوحدة الشائعة تضرب المدخلات في أوزان، وتجمع مع انحياز، ثم تطبق دالة تنشيط. ويعدل التدريب المعلمات لتقليل الخسارة، بينما يقيس اختبار منفصل مدى التعميم ولا يضمن صحة كل تنبؤ. يستخدم التعلم العميق شبكات متعددة الطبقات قد تتعلم تمثيلات متتابعة؛ وقد ساعدت تطورات البيانات والحوسبة والخوارزميات والمعماريات والأدوات على نجاحه في مهام محددة، لكن زيادة العمق أو البيانات وحدها لا تضمن أداءً أفضل.

nezakr.natega4dk.net

تمارين

1 اختبر فهمك: أجب عن الأسئلة الآتية دون الرجوع إلى قسم «ملحوظة مهمة»، ثم راجع إجاباتك.

2 اقرأ المقطع التالي وأجب عن كل سؤال.

في النموذج المبسط، تتكون الشبكة العصبية من وحدات حسابية منظمة في (1). يمثل (2) القيم الداخلة، وتحول (3) التمثيلات، وتنتج (4) قيمًا تناسب المهمة. تحسب الوحدة مجموعًا موزونًا باستخدام (5)، وتضيف (6)، ثم تطبق غالبًا (7). يعدل التدريب هذه المعلمات لتقليل الخسارة، ويقاس (8) على بيانات منفصلة لمعرفة الأداء على أمثلة جديدة.

1. املأ الفراغات (1) - (8).

3 أجب عن الأسئلة التالية.

(1) من بين الخيارات التالية (أ - د)، اختر الخيار الذي يصف كيفية عمل الخلية العصبية في الشبكة العصبية بأنسب شكل.

أ تمرر كل خلية عصبية مدخلًا واحدًا كما هو إلى الطبقة التالية.

ب تقوم كل خلية عصبية بضرب مدخلات متعددة في الأوزان، وجمعها، وتمرير النتيجة إلى الطبقة التالية.

ج تحسب كل خلية عصبية متوسط قيم المدخلات وتمرر النتيجة إلى الطبقة التالية.

د تمرر كل خلية عصبية أكبر مدخل فقط إلى الطبقة التالية.

(2) من بين الخيارات التالية (أ - د)، اختر الخيار الذي يصف خلفية جعل التعلم العميق عمليًا بأنسب شكل.

أ أصبحت الحوسبة أبسط عن طريق تقليل عدد الطبقات المخفية إلى واحدة.

ب أصبحت كميات كبيرة من البيانات وقدرة حوسبية عالية متاحة.

ج تم إلغاء الاتصالات بين الخلايا العصبية، مما بسّط المعالجة.

د تم إرساء طريقة يقوم فيها البشر بضبط الأوزان واحدة تلو الأخرى يدويًا.

(3) من بين الخيارات التالية (أ - د)، اختر الخيار الذي يصف لماذا أحدث التعلم العميق قفزة في الذكاء الاصطناعي بأنسب شكل.

أ يمكن تعلم السمات تلقائيًا من البيانات دون أن يصممها البشر واحدة تلو الأخرى.

ب أصبحت الحواسيب قادرة على التفكير بنفس آلية عمل الدماغ البشري.

ج أصبحت الحواسيب قادرة على فهم معنى الكلمات فهمًا كاملًا.

د أصبحت الحواسيب قادرة على إصدار أحكام دقيقة تمامًا في أي مسألة.

⭐ الخلاصة

تذكّر:
الشبكة العصبية = طبقات من الخلايا العصبية (إدخال ← مخفية ← إخراج)؛ كل خلية عصبية ترجّح مدخلاتها وتجمعها، ويضبط التدريب الأوزان بناءً على البيانات. يقوم التعلم العميق بتكديس العديد من الطبقات المخفية لبناء السمات (حواف ← أجزاء ← كل)، بفضل البيانات الضخمة ومعالجات GPU.

⭐ فكّر وتحدَّ نفسك

تأمل: أين استخدمت أداة مبنية على التعلم العميق (الكتابة الصوتية، البحث عن الصور، الترجمة)؟ ما المجال الذي كانت تنتمي إليه؟

تحدٍّ: يقوم فريق بجعل شبكة أعمق فأعمق لكن الدقة تتوقف عن التحسن ويصبح التدريب بطيئًا جدًا. اذكر سببًا واحدًا يجعل زيادة العمق ليست دائمًا أفضل، وأمرًا واحدًا – غير إضافة المزيد من الطبقات – يجب عليهم التحقق منه أولًا.

nezakr.natega4dk.net



الدرس 7-3

نماذج اللغة الكبيرة (LLM) والذكاء الاصطناعي التوليدي

أهداف التعلم

- شرح كيف يولّد نموذج اللغة النص عن طريق التنبؤ بالكلمة التالية، ولماذا يؤدي هذا إلى الهلوسة.
- شرح ما الذي يجعل نموذج اللغة كبيرًا، موضحًا دور ضبط RLHF وتعدد الوسائط.

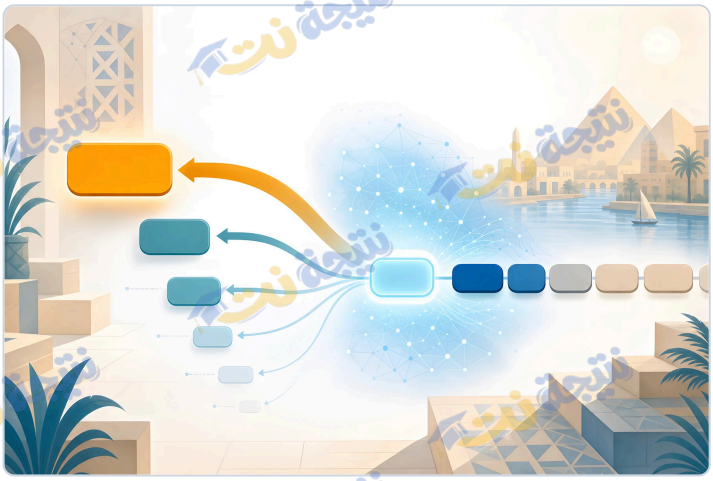
nezakr.natega4dk.net

عندما يكتب نظام محادثة فقرة، يولد النموذج اللغوي الأساسي عادة رمزًا نصيًا تاليًا (Token) ثم يكرر العملية؛ وقد يكون الرمز كلمة أو جزءًا منها أو علامة ترقيم. يحسب النموذج توزيع احتمالات تعلمه من بيانات التدريب وسياق الطلب، ثم تختار طريقة التوليد رمزًا وفق إعداداتها. ويمكن لأنظمة مبنية على نماذج لغوية تنفيذ مهام مثل التلخيص والترجمة والإجابة وتوليد البرمجيات، وقد تتصل بعض الأنظمة بأدوات أو مصادر خارجية. لكن هدف التنبؤ الاحتمالي لا يتضمن ضمانًا مدمجًا لصحة الادعاءات؛ لذلك قد ينتج النظام معلومة غير صحيحة أو مختلقة أو غير مدعومة بصياغة تبدو مقنعة، وهي ظاهرة تسمى الهلوسة. يشرح الدرس كذلك كيف قد تحسن طرق ضبط مثل RLHF اتباع التعليمات، من دون أن تضمن الحقيقة أو السلامة، وكيف تتعامل بعض الأنظمة مع أكثر من وسيط مثل النص والصورة والصوت بحسب تصميمها.

السؤال الموجِّه: كيف يولد نموذج لغوي النص، وماذا يعني «كبير» في هذا الدرس، ولماذا لا تضمن الطلاقة صحة المعلومات؟

ملحوظة مهمّة

1. نماذج اللغة والتنبؤ بالرمز التالي



يتنبأ النموذج اللغوي التوليدي بالرمز النصي التالي وفق السياق.

(1) **نموذج اللغة** نموذج يتعلم أنماط تسلسل الرموز في البيانات. ويمكن للنموذج التوليدي تقدير توزيع احتمال الرمز النصي التالي بالاعتماد على السياق؛ والرمز قد يكون كلمة أو جزء كلمة أو علامة ترقيم.

(2) يتحقق توليد النص بإضافة رمز بعد آخر وتحديث السياق في كل خطوة، حتى يصل النظام إلى حد التوقف أو الطول المحدد.

الخطوة	الجملة المُدخلة	الكلمة المتوقعة التالية
الأولى	"طقس اليوم"	"مشمس"
الثانية	"طقس اليوم مشمس"	"و"
الثالثة	"طقس اليوم مشمس و"	"دافئ"

بهذه الطريقة يبنى النص رمزًا بعد آخر. قد يطابق الرمز كلمة كاملة، أو جزء كلمة، أو علامة ترقيم وفق طريقة تقسيم النص المستخدمة.

⭐ خريطة الدرس

الفكرة الأساسية:

يولد النموذج اللغوي التوليدي النص عادة بإضافة رمز (Token) بعد آخر وفق سياقه. وقد تحسن النماذج الكبيرة أداء مهام متعددة، لكن الحجم لا يضمن الحقيقة. وقد تستخدم بعض الأنظمة RLHF أو طرق ضبط أخرى، وقد تدعم أكثر من وسيط؛ وتختلف القدرات والحدود بحسب النظام.

المفاهيم الأساسية:

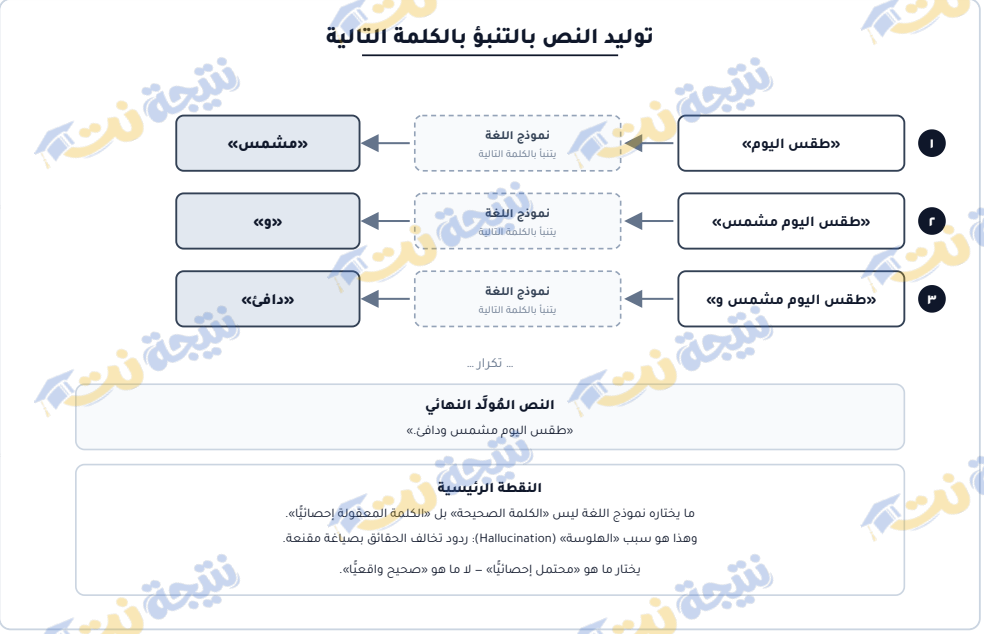
- نموذج اللغة، التنبؤ بالرمز التالي
- الهلوسة
- نموذج اللغة الكبير (LLM)، مَعلمات النموذج (Parameters)
- RLHF
- تعدد الوسائط

مصطلحات أساسية

نموذج اللغة – يتعلم احتمالات تسلسل الرموز، ويمكن للتوليدي توقع الرمز التالي.

الهلوسة – معلومة غير صحيحة أو مختلقة أو غير مدعومة قد تبدو مقنعة.

nezakr.natega4dk.net



الشكل 7.3.1 – يولد النص بإضافة رمز نصي بعد آخر وفق السياق.

(3) يحسب النموذج لكل رمز مرشح درجة أو احتمالًا مشروطًا بالسياق. ثم تختار خوارزمية التوليد الرمز التالي وفق إعدادات مثل أخذ العينات ودرجة الحرارة. ولا يكتفي النموذج بعدّ تكرار الكلمات؛ بل تتعلم الشبكة تمثيلات ومعلمات معقدة من سياقات كثيرة أثناء التدريب.

(4) ينتج النموذج رمزًا وفق توزيعه الاحتمالي، ولا يملك في هدف التنبؤ وحده ضمانًا لصحة الادعاءات. لذلك قد تظهر **الهلوسة**: معلومة غير صحيحة أو مختلقة أو غير مدعومة بالمصدر، مع صياغة قد تبدو مقنعة. وللهلوسة أسباب متعددة، فلا تنسب إلى عامل واحد فقط.

2. نماذج اللغة الكبيرة (LLM)

(1) **نموذج اللغة الكبير (LLM)** نموذج لغوي ذو عدد كبير من المعلمات ومدرب على بيانات واسعة قد تشمل نصوصًا وبرمجيات ومصادر متنوعة بحسب النموذج. وكثير من النماذج الحديثة مبني على بنية المحوِّل (Transformer). ولا يوجد حد رسمي واحد يجعل النموذج «كبيرًا».

(2) في هذا الدرس نصف «الكبر» بثلاثة جوانب مترابطة: عدد معلمات النموذج، وحجم بيانات أو رموز التدريب، والقدرة الحاسوبية المستخدمة. وتختلف هذه الجوانب كثيرًا من نموذج إلى آخر.

الجانب	الوصف
كمية بيانات التدريب	قد تكون بيانات التدريب ضخمة جدًا، ويقاس حجم النصوص الحديثة عادة بعدد الرموز (Tokens) لا الكلمات فقط
عدد مَعلمات النموذج	معلمات عددية تتعلم أثناء التدريب؛ يحتوي كثير من النماذج الكبيرة على مليارات منها، وتختلف الأحجام بين النماذج

(3) قد تتحسن قدرات متعددة مع زيادة مناسبة في حجم النموذج والبيانات والحوسبة وجودة التدريب، لكن لا يوجد حد ثابت يضمن ظهور قدرة بعينها. وقد يظل النموذج يخطئ في الاستدلال أو الحقائق حتى إذا نجح في التلخيص والترجمة وتوليد البرمجيات.

(4) توجد أنظمة محادثة عديدة مبنية على نماذج لغوية كبيرة، مثل ChatGPT Claudeg Geminig. وهي أسماء منتجات قد تتغير قدراتها وإصداراتها؛ لذلك نركز على المفهوم العام ونختبر كل نظام وحدوده قبل استخدامه.

3. الضبط عبر RLHF

(1) قد لا يتبع النموذج الأساسي التعليمات أو قصد المستخدم بصورة موثوقة، وقد ينتج في بعض الحالات مخرجات غير مناسبة أو غير آمنة. لذلك تستخدم مراحل ضبط لاحقة إلى جانب تصميم النظام واختباره.

مصطلحات أساسية

نموذج اللغة الكبير (LLM) – شبكة ضخمة مدرَّبة على نصوص هائلة.

المعامل – قيمة وزن؛ تحتوي نماذج اللغة الكبيرة على مليارات منها.

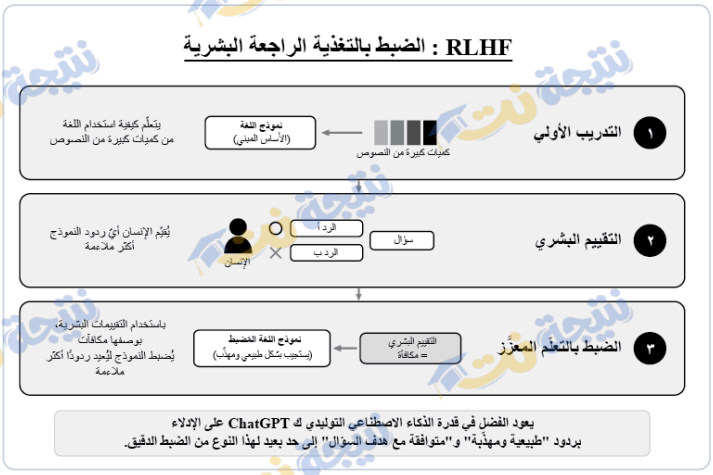
RLHF – التعلم المعزز من خلال التغذية الراجعة البشرية: الضبط بجعل التقييم البشري مكافأة.

nezakr.natega4dk.net

(2) **التعلم المعزز من خلال التغذية الراجعة البشرية (RLHF - Reinforcement Learning from Human Feedback)** إحدى طرق الضبط، وليست الوحيدة. في مسارها التقليدي تستخدم تفضيلات بشرية لتدريب نموذج مكافأة، ثم يستخدم التعلم المعزز هذه الإشارة لتحسين سلوك النموذج.

(3) تسير RLHF وفق التدفق التالي.

المرحلة	الوصف
① **التدريب الأولي**	بناء نموذج لغة تعلّم استخدام الكلمات من كميات كبيرة من النصوص
② **التقييم البشري**	يقارن مقيّمون مخرجات أو يصنفونها لتكوين بيانات تفضيل وفق معايير محددة
③ **الضبط عبر التعلم المعزز**	يدرَّب في المسار التقليدي نموذج مكافأة من بيانات التفضيل، ثم تستخدم إشارته في التعلم المعزز لضبط النموذج



الشكل 7.3.2 – RLHF: تصبح التقييمات البشرية مكافآت تضبط النموذج

(4) قد يحسن RLHF اتباع التعليمات والأسلوب وبعض جوانب السلامة، لكنه لا يضمن الحقيقة أو الأمان أو ملاءمة كل مخرج. وتعتمد جودة المساعد أيضًا على التدريب المسبق وطرق الضبط الأخرى وتصميم النظام والأدوات والتقييم المستمر.

4. تعدد الوسائط



قد يستقبل النظام متعدد الوسائط أو ينتج أكثر من نوع من البيانات، بحسب تصميمه.

(1) **تعدد الوسائط**: قدرة نموذج أو نظام على معالجة أكثر من وسيط، مثل النص والصورة أو الصوت. وقد يتعلق ذلك بالمدخلات أو المخرجات أو كليهما، ولا يعني أن كل نظام يدعم جميع الأنواع.

(2) كانت نماذج لغوية كثيرة نصية، ثم ظهرت نماذج وأنظمة تربط النص بوسائط أخرى. وقد تدعم بعض الأنظمة وصف صورة بالنص، أو تحويل صوت إلى نص، أو توليد صورة من

توقّف وفكّر

عند تركه بدون ضبط، قد يعطي النموذج الضخم إجابة غير مناسبة أو خارج الموضوع. ما الخطوة التي تصحح ذلك؟



التغذية الراجعة البشرية تصنّف المخرجات لضبط النموذج.

مصطلحات أساسية

تعدد الوسائط – نموذج واحد يتعامل مع أنواع متعددة من البيانات.

الوسيط – نوع واحد من البيانات (نص، صورة، صوت...).

nezakr.natega4dk.net

وصف، أو تحليل فيديو؛ وتختلف القدرات المتاحة بحسب النظام.

(3) تمثل البيانات داخل الحاسوب بقيم عددية، لكن طريقة ترميز النص والصورة والصوت ومعالجتها قد تختلف. تستخدم الأنظمة متعددة الوسائط بنى تربط هذه التمثيلات أو تدمجها، وقد تكون نموذجًا واحدًا أو نظامًا يجمع مكونات متخصصة.

(4) يوسع تعدد الوسائط بعض تطبيقات الذكاء الاصطناعي التوليدي عبر الربط بين أنواع من البيانات، لكنه لا يلغي الحاجة إلى فحص الدقة والخصوصية وحقوق الاستخدام والتحيز والسلامة في كل وسيط.

الفكرة الرئيسة

يولد النموذج اللغوي التوليدي رمزًا بعد آخر وفق توزيع احتمالي، وهذا لا يضمن صحة الحقائق. يصف «الكبر» جوانب مثل عدد المعلمات وحجم التدريب والحوسبة، من دون حد ثابت يضمن قدرة بعينها. وقد يحسن RLHF اتباع التعليمات وبعض جوانب السلامة، لكنه لا يضمن الصحة أو الأمان. أما النظام متعدد الوسائط فقد يعالج أكثر من وسيط بحسب تصميمه. لذلك: الطلاقة لا تساوي الحقيقة، والتحقق ضروري.

توقّف وفكّر

إذا بدا النموذج واثقًا وطليقًا، فهل هذا يضمن أنه صحيح؟

مثال محلول

(1) ضع علامة على كل عبارة (أ - د) بـ○ إذا كانت صحيحة أو × إذا كانت خاطئة.

أ نموذج اللغة هو نموذج يتنبأ بالكلمة المرجّح أن تأتي بعد ذلك في الجملة.

ب ينتهي نموذج اللغة من المعالجة بعد التنبؤ بكلمة واحدة فقط ولا يقوم بتوليد النص.

جـ قد تتحسن بعض القدرات مع زيادة مناسبة في حجم النموذج والبيانات والحوسبة، لكن لا يوجد حد ثابت يضمن قدرة جديدة.

د تعدد الوسائط هو خاصية التعامل مع النص فقط.

(2) طابق كل وصف (1 - 3) مع المصطلح الأنسب من الخيارات أدناه (أ - جـ).

1 نموذج لغوي ذو عدد كبير من المعلمات ومدرب على بيانات واسعة من مصادر متنوعة

2 طريقة ضبط تستخدم تفضيلات بشرية لبناء إشارة مكافأة ثم تحسن سلوك النموذج بالتعلم المعزز

3 إنتاج معلومة غير صحيحة أو مختلقة أو غير مدعومة بصياغة قد تبدو مقنعة

[الخيارات] أ الهلوسة ب نموذج اللغة الكبير جـ RLHF

الحلّ

(1)

أ تعريف صحيح لنموذج اللغة. لذلك، ○.

ب يولد النموذج النص عادة بإضافة رمز نصي بعد آخر وفق السياق. لذلك، ×.

جـ وصف صحيح لنماذج اللغة الكبيرة. لذلك، ○.

د تعدد الوسائط قدرة نموذج أو نظام على معالجة أكثر من وسيط، وقد تتعلق بالمدخلات أو المخرجات أو كليهما. لذلك، ×.

(2)

1: نموذج لغوي ذو معلمات كثيرة وتدريب واسع هو نموذج لغة كبير، من دون حد رسمي واحد لكلمة «كبير». ب

2: الضبط الذي يستخدم تفضيلات بشرية وإشارة مكافأة مع التعلم المعزز هو RLHF. جـ

3: إنتاج معلومات غير صحيحة أو مختلقة أو غير مدعومة بصورة قد تبدو مقنعة هو الهلوسة. أ

nezakr.natega4dk.net

تدرّب

1 أجب عن الأسئلة التالية.

1. ما مصطلح النموذج الذي يتعلم أنماط تسلسل الرموز ويمكن استخدامه لتقدير احتمال الرمز التالي؟
2. ما مصطلح النموذج اللغوي ذي العدد الكبير من المعلمات والمدرب على بيانات واسعة؟
3. ما اسم طريقة الضبط التي تستخدم تفضيلات بشرية وإشارة مكافأة مع التعلم المعزز؟
4. ما مصطلح قدرة نموذج أو نظام على معالجة أكثر من وسيط، مثل النص والصورة أو الصوت؟

2 أجب عن الأسئلة التالية.

(1) من بين الخيارات التالية (أ - د)، اختر الخيار الذي يصف كيفية عمل نموذج اللغة بأنسب شكل.

أ يحفظ الجملة المُدخلة كما هي ويعيد نفس الجملة.
ب يولد النص بإضافة رمز نصي بعد آخر وفق توزيع احتمالي مشروط بالسياق.
ج يختار ويعيد أقرب تطابق من جمل قالبية مُعدّة مسبقًا.
د يفهم معنى الجملة فهمًا كاملًا ويعيد الإجابة الصحيحة الوحيدة.

(2) من بين الخيارات التالية (أ - د)، اختر الخيار الذي يصف سبب الهلوسة بأنسب شكل.

أ لأن نموذج اللغة يسترجع ويعيد دائمًا جملًا من بيانات التدريب كما هي.
ب لأن نموذج اللغة يختار ويعيد كلمات محتملة إحصائيًا وليست بالضرورة صحيحة.
ج لأن نموذج اللغة يتجنب الإجابة عن الأسئلة ويعيد نصًا غير ذي صلة.
د لأن نموذج اللغة مصمم ليعيد إجابات قائمة على الحقائق فقط.

(3) من بين الخيارات التالية (أ - د)، اختر الخيار الذي يصف RLHF بأنسب شكل.

أ طريقة لبناء نموذج لغة من الصفر باستخدام كمية كبيرة من النصوص.
ب طريقة ضبط تستخدم تفضيلات بشرية وإشارة مكافأة مع التعلم المعزز لتحسين اتباع التعليمات.
ج طريقة تزيل بنية الشبكة العصبية من نموذج اللغة وتبسّط المعالجة.
د طريقة تجعل نموذج اللغة يحل مسائل رياضية لتحسين دقته الحسابية.

(4) من بين الخيارات التالية (أ - د)، اختر الخيار الذي لا يُعد تطبيقًا مناسبًا للذكاء الاصطناعي متعدد الوسائط.

أ عرض صورة وجعلها تُوصف بالنص.
ب الاستماع إلى صوت وتحويله إلى نص.
ج إعطاء تعليمات نصية وتوليد صورة.
د حساب نتيجة قيم عددية مُدخلة باستخدام قواعد العمليات الحسابية الأربع فقط وإعادتها.

📝 سؤال على نمط الامتحان

اشرح لماذا يمكن لنموذج لغوي أن يعطي إجابة سلسة لكنها خاطئة من حيث الحقائق (هلوسة). **[6]**

ارجع إلى: التنبؤ الاحتمالي بالرمز التالي وعدم وجود تحقق مدمج من الحقائق.

💡 توقّف وفكّر

هل يختار النموذج فعليًا ما هو «معقول» أم ما هو «صحيح»؟

nezakr.natega4dk.net

مصطلحات أساسية

الذكاء الاصطناعي التوليدي – ذكاء اصطناعي يُنشئ نصًا أو صورًا أو صوتًا جديدًا.

التحقق من الحقائق – التحقق من صحة ادعاء ما بمقارنته بمصدر موثوق.

⚙️ فكّر كمهندس: ابحث ثم قرر

يريد نادٍ مدرسي استخدام روبوت محادثة قائم على الذكاء الاصطناعي التوليدي للمساعدة في كتابة مقالات قصيرة عن التاريخ المحلي لموقع المدرسة الإلكتروني. أنت تقدّم المشورة حول استخدامه بمسؤولية.

1 · **حدد أين يفيد.** اذكر جزءًا واحدًا من المهمة يوفر فيه روبوت المحادثة الجهد فعليًا، وجزءًا واحدًا لا يجب فيه الوثوق بمخرجاته كما هي. اشرح كلًا منهما في سطر واحد.

2 · **حلّل المخاطر.** باستخدام طريقة اختيار نموذج اللغة للكلمات، اشرح لماذا قد يذكر روبوت المحادثة تاريخًا خاطئًا أو اسمًا مُختلقًا لمعلَم محلي بثقة كبيرة. وصف بالتحديد كيف سيكتشف النادي مثل هذه الهلوسة قبل النشر – وهو تحقق لا يوضحه الدرس لك.

3 · **حدد الضمانات.** اذكر القاعدة التي يجب على النادي اتباعها لكل مقال مكتوب بواسطة الذكاء الاصطناعي قبل نشره، وقل ما الذي يحله الضبط على طراز RLHF وما لا يحله هنا. **في أزواج،** قارنوا القواعد واتفقوا على التحقق الواحد الذي لن تتخطوه أبدًا.

☐ **نشر كل ما يكتبه الذكاء الاصطناعي**

☐ **استخدام الذكاء الاصطناعي، مع التحقق قبل النشر**

قدّم دليلًا على إجابتك.

هل تواجه صعوبة؟ تذكّر أن التنبؤ الاحتمالي لا يوفّر تحققًا مدمجًا من كل حقيقة. راجع التواريخ والأسماء والأماكن في مصادر موثوقة ومستقلة، وسجّل المصدر، واجعل شخصًا مسؤولًا يراجع النص قبل النشر. وقد تساعد أدوات الاسترجاع، لكنها لا تلغي التحقق.

🌍 طبّق ما تعلمته

يرفع طالب صورة لملاحظة مكتوبة بخط اليد ويطلب من مساعد قراءتها بصوت عالٍ ثم تلخيصها كتابة. اذكر الخاصية التي تتيح للنظام الربط بين الصورة والنص والصوت، واشرح أن كل وسيط يرمز عدديًا بطريقة مناسبة ثم تربط المكونات بين التمثيلات. وضح أيضًا خطرًا متعلقًا بخصوصية الصورة أو خطأ القراءة.

❓ إجابة السؤال الرئيس

شرح الدرس أن النموذج اللغوي التوليدي يضيف رمزًا نصيًا بعد آخر وفق توزيع احتمالي مشروط بالسياق، وأن الرمز قد يكون كلمة أو جزء كلمة أو علامة. يصف «الكبر» هنا عدد المعلمات وحجم التدريب والحوسبة، من دون حد رسمي يضمن قدرات بعينها. وقد تنتج الهلوسة حين يقدم النظام معلومة غير صحيحة أو مختلقة أو غير مدعومة بصياغة مقنعة؛ فالطلاقة لا تضمن الحقيقة. RLHF إحدى طرق الضبط التي تستخدم تفضيلات بشرية وإشارة مكافأة لتحسين سلوك النموذج، لكنها لا تضمن الصحة أو الأمان. أما تعدد الوسائط فهو قدرة نموذج أو نظام على معالجة أكثر من وسيط، وتختلف الأنواع والمدخلات والمخرجات التي يدعمها كل نظام. لذلك يجب التحقق من الادعاءات المهمة وحماية الخصوصية وحقوق الاستخدام قبل النشر أو اتخاذ قرار.

nezakr.natega4dk.net

تمارين

1 اختبر فهمك: أجب عن الأسئلة الآتية دون الرجوع إلى قسم «ملحوظة مهمة»، ثم راجع إجاباتك.

2 اقرأ المقطع التالي وأجب عن كل سؤال.

يسمى النموذج الذي يتعلم أنماط تسلسل الرموز (1). يولد النص بإضافة (2) بعد آخر وفق السياق. يسمى النموذج اللغوي ذو المعلمات الكثيرة والتدريب الواسع (3). ويعد عدد (4) وحجم بيانات التدريب والحوسبة من جوانب حجمه. وتعد (5) إحدى طرق الضبط التي تستخدم تفضيلات بشرية وإشارة مكافأة مع التعلم المعزز. أما قدرة نموذج أو نظام على معالجة أكثر من نوع من البيانات فتسمى (6).

1. املأ الفراغات (1) - (6).

3 أجب عن الأسئلة التالية.

(1) من بين الخيارات التالية (أ - د)، اختر الخيار الذي يصف آلية توليد النص بأنسب شكل.

أ يتم توليد النص عادة بإضافة رمز نصي بعد آخر وفق السياق.

ب يتم توليد النص من خلال الاختيار من بين آلاف القوالب المُعدّة مسبقًا.

ج يتم توليد النص من خلال تحديد طول الجملة أولًا ثم تحديد جميع الكلمات في وقت واحد.

د يتم توليد النص من خلال الترجمة الآلية للجملة المُدخلة.

(2) من بين الخيارات التالية (أ - د)، اختر الخيار الذي يصف خلفية اكتساب نماذج اللغة الكبيرة قدرات جديدة بأنسب شكل.

أ لأن تقليل عدد مَعلمات النموذج وحده يضمن ظهور قدرات جديدة.

ب لأن قصر بيانات التدريب على كمية صغيرة من النصوص عالية الجودة رفع الجودة.

ج لأن بعض القدرات قد تتحسن مع الحجم والتدريب المناسبين، من دون حد ثابت يضمن ظهورها.

د لأن البشر كتبوا قواعد اللغة واحدة تلو الأخرى وعلّموها للنموذج.

(3) من بين الخيارات التالية (أ - د)، اختر الخيار الذي يصف الغرض من إجراء RLHF بأنسب شكل.

أ لزيادة عدد الكلمات التي يمكن لنموذج اللغة التعامل معها.

ب لجعل إجابات نموذج اللغة طبيعية ومتوافقة مع قصد السؤال.

ج لتمكين نموذج اللغة من التعامل مع بيانات غير النص.

د لتقليل عدد مَعلمات النموذج بوصفه هدفًا لازمًا لعملية الضبط.

⭐ الخلاصة

تذكّر:
يولد النموذج اللغوي التوليدي النص عادة بإضافة رمز بعد آخر وفق سياقه، ولا يضمن ذلك صحة الحقائق. يعتمد أداء النماذج الكبيرة على المعلمات والبيانات والحوسبة وجودة التدريب. وقد تستخدم طرق ضبط مثل RLHF. أما الأنظمة متعددة الوسائط فتعالج أكثر من وسيط بحسب تصميمها. الطلاقة لا تساوي الحقيقة: تحقق من الادعاءات المهمة في مصادر موثوقة.

⭐ فكّر وتحدَّ نفسك

تأمل: عندما استخدمت أداة ذكاء اصطناعي توليدي آخر مرة، هل تحققت من حقائقها؟ ما الذي كان يمكن أن تكلفه إجابة خاطئة واثقة؟

تحدٍّ: يقول أحد زملائك في الصف "إن RLHF يجعل الذكاء الاصطناعي يقول الحقيقة دائمًا." اشرح لماذا هذا غير صحيح، واذكر ما الذي يحسّنه RLHF فعليًا.

nezakr.natega4dk.net

وزارة التربية والتعليم والتعليم الفني

فريق الإعداد والاعتماد

قائد المشروع والمدير التحريري

Masaki Iisaka

المحررون والمؤلفون

Biichi Nakajima
Hiromi Fujii

المراجعون

أ.د محمد سيد فرج
أ.د إيمان سراج الدين بكر
د.عبير حامد احمد
د. طاهر عبد الحميد العدلي
د. منال زيادة عبد الفضيل

الإدارة المركزية لتطوير المناهج

الإشراف العام

د. جبريل أنور حميدة

رئيس الإدارة المركزية لتطوير المناهج

جميع الحقوق محفوظة لوزارة التربية والتعليم والتعليم الفني - 2026 / 2027

لا يجوز إعادة إنتاج أي جزء من هذا الكتاب أو تخزينه في نظام لاسترجاع المعلومات أو نقله بأي صورة أو وسيلة، إلكترونية أو بالتصوير أو التسجيل أو غير ذلك، من دون إذن كتابي مسبق من وزارة التربية والتعليم والتعليم الفني.

nezakr.natega4dk.net

EDUCATION SPRINGS NEW LIFE

SPRIX

CACD

class | name

nezakr.natega4dk.net