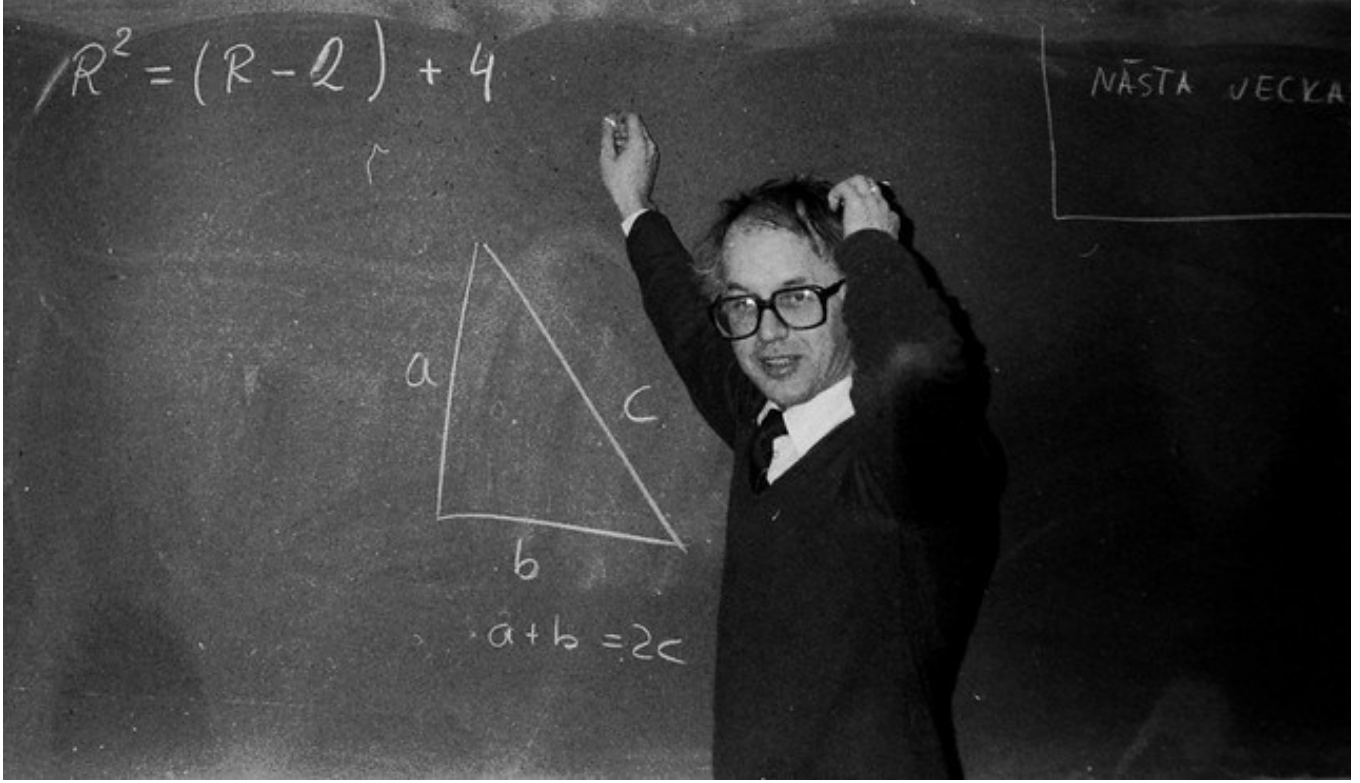


# أساسيات الرياضيات لعلم البيانات: "لماذا" و "كيف"

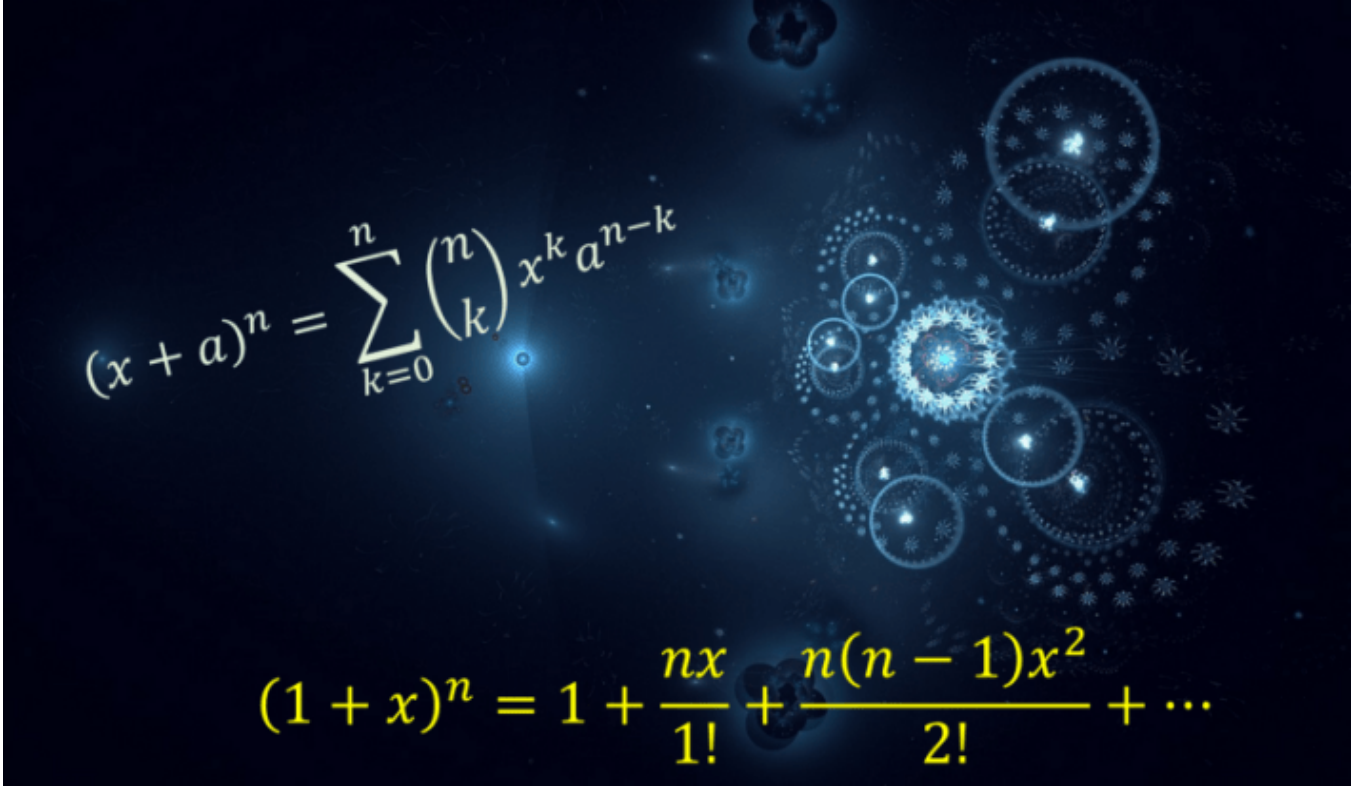
[Haya Aldawood](#)



كتابة: Tirthajyoti Sarkar ترجمة ومراجعة: يحيى خوجه وهيا الداود

[المقال الأصلي بالإنجليزي [اضغط هنا](#)]

الرياضيات هي حجر الأساس لأي مادة علمية معاصرة. لذلك أمر غير مستغرب أن تكون جميع أساليب علم البيانات الحديثة (و التي تشمل تعلم الآلة) مبنية على مبادئ رياضية عميقة. في هذا المقال سنتحدث عن مواضيع الرياضيات الأساسية لتصبح عالم بيانات أفضل من جميع النواحي.



## مقدمة

الرياضيات هي حجر الأساس لأي مادة علمية معاصرة. لذلك، أمر غير مستغرب أن تكون جميع أساليب علم البيانات الحديثة (و التي تشمل تعلم الآلة) مبنية على مبادئ رياضية عميقة.

أحياناً كعالم بيانات (أو حتى كمحلل بيانات في فريق) يجب عليك تعلم أساسيات رياضية عن ظهر قلب لاستعمال أو تطبيق التقنيات بشكل جيد. أحياناً أخرى تستطيع تجاوز ذلك باستعمال API أو الخوارزميات الجاهزة.

على أي حال، حصولك على فهم قوي لمبادئ الرياضيات في الخوارزميات التي تستعملها لبناء منصة اقتراحات شرائية لعملاءك مثلاً لن يضررك أبداً. في معظم الأحيان سيجعلك هذا متقدماً عن زملائك وسيعطيك ثقة أكبر. معرفة كيف يعمل محرك السيارة (حتى وإن كان بشكل سطحي) يجعلك أفضل ممن يقودها دون أي علم بكيفية عملها.

كما أنه أمر مفروغ منه أنك تحتاج قدر كبير من العلم والقدرة على البرمجة والبعض من المعرفة بإدارة الأعمال وعقليتك المتميزة لتحليل واستكشاف البيانات لتكون من أفضل عالمي البيانات. كل ما أحاول فعله من خلال هذا المقال هو تقديم مؤشرات لأهم مهارات الرياضيات التي ستساعدك في الوصول لذلك المستوى.

## نقطة مهمة للمبتدئين

معرفة أساسيات الرياضيات تعتبر ذات أهمية عالية بالذات لمن يحاول دخول مجال علم البيانات بعد قضاء وقت

طويل في مجال آخر مثل الهندسة أو البيع بالتجزئة أو صناعة الكيماويات أو الطب والقطاع الصحي أو إدارة الأعمال أو غيرها.

قد تعتقد أنك عملت بما فيه الكفاية على جداول البيانات والحسابات الكمية والتنبؤية في عملك الحالي، لكن الطلب على مهارات الرياضيات الأساسية يختلف بشكل كبير في تطبيقات علم البيانات.

## لماذا وكيف تختلف - السبب في المنهجية العلمية وليست البيانات

خذ على سبيل المثال مطور ويب (أو محلل أعمال). ربما يتعاملون مع الكثير من البيانات والمعلومات بشكل يومي، لكن ذلك قد لا يتطلب تطوير أي نماذج دقيقة لمعالجة هذه البيانات. أغلب الأحيان يكون هناك ضغط زمني كبير ويكون التركيز على "استعمال البيانات لاحتياجاتك الحالية" دون التعمق في خواص البيانات نفسها بمنهجية علمية. شئت أم أبيت، يجب أن يكون تركيز عالم البيانات على المنهجية العلمية (وليس البيانات)، وعلى هذا السياق تصبح بعض الأدوات والتقنيات جوهريّة. معظمها تشكل أساسيات المنهجية العلمية السليمة:

- بناء نموذج رياضي لآلية (لموسسة أو معلوماتية) من خلال فهم الديناميكيات التي تشكلها
- بناء الفرضيات
- تنبؤ جودة مصدر البيانات بتمعن
- تحديد مدى الشك في البيانات والتنبؤات
- تدريب حدس الشخص على تحديد الأنماط المخفية في مجموعة البيانات
- فهم حدود ونقاط ضعف النموذج الرياضي
- (أحياناً) فهم إثبات رياضي والمنطق المجرد الذي خلفه

هذا النوع من التدريب، أو معظمه - القدرة على التفكير، ليس فقط عن طريق أرقام جافة بل عن طريق مبادئ رياضية مجردة (وخواصها والعلاقات التي تربطها ببعض)، جزء من أي برنامج جامعي للتخصصات العلمية. لا يحتاج أن يكون الشخص الأول على الدفعة من جامعة عليا حتى يتعرض لهذه المبادئ ولكن للأسف تعرضنا لهذه المبادئ غالباً يتوقف بعد الجامعة ولا يكتمل نموه ليرسخ في عقليتنا، ولا أتحدث عن مادة التفاضل والتكامل التي قد درستها في سنتك الأولى بالجامعة. أنا أتحدث عن ما هو أبسط من ذلك ... مثل الرقم 2 ...

“افترض أنك تجلس على مكتبك في الصباح - مستعد لمواجهة تحديات العمل المعقدة. فجأة يصلك بريد إلكتروني من مديرك (أو صديقك المحب للرياضيات) بهذا التحدي - أوجد الإثبات الرياضي أن الجذر التربيعي للرقم 2 عدد غير نسبي في أقل من دقيقتين“

هذه هي الفكرة ...

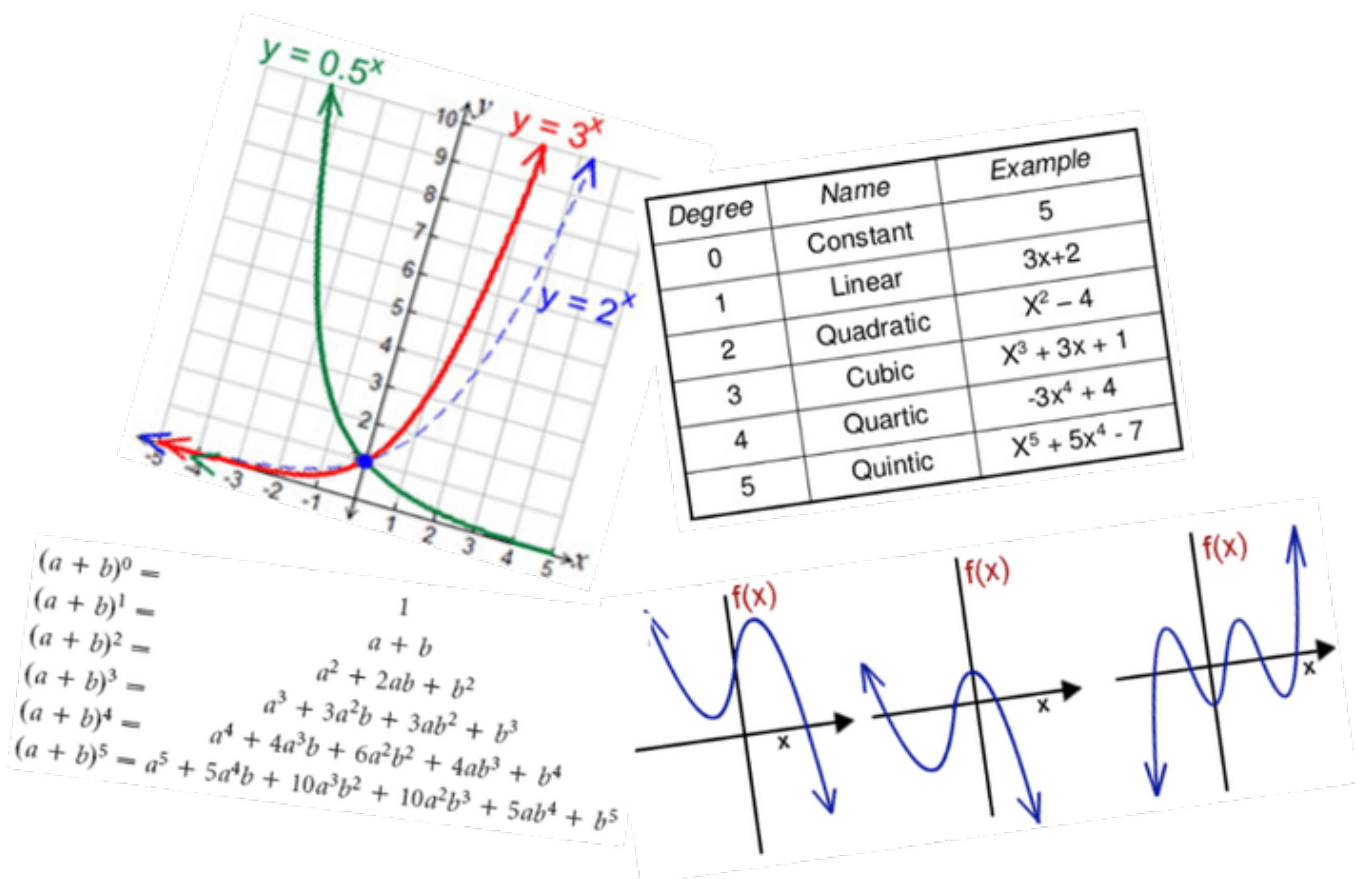
## كفانا حديثاً - أريد أن أرى المخطط الأساسي للنجاح

هذه مشكلة. لا يوجد مخطط معروف. علم البيانات بطبيعته غير مرتبط بأي موضوع محدد فقد يتعامل مع ظواهر متنوعة مثل تشخيص مرض السرطان وتحليل التعاملات الاجتماعية في نفس المشروع. هذا يؤدي إلى تقاطع تشكيلة معقدة ذات أبعاد كثيرة من العوامل الرياضية و التوزيعات الإحصائية ودوال الهدف التحسيني وغيرها. توقف رجاءاً.

ماهي هذه الأشياء المذكورة أعلاه؟ بدقة وبشكل جاد.

هذه اقتراحتي المنتقاة للمواضيع التي نحتاج دراستها أو التشبع منها حتى نكون من الصفوة في مجال علم البيانات.

### دوال، متغيرات، معادلات، رسومات بيانية



ماذا؟ سنبدأ من أهم المبادئ مثل معادلة الخط المستقيم لنظرية المعادلات ذات الحدين وخواصها:

- اللوغاريتمات، الأس، دوال كثيرات الحدود، الأرقام النسبية
- أساسيات علم الهندسة ونظرياتها، خواص الدوال المثلثية

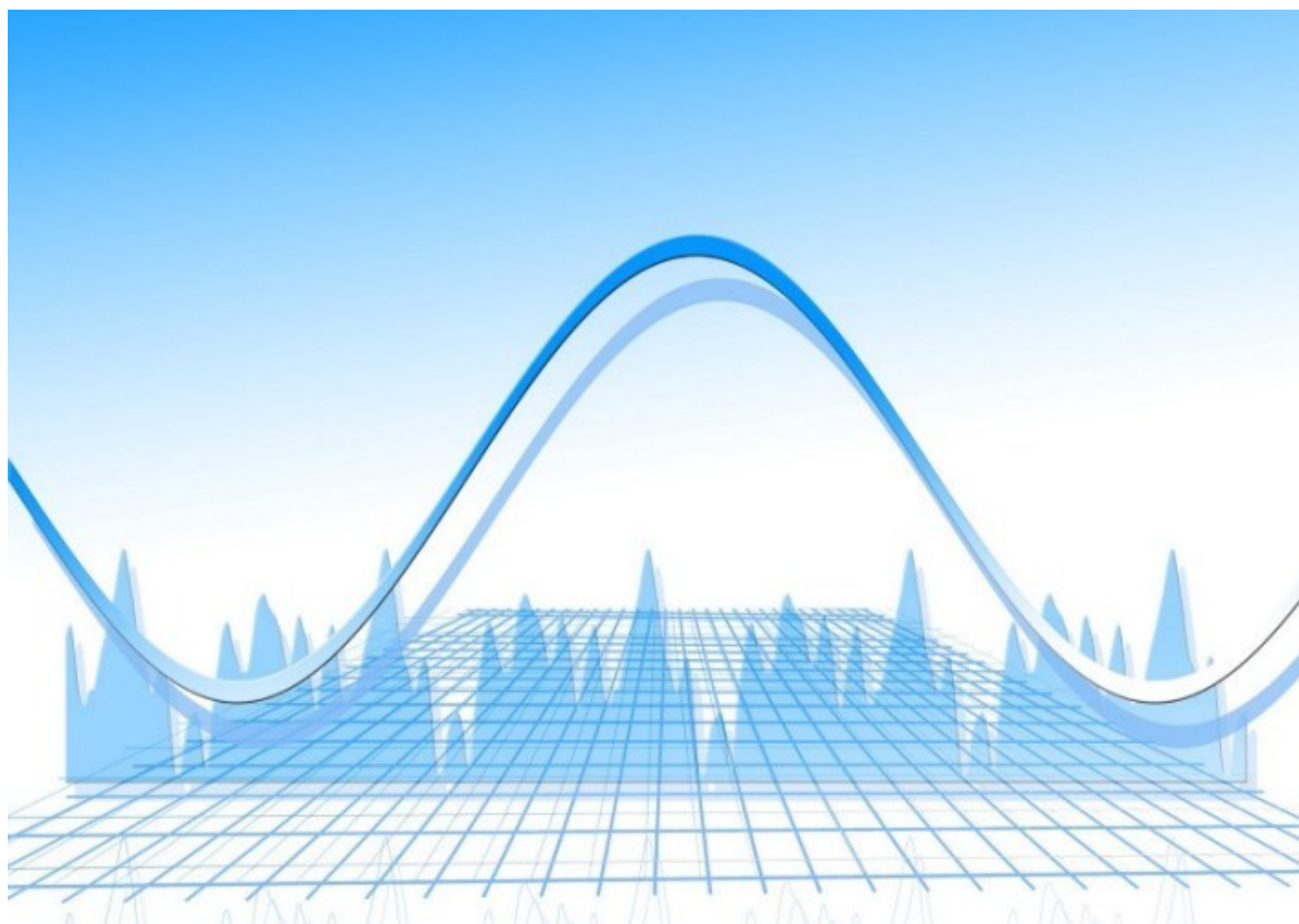
- الأرقام الحقيقية والمركبة وخواصها الأساسية
- المتسلسلات والجمع والمتباينات
- الرسم البياني، الإحداثيات الكارتيزية والقطبية، المقطوعات المخروطية

**مثال (أو اثنين) لكيفية استعمالها:** إذا أردت أن تفهم كيف يمكن أن تسرع عملية البحث في قاعدة بيانات تحتوي على مليون وحدة بعد أن تم فرزها، ستتعامل مع مبدأ البحث الثنائي (binary search). حتى تفهم خواصه، يجب أن تفهم اللوغاريتمات والمعادلات التكرارية. أو إذا أردت تحليل متسلسلة زمنية فقد تواجه مبادئ مثل الدوال الدورية والتضائل الأسي.

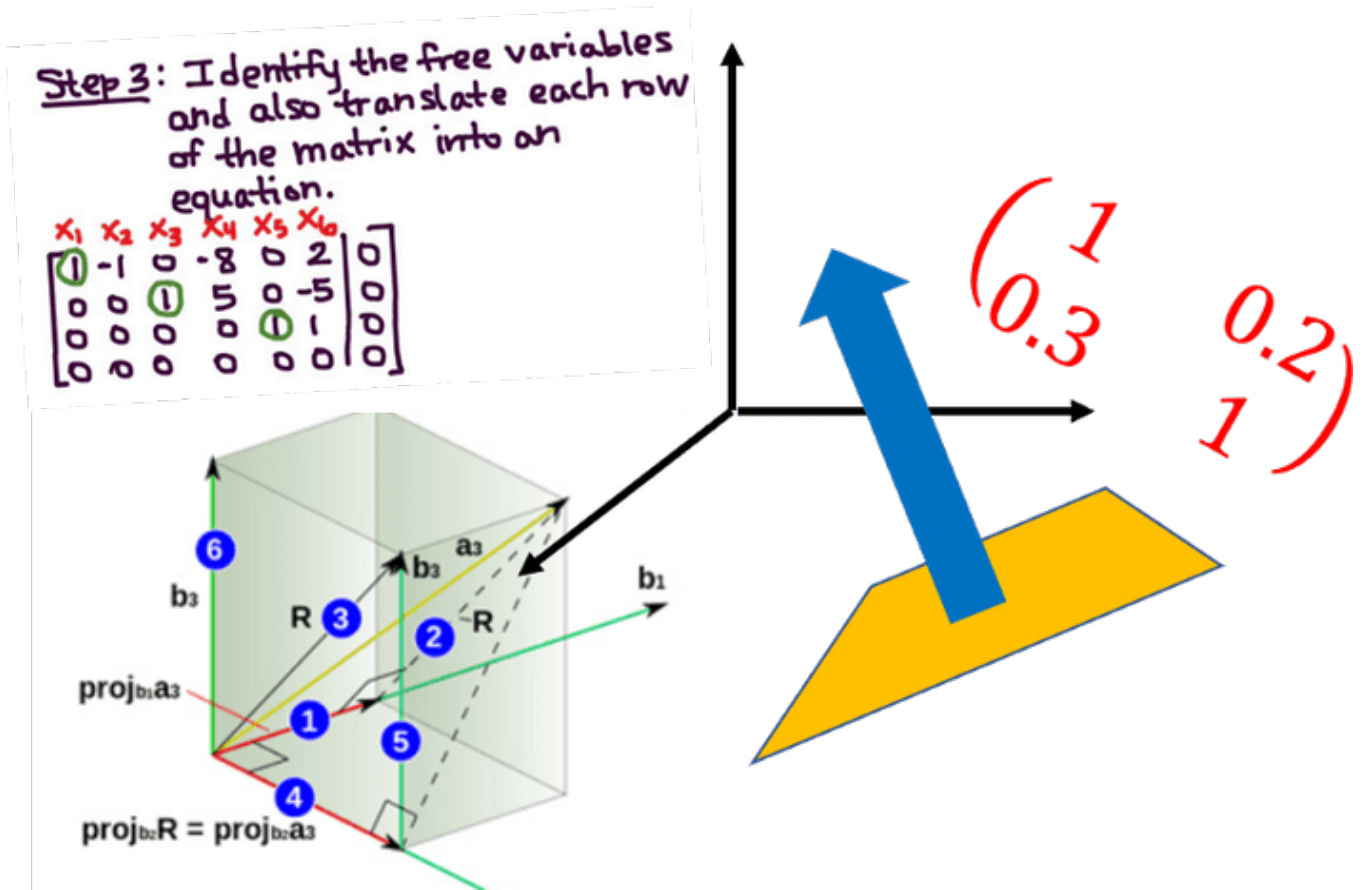
**أين تتعلم:**

- [مهارات الرياضيات لعلم البيانات - Coursera](#)
- [مقدمة في الجبر - edX](#)
- [الجبر - Khan Academy](#)

## الإحصاء







ماذا؟! اقتراح إضافة صديق على Facebook. اقتراح فيديو جديد على Youtube. تحويل صورتك الشخصية لرسم شخصية من الفنان سلفادور دالي باستعمال التعلم العميق. ماهو العامل المشترك؟ المصفوفات والجبر الخطي. هذا فرع أساسي من الرياضيات لفهم كيف تعمل معظم خوارزميات تعلم الآلة لإيجاد شيء جديد في مجموعة بيانات. هذه أهم المواضيع

- الخواص الأساسية للمصفوفات والمتجهات - الضرب العددي، التحول الخطي، المنقول (transpose)، الرتبة (rank)، المحدد (determinant)، المرافق (conjugate)
- الضرب الداخلي والخارجي، قانون ضرب المصفوفات، معكوس المصفوفة
- المصفوفات الخاصة - المصفوفة المربعة، مصفوفة الوحدة، المصفوفة المثلثة، المصفوفات المتناثرة والكثيفة، المتجهات الأحادية، المصفوفات المتماثلة، المصفوفات الهرماتية
- مفهوم تحليل المصفوفات الى عوامل باستخدام التفكيك العلوي والسفلي، الإلغاء الجاوسي وجاوس-جوردان، حل نظام معادلات خطي  $Ax = b$
- فضاء المتجهات، قاعدة المصفوفات، مدى المصفوفات، تعامد المصفوفات، التعامد الطبيعي، طريقة المربعات الصغرى
- قيم آيقن، متجهات آيقن، تفكيك القيم الأحادية، استقطار المصفوفة

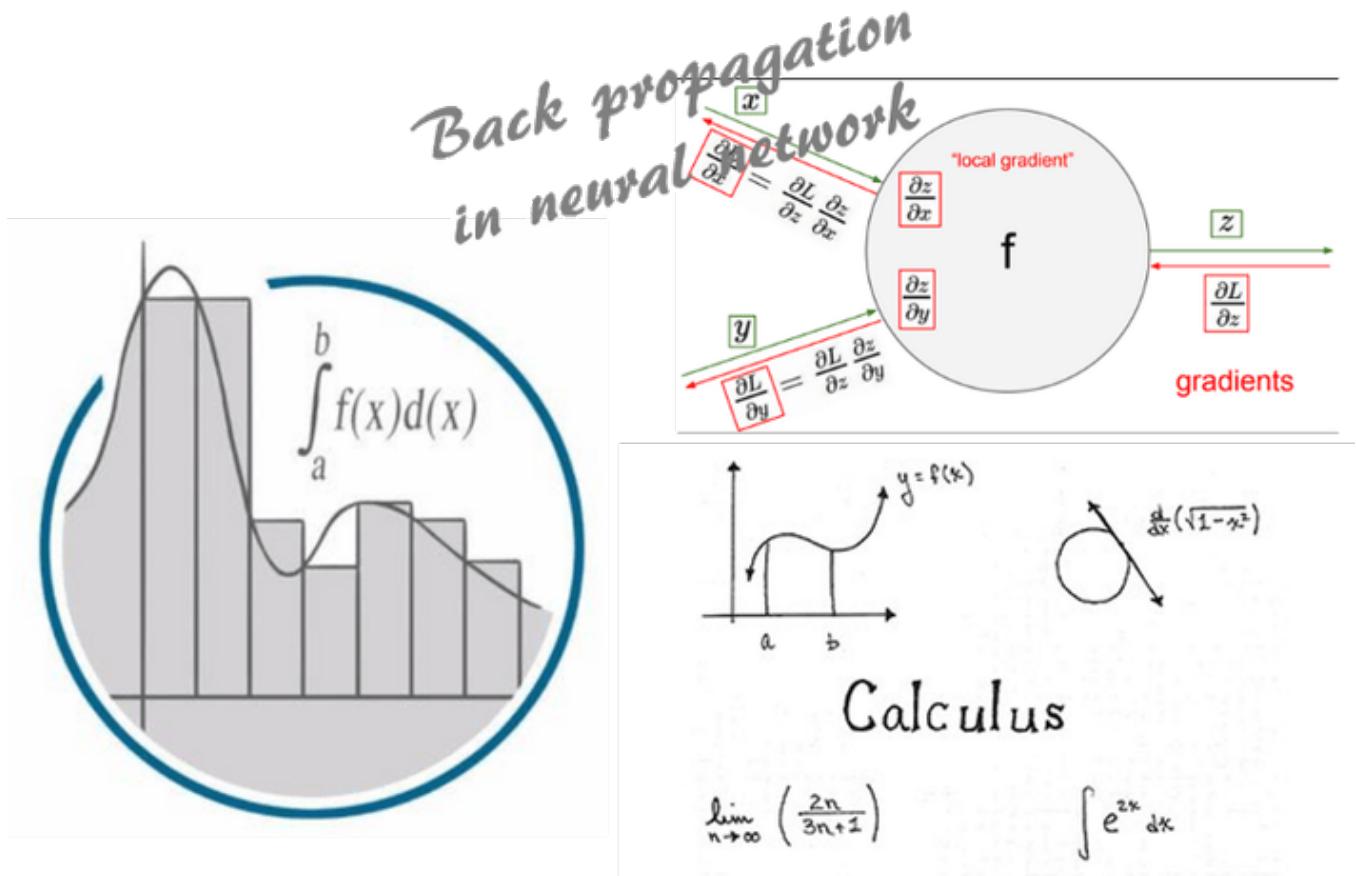
مثال (أو اثنين) لكيفية استعمالها: إذا كنت تستعمل طريقة لتقليص الأبعاد مثل تحليل المكونات الرئيسية

(PCA)، فإنك بالتأكيد تستعمل تفكيك القيم الأحادية للوصول إلى تمثيل لمجموعة البيانات بأبعاد وقيم أقل. كما أن جميع خوارزميات الشبكات العصبية تستعمل الجبر الخطي لتمثيل ومعالجة تركيبية الشبكة والقيام بعمليات التعلم.

**أين تتعلم:**

- [الجبر الخطي: من المقدمة إلى الأمام - edX](#)
- [الرياضيات لتعلم الآلة: الجبر الخطي - Coursera](#)

## علم التفاضل و التكامل



**ماذا: ها قد عاد!** سواء أحببته أم كرهته خلال أيام الجامعة، الحقيقة هي أن مفهوم وتطبيقات علم التفاضل والتكامل يعاود الظهور مراراً وتكراراً في مواضيع مختلفة في مجال علم البيانات و تعلم الآلة. انه متواجد خلف الحلول التحليلية البسيطة لمشكلة المربعات الصغرى العادية في الانحدار الخطي، و مضمن في كل انتشار عكسي في شبكتك العصبية مما سيجعلها تتعرف على نمط جديد. إنها مهارة في غاية الأهمية وإضافة مهمة لمخزونك المعرفي. هنا بعض مواضيع التعلم:



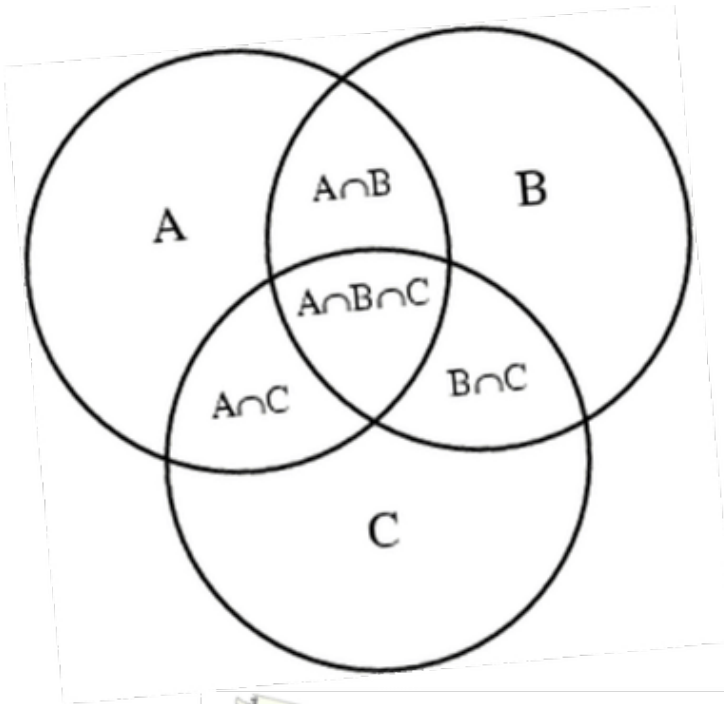
1. دالة المتغير الوحيد، النهاية (Limit)، الاستمرارية و التفاضل
2. نظرية القيمة المتوسطة، الصيغة غير المعينة وقاعدة لوبيتال
3. النقاط العظمى والصغرى
4. قاعدة الضرب وقاعدة السلسلة
5. متسلسلة تايلور، مفاهيم جمع و تكامل المتسلسلات اللانهائية
6. نظرية القيمة الأساسية و نظرية القيمة المتوسطة لحساب التكامل، التكامل المحدد والتكامل غير المحدد
7. دالتى غاما و بيتا
8. دوال القيم المتعددة، النهاية، الاستمرارية، المشتقات الجزئية
9. أساسيات المعادلة التفاضلية العادية و الجزئية (ليس متقدماً جداً)

**مثال (أو اثنين) كيفية استعمالها:** هل سبق و تساءلت كيف بالضبط يتم تطبيق خوارزمية الانحدار اللوجستي. هناك احتمال كبير أنه يستخدم طريقة تدعى "النزول الاشتقاقي" لإيجاد دالة الحد الأدنى من الخسارة. لكي تفهم كيف تعمل عليك أن تستخدم مفاهيم مختلفة من علم التفاضل والتكامل - الإنحدار، المشتقات، النهايات وقاعدة السلسلة.

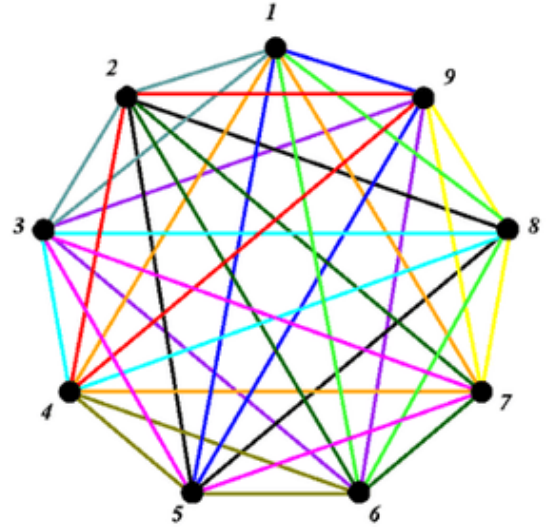
**أين تتعلم:**

- [Pre-University Calculus—edX, TU Delft](#)
- [Khan Academy Calculus all content](#)
- [Mathematics for Machine Learning: Multivariable Calculus—](#)  
Coursera, Imperial College, London

**الرياضيات المتقطعة**



$$\begin{aligned}
 {}_{105}C_3 &= \binom{105}{3} = \frac{105!}{(105-3)!3!} \\
 &= \frac{105!}{102!3!} = \frac{105 \cdot 104 \cdot 103 \cdot 102!}{102!3!} \\
 &= \frac{105 \cdot 104 \cdot 103 \cdot \cancel{102!}}{\cancel{102!}3!} \\
 &= \frac{35 \cdot 52 \cdot 103}{3 \cdot 2 \cdot 1} = 35 \cdot 52 \cdot 103
 \end{aligned}$$



**ماذا:** هذا الموضوع قليل أن يتم مناقشته في سياق "الرياضيات في علم البيانات". لكن الحقيقة هي أن علوم البيانات الحديثة يتم عملها بمساعدة أنظمة حاسوبية، والرياضيات المتقطعة تعد جوهر هذه الأنظمة. مراجعة الرياضيات المتقطعة ستزود المتعلم بمفاهيم أساسية لاستخدام الخوارزميات وهياكل البيانات في مشاريع تحليل البيانات، بعض مواضيع التعلم هنا:

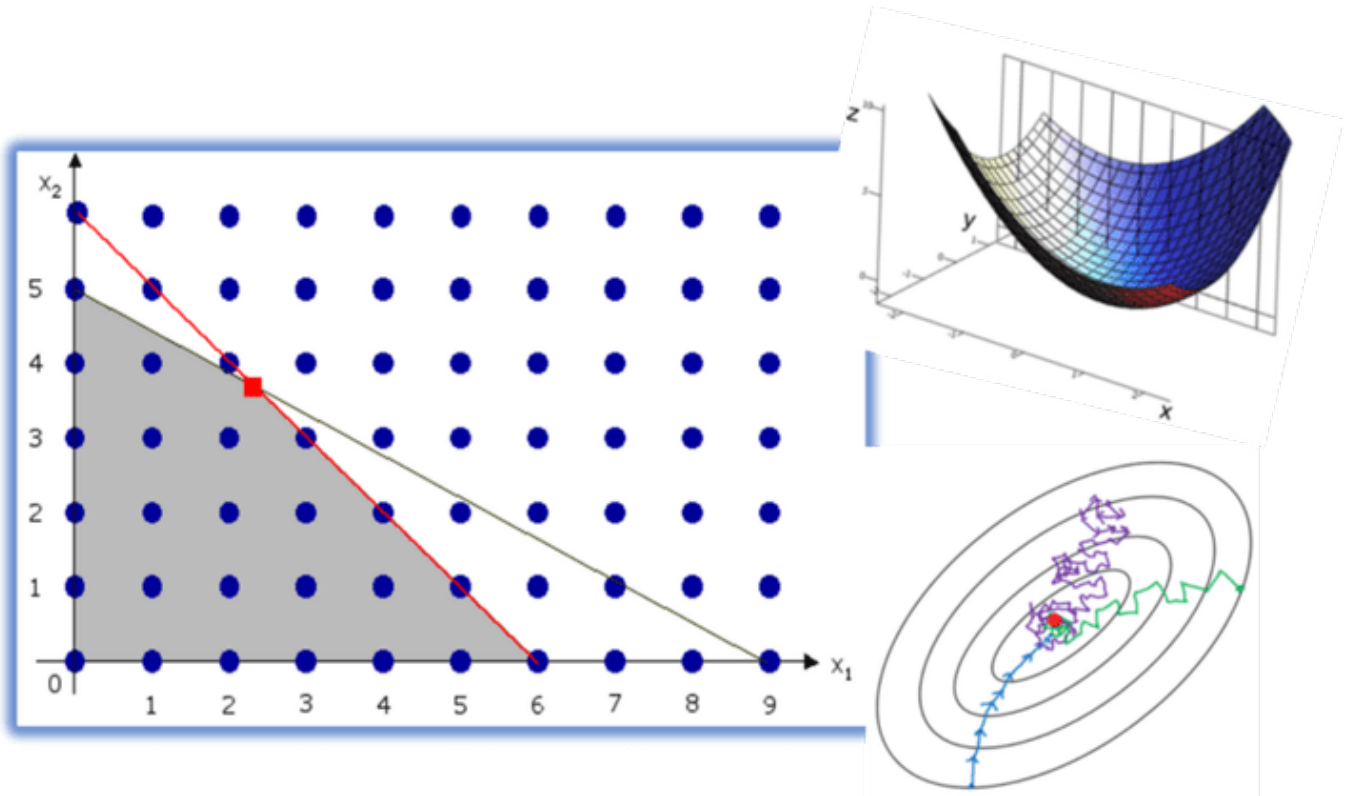
- المجموعات، المجموعات الجزئية، مجموعة القوى
- دوال التعداد، الرياضيات التوافقية، قابلية العد
- مهارات البرهان الرياضي - البرهان بالاستقراء، البرهان العكسي
- أساسيات المنطق الاستقرائي
- أساسيات هياكل البيانات - مكس (stacks)، رتل (queues)، الرسوم البيانية، المصفوفات، جداول التقطيع (hash tables)، الهيكل الشجري
- ميزات الرسوم البيانية - العناصر المترابطة، المستوى، مفاهيم الحد الأقصى للتدفق / القص الأدنى، الترميز اللوني
- العلاقات والمعادلات المتكررة
- مفاهيم نمو الدوال و مفهوم تعقيد O الكبيرة (big O notation)

**مثال (أو اثنين) لكيفية استعمالها:** في أي تحليل الشبكة الاجتماعية تحتاج أن تكون ملماً بخصائص الرسوم البيانية والخوارزميات السريعة لإجراء عمليات البحث و المسح على الشبكة. مهما كان اختيارك من الخوارزميات عليك أن تفهم تعقيدات الزمن و مساحة الذاكرة، مثلاً كيف أن زمن التشغيل والمساحة المطلوبة تكبر مع حجم البيانات المدخلة باستخدام تعقيد  $O$  الكبيرة.

**أين تتعلم:**

- [Introduction to Discrete Mathematics for Computer Science](#)  
Specialization—Coursera, Univ. of California San Diego
- [Introduction to Mathematical Thinking](#)—Coursera, Stanford
- [Master Discrete Mathematics: Sets, Math Logic, and More](#)  
—Udemy

## التحسين الرياضي (optimization) ومواضيع أبحاث العمليات



هذه المواضيع مختلفة قليلاً عن الحوارات التقليدية في مجال الرياضيات التطبيقية، بحيث أنها أكثر ارتباطاً واستخداماً في المجالات الدراسية المتخصصة - علوم الحاسب النظرية، نظرية التحكم أو أبحاث التشغيل. بالرغم من ذلك، حصولك على فهم مبدئي لهذه التقنيات القوية يمكن أن يفيدك في ممارسة تعلم الآلة. عملياً تهدف جميع خوارزميات/تقنيات تعلم الآلة إلى تقليل معدل الخطأ المقدر والذي يكون خاضع للقيود المختلفة. هذا هنا يعد

مشكلة تحسين (optimization). مواضيع للتعلم:

- مبادئ التحسين الرياضي (optimization) - كيفية صياغة المشكلة
- الحد الأقصى، الحد الأدنى، الدالة المحدبة، الحل المطلق
- البرمجة الخطية، خوارزمية التبسيط
- برمجة الأعداد الصحيحة
- البرمجة القيدية، مسألة حقيبة الظهر

**مثال (أو اثنين) لكيفية استعمالها:** مشاكل الانحدار الخطي البسيط لها حل تحليلي معين باستخدام دالة الخسارة في المربعات الصغرى لكن مشاكل الانحدار اللوجستي لا ينطبق عليها ذلك. لفهم السبب تحتاج ان تكون لديك معرفة بمبدأ التحدب في التحسين الرياضي. مسار البحث هذا سيوضح لماذا علينا أن نرضى بالحلول التقريبية في أغلب مشاكل تعلم الآلة. هذه حقيقة عظيمة علينا أن نفهمها بعمق.

**أين تتعلم:**

- [Optimization Methods in Business Analytics—edX, MIT](#)
- [Discrete Optimization—Coursera, University of Melbourne](#)
- [Deterministic Optimization—edX, Georgia Tech](#)

روابط لمقالات ممتازة مرتبطة بهذا الموضوع:

- [Mathematics MOOCs for Data Science 15](#)
- [How to Learn Math for Data Science, The Self-Starter Way](#)
- [How much Math & Stats do I need on my Data Science resume](#)
- [MOOCs on Mathematics & Statistics for Data Science & 19](#)
- [Machine Learning](#)
- [Learning Math for Machine Learning](#)

## ختاماً، بعض الأفكار الإضافية

لا داعي للشعور بالخوف أو الضياع، هذا الموضوع يحتوي على أشياء كثيرة لتعلمها وإتقانها، خصوصاً اذا كنت لا تمارسها بشكل منتظم. لكن هناك مصادر كثيرة وفيديوهات رائعة على الإنترنت. مع القليل من الوقت والجهد تستطيع عمل خطتك الخاصة التي تتضمن مصادر تعلم حسب حاجتك و مستواك.

لكن تأكد أنك بعد أن تكون حدثت معرفتك بهذه المواضيع (العديد منها قد تكون درستها مسبقاً)، وتعلمت مفاهيم

جديدة، سوف تشعر بأنك جاهز لتسمع الألحان المخفية التي لم تكن تسمعها مسبقاً أثناء عملك على مشاريع متعلقة بتعلم الآلة أو تحليل البيانات المعتادة. وهذه تعد قفزة كبيرة نحو أن تصبح عالم أو عالمة بيانات