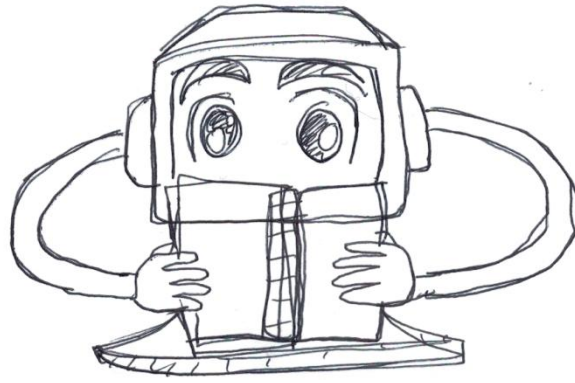


تعلم الآلة

Machine Learning



عبدالرحمن عثمان

مارس 2018

نسخة غير منقحة

قيد التطوير

آخر تحديث 4 أبريل 2018

لا تعطيني سمكة بل علمني كيف اصطادها

سنؤلف مثلاً مشابه بلسان الحاسب وهو

لا تعطيني نموذجاً بل علمني كيف اصنعه

جدول المحتويات

1. ماهو تعلم الآلة
2. انواع تعلم الآلة
3. الانحدار (regression)
4. التصنيف (classification)
5. التعلم الآلي دون إشراف (unsupervised learning)
6. التعلم الآلي شبه الخاضع للإشراف (Semi-Supervised Learning)
7. التعلم التعزيزي (reinforcement Learning)
8. أمثلة لتطبيقات تعلم الآلة

الباب الأول

ما هو تعلم الآلة

الباب الأول

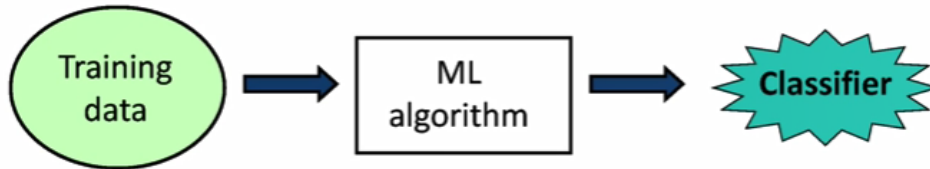
ما هو تعلم الآلة

في هذا الباب سنتحدث عن تعلم الآلة أو التعلم الآلي، والذي يمثل مقدرة الخوارزميات على التعلم الذاتي من خلال البحث والتنقيب في العلاقات بين المعطيات. حيث يعتبر تعلم الآلة وسيلة هامة ومفيدة وتساعد كثيرا في كثير من المجالات.

تعريف

تعلم الآلة هو تصميم برامج تستطيع التعلم من الخبرات السابقة، وهو إضافة مقدرة التعلم لأجهزة الحاسب. حيث يتم تصميم وتطوير الخوارزميات والتقنيات التي تمكن الحاسب من امتلاك خاصية "التعلم" [1]. المهمة الأساسية للتعلم الآلي هو استخراج معلومات قيمة من بيانات تدريبية ثم يستفيد منها في بناء نموذج قادر على التنبؤ بشكل البيانات الجديدة، شكل رقم 1. **إذا استطاع برنامج حاسوبي تنفيذ مهام معينة بالاستفادة من خبرة سابقة فيمكننا ان نقول انه تعلم**

تعلم الآلة هو المجال العلمي الذي يتعامل مع الطرق التي تتعلم فيها الآلات من التجربة. بالنسبة للعديد من العلماء ، فإن مصطلح "التعلم الآلي" مماثل لمصطلح "الذكاء الاصطناعي" ، بالنظر إلى أن إمكانية التعلم هي السمة الرئيسية للكيان الذي يسمى الذكاء بالمعنى الأوسع للكلمة [2]. هدف الذكاء الصناعي هو تصميم آلة تقلد العقل البشري في تصرفها، وللوصول لمثل هذا الهدف يجب ان يكون للآلة مقدرة على التعلم والتمثيل المنطقي للمعرفة والتفكير المجرد.



شكل رقم 1-1: تتعلم الخوارزمية من بيانات التدريب [3]

الفرق بين تعلم الآلة وتنقيب البيانات

يرتبط التعلم الآلي بدراسة وتصميم وتطوير الخوارزميات التي تعطي الحواسيب القدرة على التعلم دون برمجة صريحة (تعريف آرثر صموئيل). بينما يمكن تعريف تنقيب البيانات على أنه العملية التي تبدأ من البيانات غير المنظمة ظاهريًا تحاول استخراج المعرفة و / أو أنماط غير معروفة مثيرة للاهتمام. خلال هذه العملية يتم استخدام خوارزميات التعلم الآلي (Giovanni Di Orio من Institute for the Development of New Technologies) [4].

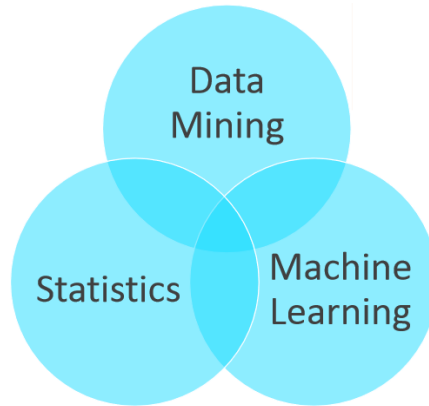
إن تقنيات التعلم الآلي هي تقنيات عامة ويمكن تطبيقها في أماكن مختلفة. يركز تنقيب البيانات على استخدام البيانات في مجال ما لفهم بعض الأسئلة في هذا المجال (مثل مجال وسائل التواصل الاجتماعي، أو بيانات المستشعر، ومقاطع الفيديو، وما إلى ذلك). وقد يستخدم تنقيب البيانات تقنيات التعلم الآلي للحصول على ما يسعى إليه (Pramod Anantharam من Wright State University) [4].

هل هنالك فرق بين تعلم الآلة وتنقيب البيانات والإحصاء؟

تنقيب البيانات والإحصاء وتعلم الآلة جميعها تعتمد على البيانات وتساعد المؤسسات على اتخاذ قرارات أفضل وتتأثر بشكل إيجابي على نمو أي عمل تجاري. حول هذا السؤال (ما هو الفرق بين تنقيب البيانات والإحصاءات والتعلم الآلي؟)، رد البروفيسور لواسرمان الأستاذ في كل من قسم الإحصاء والتعليم الآلي في جامعة كارنيجي ميلون، بإجابة مختصرة هي: لا شيء. فالثلاث مجالات تهتم بنفس السؤال: **كيف نتعلم من البيانات**؟ وبالنظر إلى إجابة ويسرمان، فإن التخصصات الثلاثة هي نفسها إلى حد كبير ولكن مع اختلافات بسيطة، بل يمكن الإشارة إليها على أنها توائم متماثلة تستخدم كلمات ومصطلحات مختلفة. فيما يلي بعض الاختلافات بين المجالات الثلاثة:

- يستخدم تنقيب البيانات قوة التعلم الآلي والإحصاءات من أجل التنقيب في بيانات ضخمة والتوصل إلى أنماط.
- الإحصاء هو أساس تنقيب البيانات وخوارزميات التعلم الآلي.
- الإحصاء هو دراسة جمع وتحليل ودراسة البيانات والتوصل إلى الاستنتاجات والتنبؤ بشأن المستقبل.
- يعتبر التعلم الآلي جزءًا من علم البيانات الذي يركز بشكل أساسي على كتابة الخوارزميات بطريقة تجعل أجهزة الحاسب قادرة على التعلم بمفردها واستخدام التعلم لمعرفة مجموعة البيانات الجديدة متى ما وجدت. يستخدم التعلم الآلي قوة الإحصاء ويتعلم من مجموعة البيانات التدريبية. على سبيل المثال، نحن نستخدم الانحدارات والتصنيفات وما إلى ذلك للتعلم من بيانات التدريب (training dataset) واستخدام هذا التعلم لتقدير مجموعة بيانات الاختبار (estimate test dataset).
- الهدف من "تنقيب البيانات والإحصاء" هو إجراء تحليل للبيانات ولكن كلاهما أدوات مختلفة. تضمنت عملية تنقيب البيانات النمذجة والتنبؤ وتحسين مجموعة البيانات (optimizing a dataset)، بينما توضح الإحصاء مدى كفاءة مجموعة البيانات [5].

الإحصاء: يقيس البيانات من العينة ويقدر سلوك المجتمع (estimates population behavior)
تنقيب البيانات: يكتشف النمط في البيانات
التعلم الآلي: يتعلم من بيانات التدريب ويتنبأ أو يقدر المستقبل [5]



شكل رقم 2: نقاط التقاطع والاختلاف

الصورة من http://data-mining-forum.de/complementarities_and_differences_between_machine_learning_and_data.pdf

تطبيقات تعلم الآلة

الآن أصبحت معظم التطبيقات إن لم يكن جميعها تستخدم تعلم الآلة بشكل أة بآخر. أمثلة لذلك:

- المجالات الطبية والتشخيصية
- مجالات الامن والحماية
- المجالات التجارية
- معالجة اللغات الطبيعية (natural language processing)
- تمييز الأنماط (syntactic pattern recognition)
- محركات البحث (search engines)
- التشخيص الطبي والمعلوماتية الحيوية والمعلوماتية الكيميائية، تصنيف تسلسلات الحمض النووي
- تمييز الكلام (speech recognition)
- تمييز الكتابة (handwriting recognition)
- تمييز الأشياء (object recognition)
- رؤية بالحاسب (computer vision)
- حركة الروبوت (robot locomotion).

فئات مهام التعلم الآلي

تصنف مهام التعلم الآلي عادة إلى ثلاث فئات رئيسية هي:

1. التعلم تحت الإشراف (Supervised Learning) ، حيث ينشأ النظام عن وظيفة من بيانات التدريب (إعطاء أمثلة المدخلات والمخرجات المقابلة لها، ليجاد علاقة تربط بين المدخلات بالمخرجات)، فهو يستفيد من بيانات معلمة (مصنفة).
2. التعلم بدون إشراف (Unsupervised Learning) ، حيث يحاول نظام التعلم استنتاج هيكل البيانات غير المعلمة (ترك خوارزمية التعلم للاعتماد على نفسها في استكشاف هيكل مدخلاتها. وذلك لاكتشاف الأنماط الخفية في البيانات)
3. التعلم شبه الخاضع للإشراف (Semi-Supervised Learning): يستخدم هذا النوع خليط من البيانات المعلمة وغير المعلمة في بناء نماذجه.
4. تعلم التعزيز (Reinforcement Learning) ، الذي يتفاعل فيه النظام مع بيئة ديناميكية (تفاعل البرنامج مع بيئة ديناميكية. الهدف، هو تحقيق غاية معينة دون معلم ينبئه حتى باقترابه من غايته تلك)

أهمية تعلم الآلة

هذا المجال يعتبر من المجالات الحديثة والهامة والتي برزت مع عصر البيانات الضخمة والحاجة الى التنقيب فيها والاستفادة منها [6]. فيما يلي بعض مسببات بروز تعلم الآلة:

- بروز عصر البيانات الضخمة، حيث تتوفر الآن كمية ضخمة من البيانات مختلفة ومتنوعة
- توفر الحاسبات القوية والسريعة التي تستطيع معالجة هذا الكم الكبير من البيانات
- زيادة القدرة التخزينية للأجهزة وتوفرها ورخصها ادي الى تراكم كم كبير من البيانات
- يهدف التعلم الآلي إلى تحليل بيانات أكبر حجمًا وأكثر تعقيدًا، تلقائيًا
- يستطيع تعلم الآلة الوصول إلى نتائج أسرع وأكثر دقة تساعد في إنتاج تنبؤات ذات قيمة عالية
- يستفيد اصحاب القرار من النتائج والتبؤات في اتخاذ القرار الصائب
- تعلم الآلة تستطيع إنتاج آلاف النماذج في أقل وقت (يعتمد على قوة الحوسبة)
- تعلم الآلة تستطيع التنبؤ بالسلوك والتعرف على الأنماط بكفاءة أعلى من البشر

بعض الأمثلة لاستخدامات تعلم الآلة

هنالك الكثير من المجالات التي يستخدم فيها تعلم الآلة مثل الخدمات المالية والرعاية الصحية والمبيعات.

هنالك الكثير من الجهات التي تستخدم تعلم الآلة في عملها مثل فيسبوك، خرائط جوجل، محركات بحث جوجل، gmail، paypal، Netflix، Uber [7].

وفيما يلي بعض الامثلة في استخدام تعلم الآلة:

- استخدام موقع أمازون لتعلم الآلة في التنبؤ بما يريده العملاء حيث ساعد ذلك في زيادة ارباحه
- استخدام الفيسوك لتقنيات تعلم الآلة في التعرف على الوجوه (Face Recognition) في الصور.
- استخدام تعلم الآلة من قبل شركة جوجل للتحكم في القيادة الذاتية للسيارة.

تسريع تعلم الآلة في السحابة

أكثر من 60 في المئة من المؤسسات تتبنى السحابة من أجل قابليتها للتوسع المنقطعة النظير. حيث تشجع امكانية التوسع الاقتصادي الشركات إلى اتباع أساليب أفضل وأكثر للتعلم الآلي لإنشاء النماذج الخاصة بها.

1. تعلم الآلة، ويكيبيديا. Available from: https://ar.wikipedia.org/wiki/تعلم_الآلة.
2. Kavakiotis, I., et al., *Machine learning and data mining methods in diabetes research*. Computational and structural biotechnology journal, 2017. **15**: p. 104-116.
3. Witten, I.H. *Data Mining with Weka*. Available from: <https://weka.waikato.ac.nz/dataminingwithweka/preview>.
4. Sadeg, S. *What is the difference between machine learning and data mining*. 2014; Available from: https://www.researchgate.net/post/What_is_the_difference_between_machine_learning_and_data_mining.
5. dezyre. *Data Mining vs. Statistics vs. Machine Learning*. 2017; Available from: <https://www.dezyre.com/article/data-mining-vs-statistics-vs-machine-learning/349>.
6. Nora Elshwemy, ر.ر. مقدمة عن تعلم الآلة. Available from: <http://www.egyres.com/articles/مقدمة-عن-تعلم-الآلة-machine-learning>.
7. Feldman, M. *10 Real-World Examples of Machine Learning and AI [2018]*. 2018; Available from: <https://www.redpixie.com/blog/examples-of-machine-learning>.

الباب الثاني

أنواع تعلم اللغة

الباب الثاني

أنواع تعلم الآلة

التعلم الآلي عادة يقسم إلى أربع فئات رئيسية هي التعلم تحت الإشراف (Supervised Learning)، والتعلم بدون إشراف (Unsupervised Learning)، والتعلم شبه الخاضع للإشراف (Semi-Supervised Learning)، وتعلم التعزيز (Reinforcement Learning). سنتحدث في هذا الباب عن هذه الفئات الأربع.

فئات مهام التعلم الآلي

تحدثنا مسبقاً عن تعلم الآلة وذكرنا بأنه يمكن الحواسيب من امتلاك خاصية "التعلم"، وهو يركز على إنشاء نموذج حقيقي من بيانات موجودة. وقد يستخدم لتحسين نموذج موجود أو إنشاء نموذج جديد. يمكن تقسيم مهام تعلم الآلة إلى أربعة أنواع هي [1]:

1. التعلم تحت الإشراف (Supervised learning)، حيث ينشأ النظام عن وظيفة من بيانات التدريب (إعطاء أمثلة المدخلات والمخرجات المقابلة لها، ليجاد علاقة تربط بين المدخلات والمخرجات)
2. التعلم بدون إشراف (Unsupervised learning)، حيث يحاول نظام التعلم استنتاج هيكل البيانات غير المعلمة (ترك خوارزمية التعلم للاعتماد على نفسها في استكشاف هيكل مدخلاتها. وذلك لاكتشاف الأنماط الخفية في البيانات)
3. التعلم شبه إشراف (Semi-Supervised learning): يستخدم هنا النظام بيانات معلمة وغير معلمة لبناء نموذج التنبؤ.
4. تعلم التعزيز (Reinforcement learning)، الذي يتفاعل فيه النظام مع بيئة ديناميكية لتحقيق غاية معينة دون معلم ينبئه حتى باقترابه من غايته تلك.

2.1. التعلم تحت الإشراف (supervised learning)

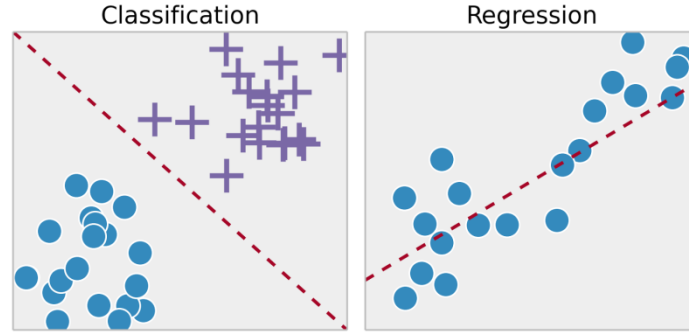
في هذا النوع قد لا تعرف العلاقات الداخلية للبيانات التي تقوم بمعالجتها، لكنك تعرف جيداً أي المخرجات التي تحتاجها من النموذج الخاص بك. فمثلاً: "التنبؤ متى سيلغي المستخدمون اشتراكاتهم".

هنا نجد أن مخرجات نموذجك محددة وهي: "هل سيقوم المستخدم X بإلغاء اشتراكه". ما قد لا تعرفه بعد هو كيفية معرفة المستخدمين الذين سيلغون اشتراكهم. لذا يمكننا استخدام مجموعة من البيانات الموجودة لتدريب نموذج ما على التنبؤ بهذا الجانب الخاص بمستخدمك. عادة ما يستخدم تدريب النموذج جزءاً من البيانات "للتعلم" (training)، وجزءاً من البيانات للتحقق من مدى دقة النموذج وقياسه (test).

مثلاً إذا كان لديك 10000 مستخدم. من هؤلاء، منها 5000 تم إلغاء اشتراكهم و 5000 لا يزالون مشتركين. هنا يمكنك أخذ بيانات من 4500 مستخدم من الذين ألغوا اشتراكهم، و 4500 من المستخدمين الذين ما زالوا مشتركين (بيانات من 9000 مستخدم). وجعل النموذج الخاص بك يتدرب على هذه البيانات، والساح له "رؤية" الذين ألغيت اشتراكهم والذين ما زالوا مشتركين (رؤية الحل). بعد أن يتم تدريب نموذجك سيكون جاهزاً لبدء التنبؤ. لذلك يمكنك الآن تغذية نموذجك ببيانات بقية المستخدمين (1000 مستخدم)، وهنا لن تسمح للنموذج بالاطلاع على الحل (من من الالف ما زال مشترك ومن تم إلغاء اشتراكه). سيعمل النموذج على إيجاد أفضل نتيجة، ويمكنك مقارنة النتيجة مع القيمة الحقيقية (الحل). إذا كان النموذج قد تنبأ بشكل صحيح مثلاً بعدد 891 من أصل 1000 مستخدم، فإن النموذج لديه دقة تبلغ 89.1 بالمائة [2].

هناك نوعان من التعلم الخاضع للإشراف هما:

1. **التصنيف (Classification):** التنبؤ بتصنيفات، مثل الألوان، أنواع الحيوانات وأنواع الفاكهة.
2. **الانحدار أو التوقع (Regression):** التنبؤ بكمية، مثل توقع درجة الحرارة بناءً على بيانات تاريخية، بالإضافة إلى اتجاه الرياح والغيوم. وكذلك توقع أسعار المنازل بناءً على معطيات كحجم البيت، مكان الحي، وعدد الغرف ودورات المياه.



شكل رقم 1-2: الفرق بين التصنيف والانحدار [1]

- Classification is the task of predicting a discrete class label.
- Regression is the task of predicting a continuous quantity.

There is some overlap between the algorithms for classification and regression; for example:

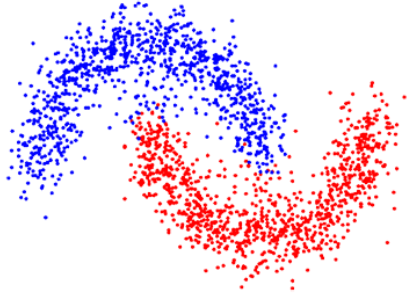
- A classification algorithm may predict a continuous value, but the continuous value is in the form of a probability for a class label.
- A regression algorithm may predict a discrete value, but the discrete value in the form of an integer quantity [3].

2.2. تعلم دون مشرف (Unsupervised Learning)

في هذا النوع أنت لا تعرف ما تريد استخراجه من النموذج. ربما تعتقد أن هناك بعض أنواع العلاقات أو الارتباط (correlation) بين البيانات المتوفرة لديك. وقد تكون البيانات معقدة للغاية ولا يمكن التكهّن بها. لذا في هذه الحالات تقوم بتنظيم بياناتك (normalize your data) في صيغة منطقية للمقارنة، ثم ترك النموذج يعمل عليها ومحاولة العثور على بعض العلاقات. تتمثل إحدى الخصائص المميزة لهذه النماذج في أنه يمكن للنموذج اقتراح طرق مختلفة لتصنيف بياناتك أو ترتيبها، والأمر متروك لك لإجراء مزيد من الأبحاث لكشف النقاب عن شيء مفيد.

على سبيل المثال، بعد معالجة جميع البيانات المتعلقة بجميع مستخدمي منتجك باستخدام خوارزمية غير خاضعة للإشراف، قد يتم إنشاء طريقة لتجميع المستخدمين إلى مجموعتين. بعد فحص ومقارنة هاتين المجموعتين، قد تجد أن المجموعة (أ) في موقع جغرافي، والمجموعة (ب) في موقع آخر. يمكنك التصرف بناءً على هذا التقسيم المحدد للبيانات (الأمر متروك لك لمعرفة ذلك)، وإذا لم تستفيد من النتيجة يمكنك إعادة ترتيب البيانات والحصول على نتائج مختلفة تكون أكثر فائدة لك [2].

في التعلم غير الخاضع للإشراف، يحاول النظام اكتشاف البنية الخفية للبيانات أو الارتباطات بين المتغيرات. وفيه يجمع البرنامج البيانات المتشابهة إلى مجموعات، ثم يتم تصنيف عينة الاختبار بناءً على قربها أو بعدها من هذه المجموعات.

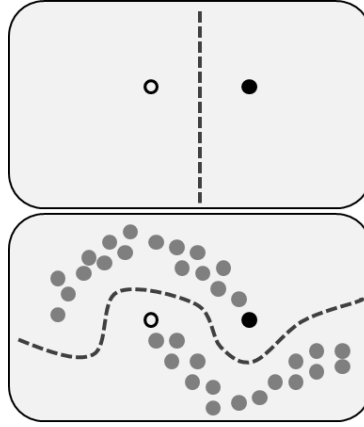


شكل رقم 2-2: دون إشراف [1]

من أشهر خوارزميات التعلم الغير خاضع للإشراف خوارزمية (K-Means). ومن التطبيقات التي تستخدم هذا النوع من التعلم كثيرا، تجميع العملاء أصحاب التفضيلات المتشابهة في نفس المجموعات. أو اكتشاف وتصنيف المجتمعات (الأشخاص ذوو الاهتمامات المشتركة) في وسائل التواصل الاجتماعي.

2.3. التعلم شبه الخاضع للإشراف

يستخدم البيانات المعلمة (labeled data) وغير المعلمة (unlabeled data). وهذا النوع مفيد لمن لا يستطيعوا تصنيف بياناتهم المعلمة. تسمح لنا هذه الطريقة بتحسين الدقة إلى حد كبير ، لأننا نستطيع استخدام البيانات غير المعلمة في مجموعة التدريب مع كمية صغيرة من البيانات المعلمة.



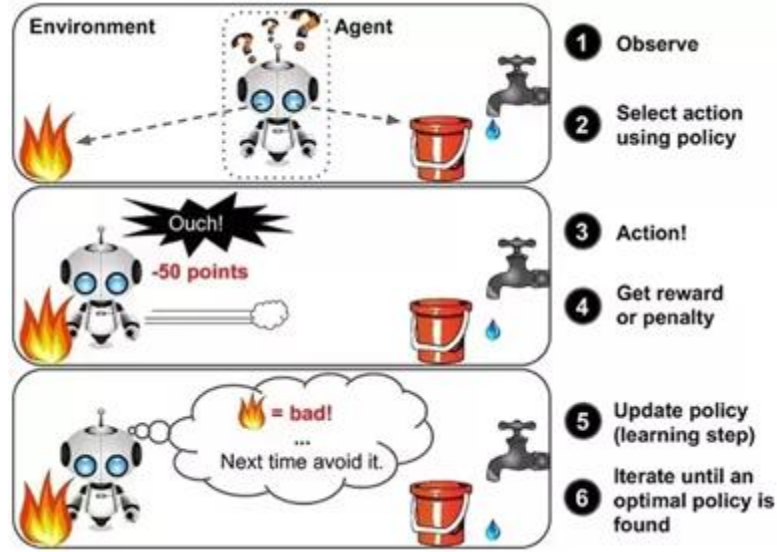
شكل رقم 2-3: شبه الإشراف [1]

2.4. تعلم التعزيز (Reinforcement Learning)

يعتقد البعض ان خوارزميات تعلم الآلة هي خوارزميات التعلم تحت الإشراف و التعلم بدون إشراف فقط، ولكن هنالك خوارزميات تعلم آلة أخرى مثل خوارزميات "تعلم التعزيز".

تعلم التعزيز هو شبيه بالتعلم البشري، حيث تتعلم الخوارزمية سياسة كيفية التصرف في بيئة معينة. كل تصرف له بعض التأثير في البيئة ، وتوفر البيئة المكافآت (rewards) التي توجه خوارزمية التعلم.

ينبئ نموذج تعلم التعزيز التفاعل بين عنصرين هما البيئة ووكيل التعلم (learning agent)، حيث تقوم البيئة بمكافأة الوكيل على التصرفات الصحيحة ، وهي إشارة التعزيز. حيث يستفيد الوكيل من المكافآت التي حصل عليها في تحسين معرفته بالبيئة لتحديد الإجراء التالي [4].



شكل رقم 4-2: مثال لتعلم التعزيز [4]

يمكن أحياناً استخدام أكثر من خوارزمية لأداء مهمة واحدة، كاستخدام خوارزمية التجميع، ثم تدريب خوارزمية تصنيف بناء عليها.

بعض خوارزميات تعلم الآلة

الجدول 1-2 يبين بعض خوارزميات تعلم الآلة.

الجدول 1-2: مقارنة عينة من خوارزميات تعلم الآلة [5]

التعلم دون اشراف (USL)	التعلم تحت الاشراف (SL)	
<ul style="list-style-type: none"> Clustering & Dimensionality Reduction <ul style="list-style-type: none"> SVD PCA K-means 	<ul style="list-style-type: none"> Regression <ul style="list-style-type: none"> Linear Polynomial Decision Trees Random Forests 	Continuous
<ul style="list-style-type: none"> Association Analysis <ul style="list-style-type: none"> Apriori FP-Growth Hidden Markov Model 	<ul style="list-style-type: none"> Classification <ul style="list-style-type: none"> KNN Trees Logistic Regression Naive-Bayes SVM 	Categorical

1. Korbut, D. *Machine Learning Algorithms: Which One to Choose for Your Problem*. 2017; Available from: <https://blog.statsbot.co/machine-learning-algorithms-183cc73197c>.
2. Machado, G. *ML Basics: supervised, unsupervised and reinforcement learning*. 2016; Available from: <https://medium.com/@machadogj/ml-basics-supervised-unsupervised-and-reinforcement-learning-b18108487c5a>.
3. Brownlee, J. *Difference Between Classification and Regression in Machine Learning*. 2017; Available from: <https://machinelearningmastery.com/classification-versus-regression-in-machine-learning>.
4. Akiwatkar, R. *What is reinforcement learning*. 2017; Available from: <https://www.quora.com/What-is-reinforcement-learning>.
5. *Differences Between Supervised Learning and Unsupervised Learning*. Available from: <http://www.differencebetween.net/technology/differences-between-supervised-learning-and-unsupervised-learning>.

الباب الثالث

النموذج الخطي

Linear regression

الباب الثالث

الانحدار الخطي

Linear regression

تحدثنا في الباب السابق عن انواع تعلم الآلة وذكرنا من بينها التعلم تحت الاشراف (supervised) وهو النوع الذي يتعلم من بيانات تدريبية. وذكرنا أن من انواعه الانحدار (regression). سنتحدث هنا في هذا الباب عن الانحدار الخطي وكيفية استخدامه في استخراج وتحديد العلاقات من البيانات والاستفادة من هذه العلاقات في التنبؤ.

تعريف

الانحدار نوع من انواع تعلم الآلة (احصائي) ويستخدم لتحديد ومعرفة العلاقة بين متغيرين او أكثر. فهو يتعلم عادة مع بيانات رقمية ومستمرة.

استخداماته

يتم استخدامه لتقدير القيم الحقيقية (تكلفة المنازل ، عدد المكالمات ، إجمالي المبيعات وما إلى ذلك) على أساس المتغير المستمر. هنا ، نقيم علاقة بين المتغيرات المستقلة وغير المستقلة (التابعة) عن طريق تركيب أفضل خط. ويعرف هذا الخط الأكثر ملاءمة باسم خط الانحدار وتمثله معادلة خطية:

$$Y = a * X + b.$$

مثال: لنفترض اننا طلبنا من تلميذ في الصف الخامس ترتيب التلاميذ في فصله حسب اوزانهم، دون أن يطلب منهم أوزانهم! ماذا تعتقد أن الطفل سوف يفعل؟ من المحتمل أن ينظر (تحليل بصري) الى اطوالهم ويرتبهم بناءا على ذلك. هذا هو الانحدار الخطي في الحياة الحقيقية، لقد اكتشف الطفل بالفعل أن هذه الاطوال وربطها بالوزن من خلال علاقة .

وهذا ما تقوم به معادلة الانحدار ، حيث:

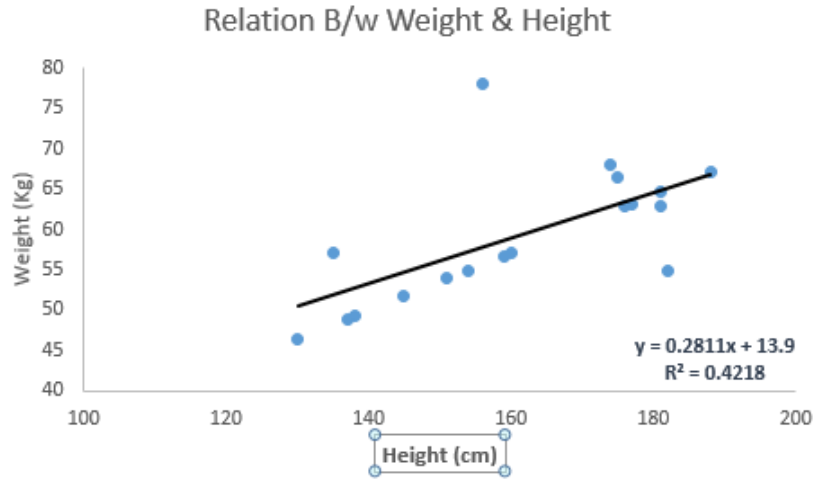
- Y – Dependent Variable (متغير تابع)
- a – Slope (الميل)
- X – Independent variable (متغير مستقل)
- b – Intercept (تقاطع)

يتم اشتقاق هذه المعاملات (coefficients) a و b على أساس تقليل مجموع فارق التربيع المرتفع بين نقاط البيانات وخط الانحدار (minimizing the sum of squared difference of distance between data points and regression line).

في الشكل التالي تم تحديد أفضل خط مناسب للعلاقة بين الوزن والطول، بمعادلة خطية:

$$y=0.2811x+13.9$$

الآن باستخدام هذه المعادلة ، يمكننا العثور على الوزن ، متى ما عرف الطول [1]..



انحدار خطي للعلاقة بين الوزن والطول [1]

الانحدار الخطي هو في الأساس نوعان: الانحدار الخطي البسيط والانحدار الخطي المتعدد. يتميز الانحدار الخطي البسيط بمتغير مستقل واحد. بينما يتميز الانحدار الخطي المتعدد بالعديد من المتغيرات المستقلة (أكثر من 1). أثناء البحث عن أفضل خط مناسب (best fit line)، يمكنك (you can fit a polynomial or curvilinear regression). وتعرف هذه بال (polynomial or curvilinear regression) [1].

في بقية الباب سنركز على الانحدار الخطي البسيط مع امثلة عملية له باستخدام ميكروسوفت اكسل.

الانحدار الخطي البسيط

لايجاد العلاقة بين متغيرين احدهما مستقل والآخر تابع (يتغير بتغير المستقل). وكلمة خطي جاءت من كون العلاقة تكون في شكل خط مستقيم (تذكر معادلة الخط المستقيم). ويمكننا تكوين نموذج رياضي لتمثيل هذه العلاقة.

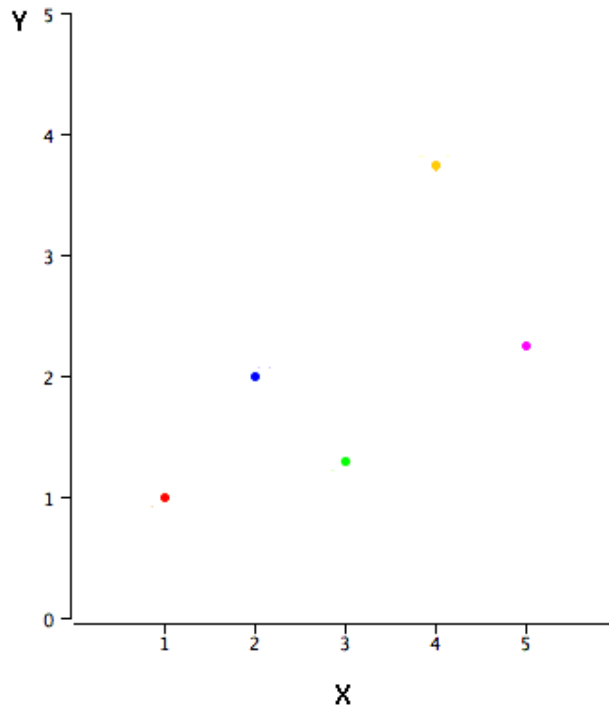
الانحدار الخطي البسيط هو دراسة العلاقة بين متغيرين فقط بحيث نحاول الوصول إلى علاقة خطية (أي معادلة خط مستقيم) بين هذين المتغيرين ، مثل العلاقة بين السعر والسعلة او بين درجة الحرارة واستهلاك الكهرباء.

في الانحدار الخطي البسيط يمكننا التنبؤ بقيم متغير معين (التابع) بناء على قيم متغير ثاني (المستقل).

جدول رقم 1: [2]

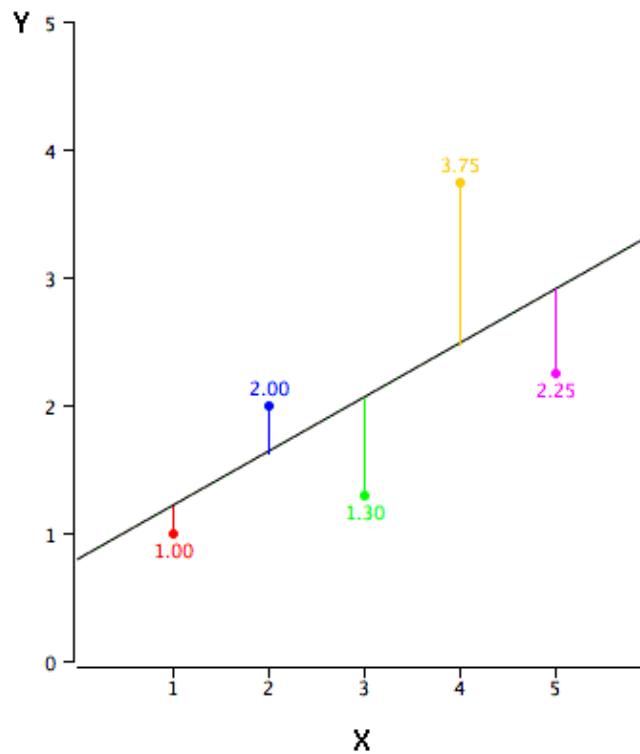
X	Y
1.00	1.00
2.00	2.00
3.00	1.30
4.00	3.75
5.00	2.25

يمكننا تمثيل بيانات الجدول رقم 1 في الشكل رقم 1.



شكل رقم 1: تمثيل بيانات جدول 1. [2]

يسعى الانحدار الخطي للعثور على أفضل خط مستقيم (best-fitting line) يمر خلال النقاط في الشكل رقم 1. لذلك فإن الخط الظاهر في الشكل رقم 2 يسمى خط الانحدار ويتكون من النتيجة المتوقعة على Y لكل قيمة ممكنة لـ X . تمثل الخطوط العمودية من النقاط إلى خط الانحدار أخطاء التنبؤ. نلاحظ أن النقطة الحمراء قريبة جداً من خط الانحدار لذلك يعتبر خطأها في التنبؤ صغير. على النقيض من ذلك، فإن النقطة الصفراء أعلى بكثير من خط الانحدار، وبالتالي فإن خطأ التنبؤ بها كبير.



شكل رقم 2: خط الانحدار. [2]

مثال حقيقي

تحصلت جامعة ما على درجات 105 طالب (معدل الطالب في المدرسة الثانوية و معدله في الجامعة) في تخصص علوم الحاسب. ثم استفادت الجامعة من هذه البيانات في التنبؤ بمعدل الطالب (GPA) بالجامعة اعتمادا على معدله السابق بالمدرسة.

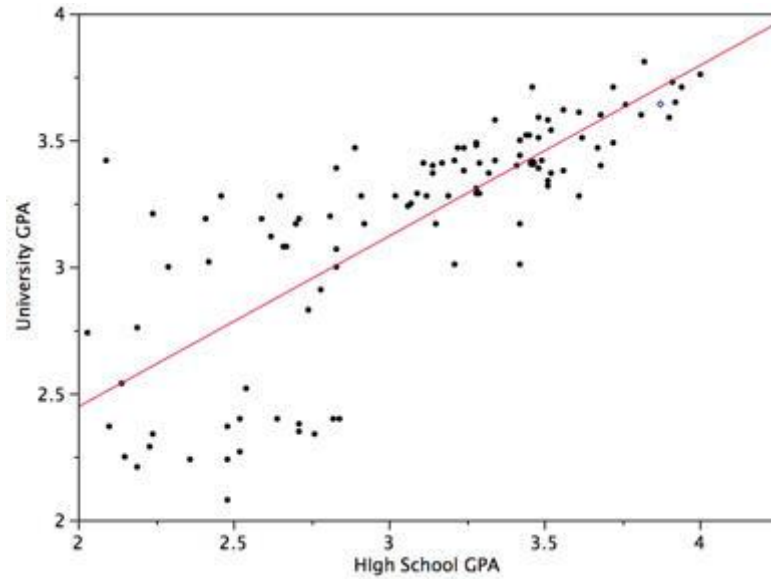
حيث تم استخدام الانحدار ليتعلم من البيانات الموجودة (وهي درجات الطلاب ال 105)، حيث استخرج الانحدار معادلة خطية تمثل الحل وهي كما يلي:

$$\text{University GPA}' = (0.675)(\text{High School GPA}) + 1.097$$

ويمثل الشكل رقم 3 الحل، حيث يتضح من الشكل ان هنالك علاقة ايجابية قوية (strong positive relationship). وكان معامل الارتباط (correlation) يساوي 0.78.

الوصول لمعامل الارتباط وإيجاد المعادلة التي تمثل العلاقة بين درجات الطلاب هو ما تعلمه الانحدار من البيانات. بعدها سيكون قادرا على التنبؤ بمعدل اي طالب في الجامعة اذا ما اعطي معدله بالثانوي. مثلا لو كان لدينا طالب ثانوي حصل على معدل 3 ، فإن معدله المتوقع له بالجامعة هو 3.12، وذلك بالتعويض في معادلة الحل، كما يلي:

$$\text{University GPA}' = (0.675)(3) + 1.097 = 3.12.$$



الشكل رقم 3 : خط الانحدار يمثل العلاقة بين معدل الثانوي ومعدل الجامعة [2]

التطبيق العملي

يستطيع اي شخص استخدام ادوات برمجية او لغة برمجة لتحديد العلاقة بين المتغيرات وإيجاد معادلة الانحدار ورسم الشكل الذي يوضح العلاقة. هنالك الكثير من الادوات المستخدمة في هذا المجال، من ابسطها ميكروسوفت اكسل الذي سنقوم باستخدامه في بقية الباب في تحليل الانحدار.

مثال باستخدام أكسل

نفترض ان لدينا البيانات الموضحة في الجدول التالي ونريد تحليلها وإيجاد العلاقات بين متغيراتها [3]. نلاحظ ان البيانات تتكون من عمودين (متغيرين اثنين فقط)، لذلك سنستخدم الانحدار الخطي بسيط بالاستفادة من أكسل.

X	Y
0	667
45	661
90	757
135	871
180	1210

الحل

افتح برنامج أكسل وادخل بياناتك فيه ثم تابع الخطوات التالية [4]:

1. On the Data tab, in the Analysis group, click Data Analysis
2. Select Regression and click OK
3. Select the Y Range. This is the predictor variable (also called dependent variable).
4. 4. Select the X Range. These are the explanatory variables (also called independent variables). These columns must be adjacent to each other.
5. Check Labels.
6. Click in the Output Range box and select a cell.
7. Check Residuals.
8. Click OK.

إذا لم تجد زر Data Analysis يمكنك تثبيته بإتباع الخطوات التالية:

1. On the File tab, click Options.
2. Under Add-ins, select Analysis ToolPak and click on the Go button
3. Check Analysis ToolPak and click on OK.

ستظهر لك نتيجة التحليل كما يلي:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.901897
R Square	0.813417
Adjusted R Square	0.751223
Standard Error	113.3243
Observations	5

ANOVA					Significance F
	df	SS	MS	F	
Regression	1	167961.6	167961.6	13.07868	0.036338
Residual	3	38527.2	12842.4		
Total	4	206488.8			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	574	87.78064	6.539028	0.00727	294.6428	853.3572	294.6428	853.3572
X	2.88	0.796362	3.616445	0.036338	0.34562	5.41438	0.34562	5.41438

تفسير النتائج

نلاحظ من المخرجات ان قيمة R Square هي 0.813417، اي انها حوالي 81% (كلما كانت قريبة من 1 تكون افضل)

أهمية F و P-values

لمعرفة ما إذا كانت نتائجك معتمدة (ذات دلالة إحصائية) ، انظر إلى Significance F (قيمتها هنا 0.036338). إذا كانت هذه القيمة أقل من 0.05 ، فأنت على ما يرام. إذا كان أكبر من 0.05 ، فمن الأفضل التوقف عن استخدام هذه المجموعة من المتغيرات المستقلة. قم بحذف متغير ذي قيمة P-value عالية (أكبر من 0.05) وإعادة تشغيل الانحدار حتى تنخفض Significance F إلى أقل من 0.05.

يجب أن تكون معظم P-values أقل من 0.05. في مثلنا هذا قيم p-values هي (0.00727 و 0.036338).

المعاملات (Coefficients)

The linear regression equation information is given in the last output set (the coefficients column). The first entry in the “Intercept” row is “a” (the y-intercept) and the first entry in the “X” column is “b” (the slope).

معادلة خط الانحدار هي:

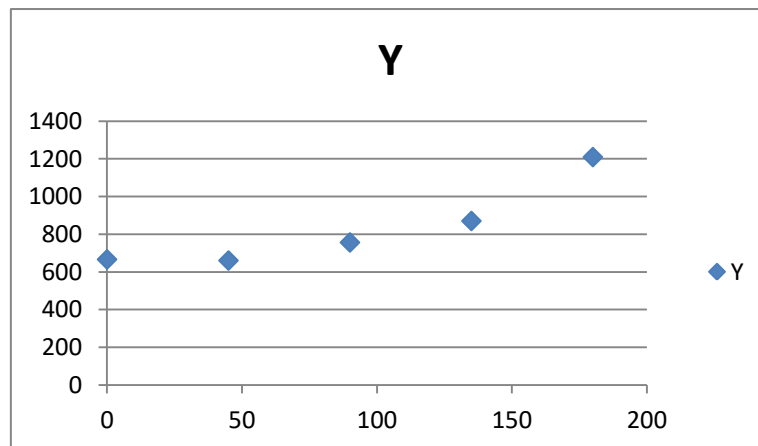
$$y' = 574 + 2.88x$$

يمكنك أيضا استخدام هذه المعاملات للقيام بتوقعات. على سبيل المثال ، إذا كان قيمة x تساوي 4، يمكنك الحصول على y بالتوضيح في المعادلة الانحدار كما يلي:

$$y' = 574 + 2.88 * (4)$$

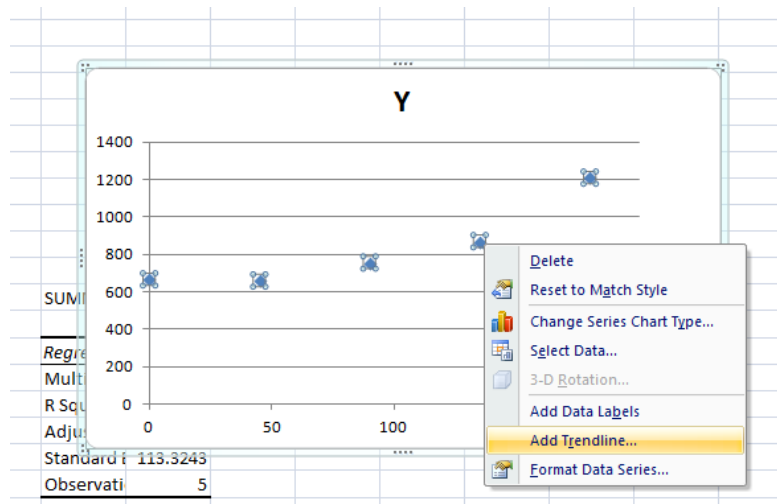
توضيح البيانات بالرسم في أكسل

ظلل البيانات التي تريد عمل رسمة لها ثم من Insert في قائمة أكسل اختار Scatter من Charts، فيظهر الشكل التالي:

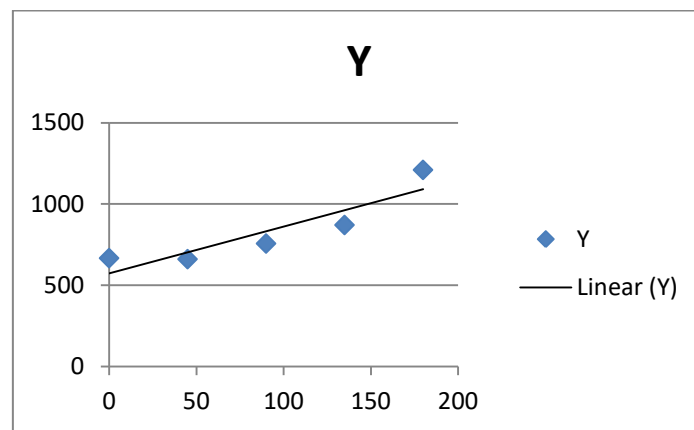


بعدها انقر على أي نقطة من النقاط التي تمثل البيانات في الشكل اعلاه، ثم انقر عليها بيمين الماوس ثم اختار Add Trendline ثم اختار Linear ثم

OK



تكون النتيجة كما يلي:



المراجع

1. Ray, S. *Essentials of Machine Learning Algorithms (with Python and R Codes)*. 2017; Available from: [./https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms](https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms)
2. Lane, D.M. *Introduction to Linear Regression*. Available from: <http://online.statbook.com/2/regression/intro.html>
3. *Doing a Linear Regression Analysis, Using Excel (version from Office '97)*. Available from: <https://msu.edu/course/psy/403/StatDemos/Regression/Regression.htm>
4. Easy, E. *Regression*. Available from: <http://www.excel-easy.com/examples/regression.html>

الباب الرابع

التصنيف

Classification

الباب الرابع

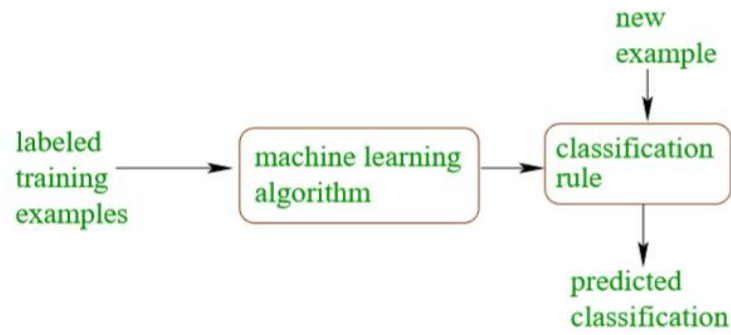
التصنيف

Classification

تحدثنا في الباب السابق عن احد نوعي التعلم تحت الاشراف (supervised) وهو الانحدار . سيكون الحديث هنا عن النوع الآخر من التعلم تحت الاشراف الا وهو التصنيف. وهو النوع الأكثر شيوعا واستخداما.

تعريف

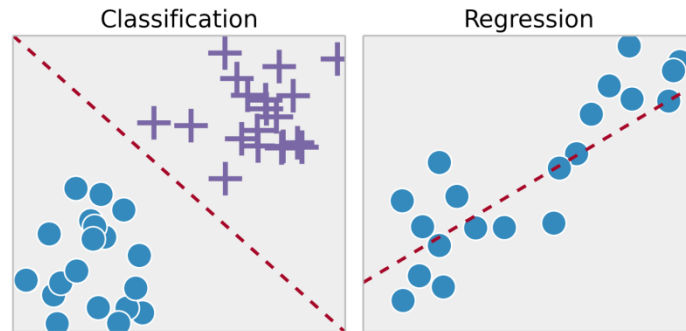
في ويكيبيديا تم تعريف التصنيف (في مجال التعلم الآلي والإحصاء) بأنه تحديد مجموعة من الفئات التي تنتمي إليها ملاحظة جديدة ، على أساس مجموعة تدريب تضم بيانات تحتوي على ملاحظات تكون عضويتها معروفة.



شكل رقم 1-4: التصنيف [1]

الفرق بين التصنيف والانحدار

الانحدار يتعامل مع بيانات رقمية مستمرة (continuous)، بنا يتعامل التصنيف مع بيانات معلمة (labelled) متقطعة (discrete). ولكن هناك بعض التداخل بين الخوارزميات التصنيف والانحدار، مثلا قد تتنبأ خوارزمية التصنيف بقيمة مستمرة ، ولكن القيمة المستمرة تكون في شكل احتمال ل class label. أيضا قد تتنبأ خوارزمية الانحدار بقيمة متقطعة، ولكن القيمة المتقطعة تكون في شكل كمية رقمية (integer quantity).



شكل رقم 1-4: مقارنة بين التصنيف والانحدار [2]

مجالات استخدام التصنيف

التصنيف من أكثر المجالات استخداما في تعلم الآلة، ومن الأمثلة ما يلي [1]:

- تصنيف النصوص (على سبيل المثال ، الرسائل غير المرغوب فيها)
- الكشف عن الغش
- التعرف الضوئي على الحروف
- رؤية الآلة (على سبيل المثال ، التعرف على الوجوه)
- معالجة اللغة الطبيعية (على سبيل المثال ، فهم اللغة المنطوقة)
- بيع التجزئة (على سبيل المثال: التنبؤ بما إذا كان العميل سيستجيب للترويج)
- المعلوماتية الحيوية (على سبيل المثال ، تصنيف البروتينات وفقاً لوظائفها)

خوارزميات التصنيف

هناك العديد من الخوارزميات التي تستخدم في التصنيف، منها [3] :

- Linear Classifiers: Logistic Regression, Naive Bayes Classifier
- Support Vector Machines
- Decision Trees
- Boosted Trees (boosting)
- Random Forest
- Neural Networks
- Nearest Neighbor
- Bagging

مصنف Naive Bayes

إنها تقنية تصنيف تستند إلى نظرية Naive Bayes مع افتراض الاستقلال بين المتنبئين. بعبارة بسيطة ، يفترض مصنف Naive Bayes أن وجود ميزة معينة في فئة (particular feature in a class) غير مرتبط بوجود أي ميزة أخرى (other feature). حتى إذا كانت هذه الميزات تعتمد على بعضها البعض أو على وجود ميزات أخرى ، فإن كل هذه الخصائص تساهم بشكل مستقل في الاحتمالية (contribute to the probability).

على سبيل المثال ، يمكن اعتبار الفاكهة تفاحة إذا كانت حمراء ومستديرة وقطرها حوالي 3 بوصات. حتى ولو كانت هذه الخصائص تعتمد على بعضها البعض أو على وجود ميزات أخرى ، فإن المصنف Naive Bayes سيعتبر كل هذه الخصائص تساهم بشكل مستقل في احتمال أن تكون هذه الفاكهة تفاحة.

من السهل بناء نموذج Naive Bayes وهو مفيد بشكل خاص لمجموعات البيانات الكبيرة جدًا (very large data sets). إلى جانب البساطة، من المعروف أن Naive Bayes تتفوق (outperform) حتى على طرق التصنيف المتطورة للغاية [3].

من السهل بناء نموذج Naive Bayesian ومفيد بشكل خاص لمجموعات البيانات الكبيرة جدًا. إلى جانب البساطة، من المعروف أن Naive Bayes تتفوق حتى على طرق التصنيف المتطورة للغاية.

نظرية بايز توفر طريقة لحساب (posterior probability)

Bayes theorem provides a way of calculating posterior probability $P(c|x)$ from $P(c)$, $P(x)$ and $P(x|c)$.

لننظرنا للمعادلة التالية:

$$P(c | x) = \frac{P(x | c) P(c)}{P(x)}$$

Likelihood
Class Prior Probability
↓
Predictor Prior Probability
Posterior Probability

$$P(c | X) = P(x_1 | c) \times P(x_2 | c) \times \dots \times P(x_n | c) \times P(c)$$

Here,

- $P(c|x)$ is the posterior probability of class (target) given predictor (attribute).
- $P(c)$ is the prior probability of class.
- $P(x/c)$ is the likelihood which is the probability of predictor given class.
- $P(x)$ is the prior probability of predictor.

مثال توضيحي

في الجداول أدناه توجد مجموعة بيانات تدريبية عن الطقس ومتغير الهدف المقابل "Play". نحتاج الآن إلى تصنيف ما إذا كان اللاعبون سيلعبون أم لا بناءً على حالة الطقس. دعونا نتابع الخطوات أدناه لأداء ذلك.

- الخطوة 1: تحويل مجموعة البيانات إلى جدول التردد (Convert the data set to frequency table)
- الخطوة 2: إنشاء جدول احتمال عن طريق إيجاد الاحتمالات مثل Overcast probability = 0.29 واحتمال اللعب هو 0.64.
- الخطوة 3: الآن، استخدم معادلة Naive Bayesian لحساب posterior probability لكل فئة (class). الفئة ذات (highest posterior probability) هو نتيجة للتنبؤ (outcome of prediction)

الجداول من [4]

Weather	Play
Sunny	No
Overcast	Yes
Rainy	Yes
Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No
Rainy	No
Sunny	Yes
Rainy	Yes
Sunny	No
Overcast	Yes
Overcast	Yes
Rainy	No

Frequency Table		
Weather	No	Yes
Overcast		4
Rainy	3	2
Sunny	2	3
Grand Total	5	9

Likelihood table		
Weather	No	Yes
Overcast		4
Rainy	3	2
Sunny	2	3
All	5	9
	=5/14	=9/14
	0.36	0.64

مثال

سوف يدفع اللاعبون إذا كان الطقس مشمسًا (sunny)، فهل هذه العبارة صحيحة؟

الحل: باستخدام الطريقة المذكورة أعلاه ، لذلك

$$P(\text{Yes} \mid \text{Sunny}) = P(\text{Sunny} \mid \text{Yes}) * P(\text{Yes}) / P(\text{Sunny})$$

$$P(\text{Yes}) = 9/14 = 0.64, P(\text{Sunny}) = 5/14 = 0.36, P(\text{Sunny} \mid \text{Yes}) = 3/9 = 0.33 \text{ هنا لدينا}$$

سنجد ان:

$$P(\text{Sunny} \mid \text{Yes}) = 3/9 = 0.33, P(\text{Sunny}) = 5/14 = 0.36, P(\text{Yes}) = 9/14 = 0.64$$

الآن النتيجة هي:

$$P(\text{Yes} \mid \text{Sunny}) = 0.33 * 0.64 / 0.36 = 0.60$$

لذي لديه احتمالية أعلى.

يستخدم Naive Bayes طريقة مشابهة للتنبؤ باحتمالية فئة مختلفة بناءً على سمات مختلفة (different class based on various attributes) [4].

تستخدم هذه الخوارزمية غالباً في تصنيف النص (text classification) وفي تصنيف مسائل متعددة الفئات (multiple classes).

الانحدار اللوجستي (Logistic Regression)

إنها طريقة إحصائية لتحليل مجموعة البيانات التي يوجد بها متغير واحد مستقل أو أكثر يحدد النتيجة. تُقاس النتيجة بمتغير ثنائي التفرع (dichotomous) (يوجد فيه نتيجتان محتملتان فقط (only two possible outcomes)). الهدف من الانحدار اللوجستي هو العثور على أفضل نموذج ملائم لوصف العلاقة بين الخاصية ثنائية التفرع المرغوبة (dichotomous characteristic of interest) (المتغير التابع = الاستجابة أو متغير النتيجة) ومجموعة المتغيرات المستقلة (التوقعية أو التفسيرية) (predictor or explanatory) [3].

قد يحصل خلط عند كثير من الناس بين Logistic Regression والـ Regression. حيث يعتبر الأول خوارزمية تصنيف بينما الثاني خوارزمية الانحدار.

يتم استخدامه لتقدير القيم المنفصلة (discrete) مثل (Binary values like 0/1, yes/no, true/false) بناءً على مجموعة معينة من المتغيرات المستقلة. وبكلمات بسيطة، يتنبأ باحتمالية حدوث حدث من خلال (fitting data to a logit function). لذلك يعوّف أحياناً بالاسم logit regression. وبما أنه ينبأ باحتمالات (predicts the probability) لذلك تكون مخرجاته بين 0 و 1.

التوضيح بمثال بسيط

لنفترض أن صديقك اعطاك لغزاً حله. لا يوجد سوى سيناريوهين للنتائج - إما حلها أو عدم حلها. تخيل الآن، أنه يتم منحك مجموعة واسعة من الألغاز / الاسئلة (puzzles / quizzes) في محاولة لفهم الموضوعات التي تجيدها. ستكون النتيجة لهذه الدراسة شيئاً كهذا - إذا تم إعطاؤك مسألة في حساب المثلثات للصف العاشر، فمن المحتمل أن تحلها بنسبة 70٪. أما إذا كان السؤال حول التاريخ للصف الخامس، فإن احتمال الحصول على إجابة هو 30٪. هذا هو ما يوفره لك الانحدار اللوجستي.

رياضياً: the log odds of the outcome is modeled as a linear combination of the predictor variables

odds = $p / (1-p)$ = probability of event occurrence / probability of not event occurrence

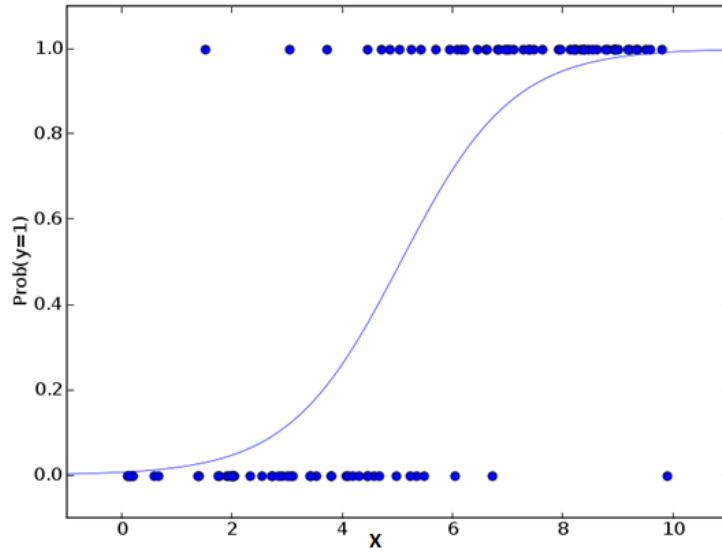
$\ln(\text{odds}) = \ln(p / (1-p))$

$\text{logit}(p) = \ln(p / (1-p)) = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_kX_k$

أعلاه، p هو احتمال وجود خاصية الفائدة. ويختار المعلمات التي تزيد من احتمال مراقبة قيم العينة بدلاً من تقليل مجموع الأخطاء المربعة (كما هو الحال في الانحدار العادي).

Above, p is the probability of presence of the characteristic of interest. It chooses parameters that maximize the likelihood of observing the sample values rather than that minimize the sum of squared errors (like in ordinary regression).

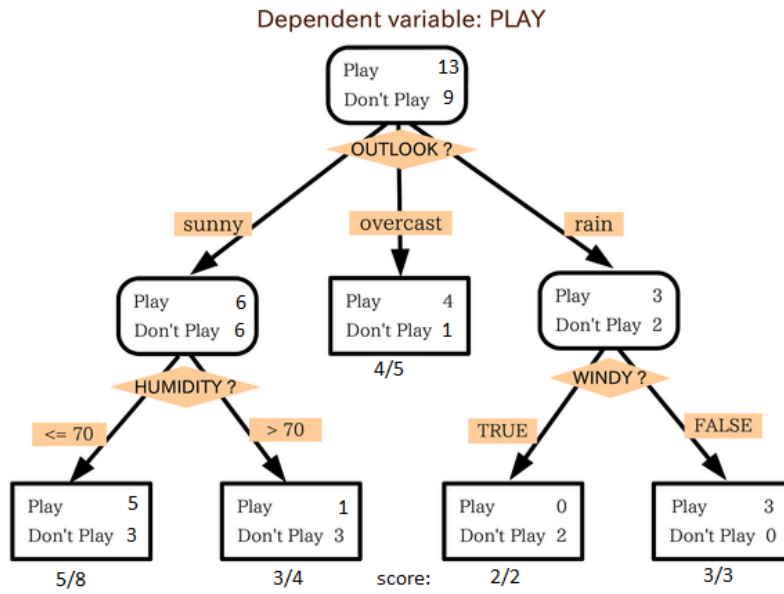
الآن، قد تسأل، لماذا تأخذ log؟ من أجل التبسيط، دعنا نقول فقط أن هذه هي واحدة من أفضل الطرق الرياضية لـ replicate a step function {Ray, 2017 #33}.



الصورة من {Ray, 2017 #33}

شجرة القرار (Decision Tree)

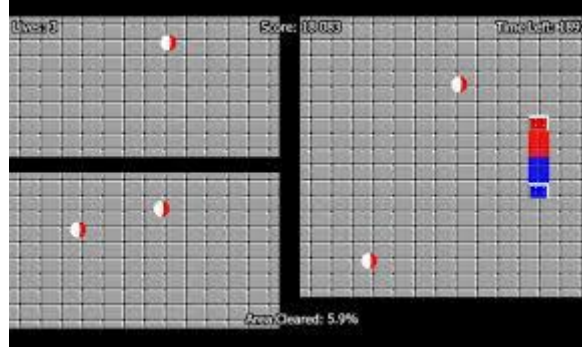
تبنى شجرة القرار تصنيفًا أو نماذج انحدار في شكل بنية شجرية. يقوم بتقسيم مجموعة البيانات إلى مجموعات فرعية أصغر وأصغر بينما يتم في نفس الوقت تطوير شجرة القرارات المرتبطة (associated decision tree is incrementally developed). النتيجة النهائية هي شجرة بها عقد قرار وعقد أوراق (decision nodes and leaf nodes). عقدة القرار (decision node) لها فرعان أو أكثر وتمثل عقدة الورقة (leaf node) تصنيفًا أو قرارًا (decision). أعلى عقدة القرار في الشجرة التي تتوافق مع أفضل مؤشر توقع يسمى الجذر (ode0root n). يمكن لأشجار القرار التعامل مع البيانات المستقلة الفئوية (categorical) والمستمرة (contineous) [3].



صورة من [4]

في الصورة أعلاه ، يمكنك مشاهدة تصنيف المجتمع (population) إلى أربع مجموعات مختلفة استنادًا إلى سمات متعددة لتحديد "إذا كانوا سيبيعون أم لا". لتقسيم المجتمع (population) إلى مجموعات مختلفة غير متجانسة ، فإنه يستخدم تقنيات مختلفة مثل Gini ، Gain Information ، Chi-square ، و entropy [4].

أفضل طريقة لفهم كيفية عمل شجرة القرارات ، هي لعب Jezzball - لعبة كلاسيكية من Microsoft (الصورة أدناه). أساسا ، لديك غرفة بها جدران متحركة وتريد إنشاء جدران لزيادة المساحة الخالية من الكرات.



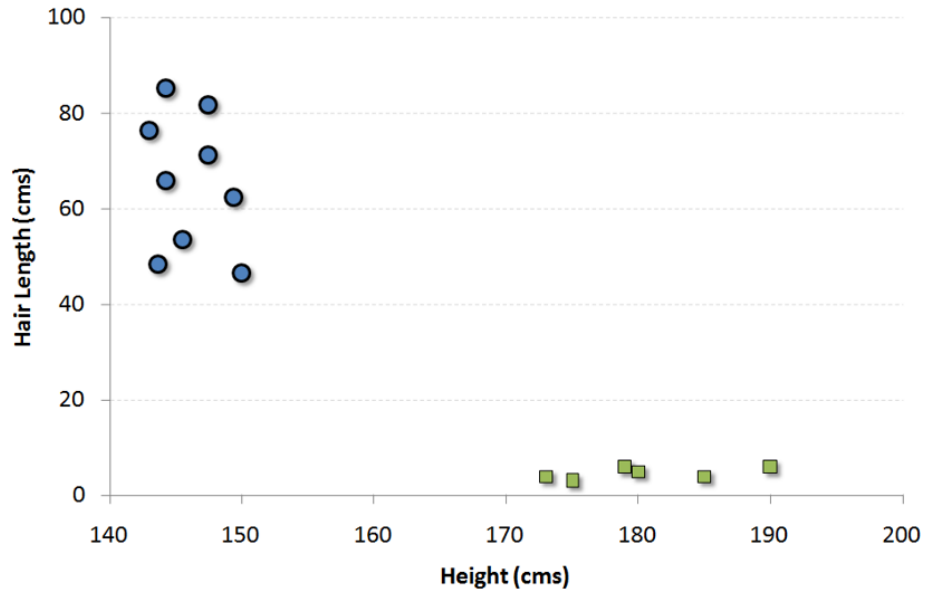
لقطة من لعبة Jezzball

لذا ، في كل مرة تقوم فيها بتقسيم الغرفة بجدار ، تحاول إنشاء مجموعتين مختلفتين في نفس الغرفة. تعمل أشجار القرار بطريقة مشابهة جدًا من خلال تقسيم المجتمع (population) في مجموعات مختلفة قدر الإمكان [4].

SVM (Support Vector Machine)

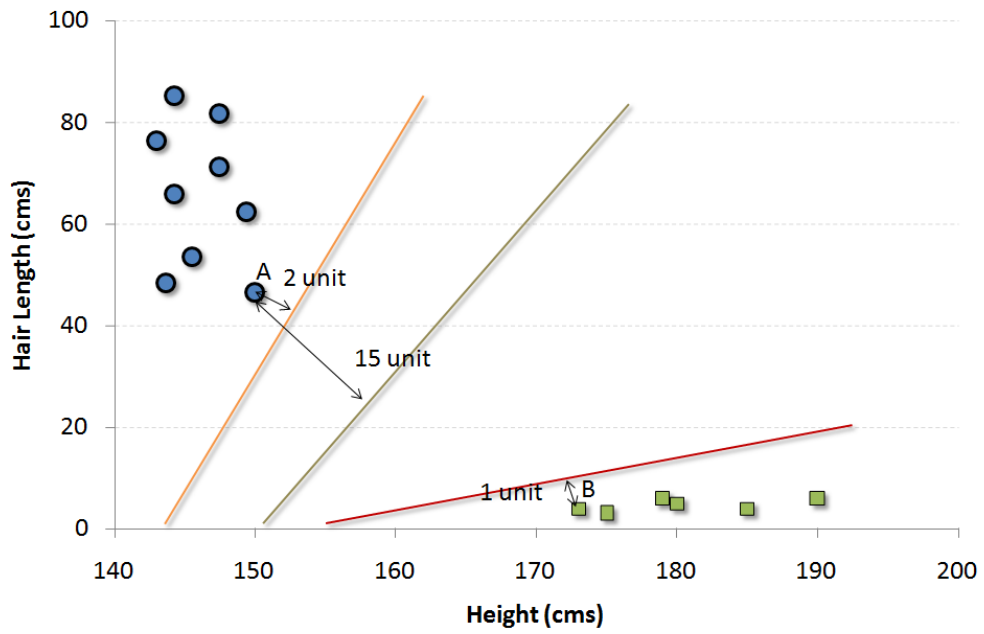
في هذه الخوارزمية ، نرسم كل عنصر بيانات كنقطة في فضاء (n-dimensional space) (حيث تمثل n عدد الميزات (features) التي لديك) مع قيمة كل ميزة هي قيمة إحداثي معين (particular coordinate)

على سبيل المثال ، إذا كان لدينا ميزتان فقط مثل الطول والطول الشعر للفرد ، فنرسم أولاً هذين المتغيرين في فضاء ثنائي الأبعاد حيث يكون لكل نقطة إثنين من إحداثيات (co-ordinates) (يُعرف هذه الإحداثيات باسم Vector Support)



الصورة من {Ray, 2017 #33}

الآن ، سوف نجد خط يقسم البيانات بين مجموعتين مصنفتين مختلفتين. سيكون هذا هو الخط بحيث تكون المسافات من أقرب نقطة في كل من المجموعتين أبعد.



الصورة من {Ray, 2017 #33}

في المثال أعلاه ، يكون الخط الذي يقسم البيانات إلى مجموعتين مختلفتين هو الخط الأسود ، لأن أقرب نقطتين هما أبعد ما تكون عن الخط. هذا الخط هو مصنفنا. بعد ذلك ، اعتمادًا على المكان الذي توجد فيه بيانات الاختبار على جانبي الخط ، ستصنف الفئة الجديدة عليه {Ray, 2017 #33}.

الغابة العشوائية (Random Forest)

الغابة العشوائية (Random Forest) هو مصطلح لمجموعة من أشجار القرار. في Random Forest ، لدينا مجموعة من أشجار القرار (المعروفة باسم "الغابة"). لتصنيف كائن جديد بناءً على سمات (attributes)، تعطي كل شجرة تصنيفًا ونقول أن الشجرة "صوتت (votes)" لتلك الفئة. تختار الغابة التصنيف الذي يضم أكبر عدد من الأصوات (على جميع الأشجار في الغابة).

كل شجرة تزرع وتتم كما يلي (Each tree is planted & grown as follows):

1. إذا كان عدد الحالات في مجموعة التدريب هو N ، يتم أخذ عينة N من الحالات عشوائياً ولكن مع الاستبدال (replacement). هذه العينة ستكون مجموعة التدريب لنمو الشجرة (growing the tree).
2. إذا كان هناك متغيرات دخل M ، فيتم تحديد رقم $m \ll M$ بحيث يتم تحديد المتغيرات m عند كل عقدة ، بشكل عشوائي من M ويتم استخدام أفضل تقسيم على m لتقسيم العقدة. يتم الحفاظ على قيمة m ثابتة أثناء نمو الغابة (forest growing).
3. تزرع كل شجرة إلى أقصى حد ممكن (Each tree is grown to the largest extent possible). لا يوجد تقليم (pruning).

تعتبر Random Forest طريقة تعلم جماعية (ensemble) للتصنيف والانحدار والمهام الأخرى التي تعمل من خلال بناء العديد من أشجار القرار في وقت التدريب وإخراج الفئة الأكثر شيوعاً (النموذج) في التصنيف، (أو متوسط التنبؤ في الانحدار) من الأشجار الفردية [3]

الشبكة العصبية (Neural Network)

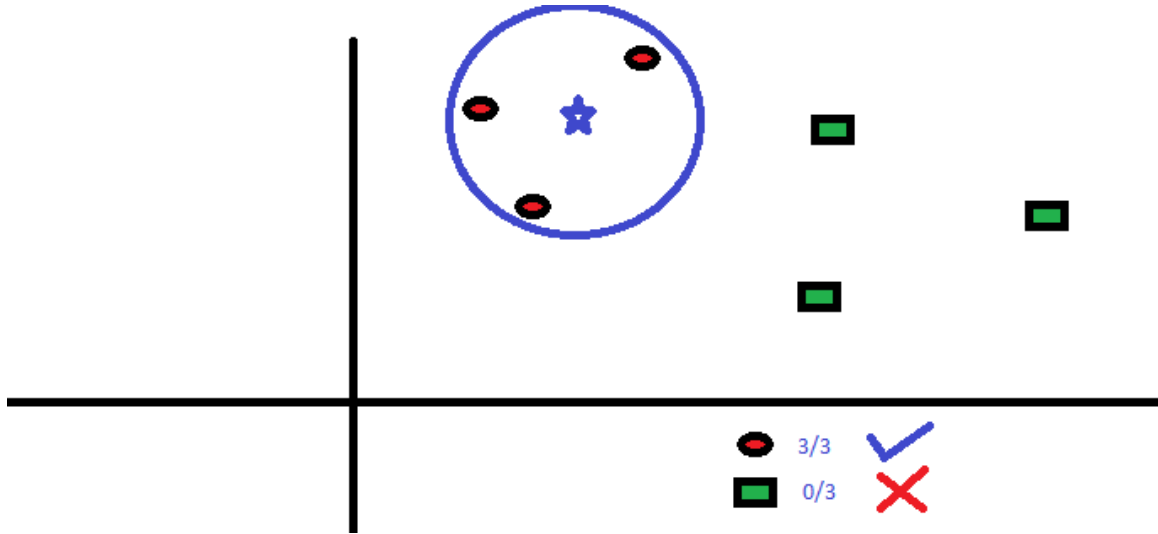
تتكون الشبكة العصبية من وحدات (عصبونات (neurons)) ، مرتبة في طبقات ، والتي تقوم بتحويل المتجه المدخل إلى بعض المخرجات. تأخذ كل وحدة مدخلات ، وتطبق وظيفة (غالبًا غير خطية) عليها ، ثم تمرر المخرجات إلى الطبقة التالية. بشكل عام ، يتم تعريف الشبكات على أنها تغذية أمامية (feed-forward): حيث تقوم الوحدة بتغذية مخرجاتها إلى جميع الوحدات في الطبقة التالية ، ولكن لا توجد أي تغذية راجعة (feedback) للطبقة السابقة. يتم تطبيق الوزن على الإشارات المارة من وحدة إلى أخرى (Weightings are applied to the signals passing from one unit to another)، ومن هذه الأوزان التي يتم ضبطها في مرحلة التدريب لتكييف الشبكة العصبية مع المشكلة المعنية الحالية [3].

اقرب جار (k- Nearest Neighbors) KNN

خوارزمية KNN تأخذ مجموعة من النقاط المعلمة (labeled points) وتستخدمها لتتعلم كيفية تسمية (label) النقاط الأخرى. لتسمية نقطة جديدة ، فإنها تنظر إلى النقاط المعلمة (labelled) الأقرب إلى تلك النقطة الجديدة (تلك هي أقرب الجيران) ، وتطلب من هؤلاء الجيران أن يصوتوا ، لذا فإن أي أكثر تسمية من الجيران تعتبر هي تسمية للنقطة الجديدة (K هو عدد الجيران الذي يقوم بالتحقق منه) [3].

يمكن استخدام KNN لكل من مشاكل التصنيف والانحدار. ومع ذلك ، فإنه يستخدم على نطاق أوسع في مشاكل التصنيف في الصناعة. وهو خوارزمية بسيطة تخزن كل الحالات المتاحة (available cases) وتصنف حالات جديدة بأغلبية أصوات جيرانها (majority vote of its k neighbors). الحالة التي يتم تخصيصها للفئة هي الأكثر شيوعًا بين الجيران الأقرب إلى K التي يتم قياسها بواسطة دالة المسافة.

يمكن أن تكون دوال المسافة Euclidean ، Manhattan ، Minkowski ، و Hamming distance. يتم استخدام الدوال الثلاثة الأولى للمتغيرات المستمرة (continuous) والرابعة (Hamming) للمتغيرات الفئوية (categorical variables). إذا كانت $K = 1$ ، فسيتم تعيين الحالة ببساطة لفئة أقرب جار لها. في بعض الأحيان ، يبدو أن اختيار K يمثل تحديًا أثناء تنفيذ نماذج kNN.



الصورة من {Ray, 2017 #33}

يمكن ربط KNN بسهولة بأمثلة حقيقية من حياتنا. إذا كنت تريد التعرف على شخص ، ليس لديك معلومات عنه ، فقد ترغب في معرفة أصدقاءه المقربين والدوائر التي ينتقل إليها للوصول إلى معلوماته.

أشياء يجب مراعاتها قبل تحديد KNN

- يعتبر KNN هو مكلف حوسبياً (KNN is computationally expensive)
- يجب أن تكون متغيرات normalized ، لان المتغيرات عالية المدى ستؤدي بنتائج متحيزة (higher range variables can bias it)
- العمل على مرحلة ما قبل المعالجة (pre-processing) أكثر قبل استخدام kNN (مثل إزالة الضجيج)

المراجع

1. Schapire, R. *Machine Learning Algorithms for Classification*. Available from: <http://www.cs.princeton.edu/~schapire/talks/picasso-minicourse.pdf>
2. Korbut, D. *Machine Learning Algorithms: Which One to Choose for Your Problem*. 2017; Available from: <https://blog.statsbot.co/machine-learning-algorithms-183cc73197c/>
3. Sifium. *Types of classification algorithms in Machine Learning*. 2017; Available from: <https://medium.com/@sifium/machine-learning-types-of-classification-9497bd4f2e14>
4. Ray, S. *Essentials of Machine Learning Algorithms (with Python and R Codes)*. 2017; Available from: <https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/>

الباب الخامس

التعليم دون اشراف

Unsupervised Learning

الباب الخامس

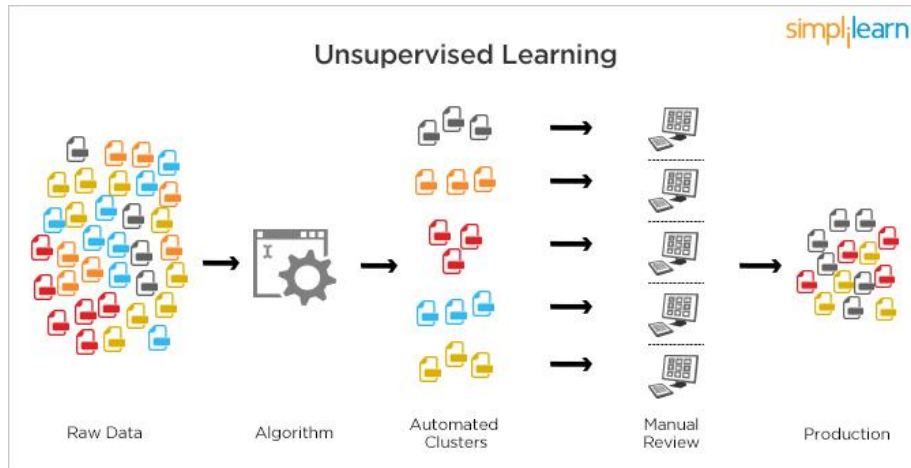
التعلم دون إشراف

UnSupervised Learning

نستحدث في هذا الباب عن التعلم دون إشراف، وتقصد به تعلم الآلة من غير بيانات تدريبية معلمة (مصنفة)، حيث تحاول الآلة هنا البحث في البيانات عن علاقات ومعرفة مخفية داخل البيانات.

تعريف

التعلم دون إشراف هو مقدرة الخوارزمية على استخلاص معرفة مخفية في البيانات، لذلك على النقيض من التعلم تحت الإشراف، فهي تتعامل مع بيانات غير معلمة (unlabeled data).



شكل رقم: التعلم بدون إشراف (المصدر [1]).

مهمتا التعلم دون إشراف هما:

- تجميع البيانات في مجموعات عن طريق التشابه (similarity)
- تقليل (تقليص) الأبعاد (reducing dimensionality) لضغط البيانات (compress) مع الحفاظ على هيكلها وفائدتها [2].

أهداف التعلم دون إشراف

الهدف في التعلم دون إشراف هو تجميع البيانات في مجموعات مختلفة. التعلم الآلي بدون إشراف هو **أكثر تحديًا** من التعلم تحت الإشراف نظرًا لغياب **المعلومات** (absence of labels). لو فصلنا هذا الهدف لوجدناه يسعى للإجابة والوصول إلى الأهداف التالية:

- كيف يمكنك العثور على البنية الأساسية لمجموعة بيانات؟
- كيف تلخصها وتجمعها بشكل مفيد؟
- كيف تمثل البيانات بفعالية وتنسيق مختصر؟

نفس البيانات يمكن تجميعها في مجموعات مختلفة بطرق مختلفة. مثلاً يمكن تجميع 16 حيواناً تم تمثيلها باستخدام 13 من الخصائص المنطقية (المظهر والنشاط)، لذلك يمكن تجميعها الى مجموعات حسب المظهر مرة، ويمكن تجميعها الى مجموعات مختلفة حسب النشاط مرة ثانية. ففي المرة الأولى تم تجميعها الى ثدييات وطيور (المظهر)، بينما تم تجميعها في المرة الثانية إلى حيوانات مفترسة (predators) وفرائس (preys) (النشاط).

تطبيقات التعلم بدون إشراف

بعض التطبيقات المعروفة على نطاق واسع للتعلم بدون إشراف [3] ، هي:

- بيع التجزئة بالسوق الذي يستهدف العملاء المناسبين.
- اكتشاف الشذوذ / الاحتيال (anomaly/fraud detection) في القطاع المصرفي
- تجزئة الصور (image segmentation)
- تجميع الجينات حسب مستويات التعبير الجيني (gene clustering for grouping gene with similar expression levels)
- استخلاص مؤشرات مناخية (deriving climate indices) مبنية على تجميع بيانات علوم الأرض (earth science data)
- تجميع الوثائق حسب المحتوى.
- تطبيق مرتبط بعلم البيئة، حيث يتم استخدام (تقنيات التجميع) لتجميع التسجيلات الصوتية التي يتم التقاطها من خلال الميكروفون الموضوع في أماكن محددة في المنطقة محل الاهتمام. ثم يتم تحليل هذه التسجيلات باستخدام تقنيات تعلم دون إشراف لقياس التنوع البيولوجي ، مثل عدد أنواع الطيور والحيوانات ، في المنطقة ذات الأهمية.

طرق التعلم دون إشراف

هنالك العديد من الطرق التي تستخدم في التعلم دون إشراف [2] ، منها :

Clustering and dimensionality reduction:

- k-means clustering
- hierarchical clustering
- principal component analysis (PCA)
- singular value decomposition (SVD)

التجميع (Clustering)

ومن الأمثلة المثيرة للاهتمام للتجميع في العالم الحقيقي نظام "أكسيسوميكس" المتخصص في مجال تسويق البيانات (Acxiom's life stage clustering system) الخاص بشركة "أكسيوم". تقسم هذه الخدمة الأسر الأمريكية إلى 70 مجموعة مميزة في 21 مجموعة من مراحل الحياة التي يستخدمها المعلنون عند استهداف إعلانات Facebook والإعلانات الصورية وحملات البريد المباشر وما إلى ذلك [2].

تجميع k-means

الهدف من التجميع هو إنشاء مجموعات من نقاط البيانات بحيث تكون النقاط في مجموعات مختلفة غير متشابهة بينما تكون النقاط داخل مجموعة متشابهة.

في تجميع k-means، فإننا نريد تجميع نقاط البيانات الخاصة بنا في مجموعات عددها k مجموعة. كلما زادت عدد المجموعات (زيادة قيمة k) كلما كانت المجموعات صغيرة (مع مزيد من التفاصيل (more granularity))، وكلما قلت عدد المجموعات (قيمة k صغيرة) هذا يعني مجموعات أكبر و أقل دقة [2].

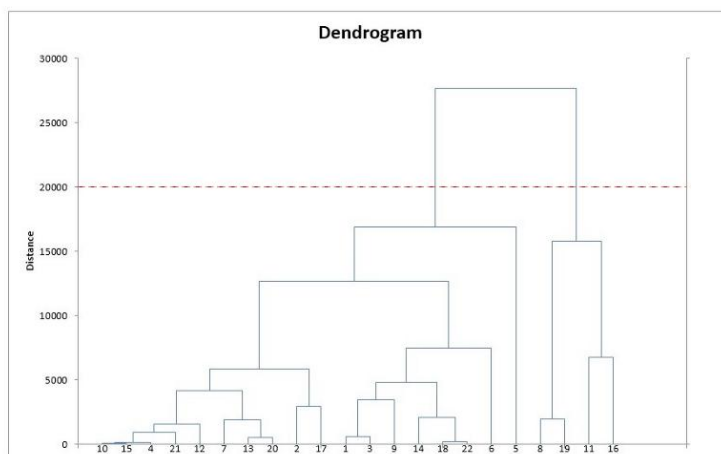
من التطبيقات الحقيقية التي تستخدم k-means هو تصنيف الأرقام المكتوبة بخط اليد (classifying handwritten digits). لنفترض أن لدينا صورًا للأرقام كمتجه لسطوع البيكسل (vector of pixel brightnesses). لنفترض أن الصور سوداء وبيضاء وحجمها 64x64 بكسل. كل بكسل يمثل بعد (dimension). لذلك فإن هذه الصور توجد في $(64 \times 64 = 4,096)$ بعدًا. في هذه الأبعاد 4,096، يسمح لنا تجميع k-means بتجميع الصور التي تكون قريبة من بعضها البعض، ونفترض أنها تمثل نفس الرقم، والذي يمكن أن يحقق نتائج جيدة للتعرف على الأرقام [2].

===

هو نوع من الخوارزميات التي تحل مشكلة التجميع. يتبع الإجراء الخاص به طريقة بسيطة وسهلة لتصنيف مجموعة بيانات معينة من خلال عدد معين من المجموعات (افترض مجموعات k). تكون نقاط البيانات داخل مجموعة متجانسة (homogeneous) ومتباينة (heterogeneous) مع المجموعات الأخرى [4].

التجميع الهرمي (Hierarchical clustering)

التجميع الهرمي يشبه التجميع العادي، إلا أنك تهدف إلى إنشاء تسلسل هرمي للتجمعات. يمكن أن يكون ذلك مرئياً أكثر في تحديد عدد المجموعات التي تريدها في النهاية. على سبيل المثال، تخيل عناصر التجميع في سوق عبر الإنترنت مثل Etsy أو Amazon. في الصفحة الرئيسية، تحتاج إلى بضع فئات واسعة من العناصر للتنقل البسيط (broad categories of items for simple navigation)، ولكن أثناء انتقالك إلى فئات تسوق أكثر تحديداً، ستحتاج إلى زيادة مستويات الدقة، أي مجموعات أكثر تميزاً للعناصر [2].



شكل رقم 5-1: التجميع الهرمي [2] (المصدر Solver.com)

تقليل الأبعاد (Dimensionality reduction)

تقليل الأبعاد يشبه إلى حد كبير الضغط (compression). يتعلق الأمر بمحاولة تقليل تعقيد البيانات مع الاحتفاظ بأكبر قدر ممكن من البنية (structure) ذات الصلة. إذا كنت تأخذ صورة بحجم $(3 \times 128 \times 128)$ نقطة (بكسل) أي 49,152 بعدا. إذا كنت قادراً على تقليل بُعد المساحة التي توجد بها هذه الصور دون تغيير كثير في المحتوى (مغزى الصورة) ، فحينئذ تكون قد قمت بتقليل الأبعاد [2].

في السنوات الأربع أو الخمس الماضية ، حدثت زيادة هائلة في البيانات التي يتم التقاطها في كل مرحلة ممكنة. لا تأتي الشركات / الوكالات الحكومية / المؤسسات البحثية بمصادر جديدة فحسب ، بل إنها تلتقط البيانات بتفاصيل دقيقة وكثيرة..

على سبيل المثال: تستحوذ شركات التجارة الإلكترونية على مزيد من التفاصيل حول العملاء مثل المعلومات السكانية (demographics)، وتاريخ تصفح الويب (web crawling history)، وما يعجبهم وما لا يعجبهم، وسجلات الشراء ، والتعليقات وغيرها الكثير، لتوجيه اهتمام خاص لهم أكثر من أقرب صاحب بقالة.

وباعتباري عالم بيانات ، فإن البيانات التي نقدمها تتكون أيضاً من العديد من الميزات (features)، وهذا يبدو جيداً لبناء نموذج قوي جيد ولكن هناك تحدٍ (challenge). كيف تحدد المتغير (المتغيرات) الهامة للغاية في 1000 أو 2000؟ في مثل هذه الحالات ، تساعدنا خوارزمية تقليل الأبعاد (dimensionality reduction) مع خوارزميات أخرى مثل PCA، Random Forest، Decision Tree، Factor Analysis، Identify، missing value ratio، based on correlation matrix، وغيرها [4] .

أخيراً

"نتوقع أن يصبح التعلم دون إشراف أكثر أهمية على المدى الطويل، لأن التعلم البشري والحيواني غير خاضع لإشراف إلى حد كبير: نحن نكتشف بنية العالم من خلال رصده ، وليس من خلال إخبارنا باسم كل كائن." (LeCun, Bengio, Hinton, Nature (2015))

المراجع

1. and Priyadharshini. *Machine Learning: What it is and Why it Matters*. 2018; Available from: <https://www.simplilearn.com/what-is-machine-learning-and-why-it-matters-article>.
2. Available from: Maini, V. *Machine Learning for Humans, Part 3: Unsupervised Learning*. 2017. <https://medium.com/machine-learning-for-humans/unsupervised-learning-f45587588294>.
3. Kumar, S. *What is an example application of unsupervised machine learning*. 2018; Available from: <https://www.quora.com/What-is-an-example-application-of-unsupervised-machine-learning>.
4. Ray, S. *Essentials of Machine Learning Algorithms (with Python and R Codes)*. 2017; Available from: <https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms>.

الباب السادس

تعلم الآلة شبه الخاضع للإشراف

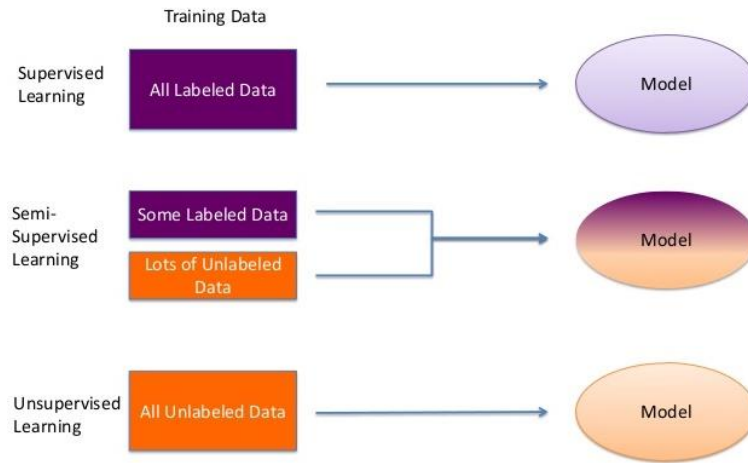
Semi-Supervised Learning

الباب الخامس

تعلم الآلة شبه الخاضع للإشراف

Semi-Supervised Learning (SSL)

نستحدث في هذا الباب عن احد تعلم الآلة شبه تحت الاشراف، وللاختصار سنستخدم الحروف الاولى من اسمه الانجليزي وهي SSL في بقية الباب. لقد عرفنا التعلم تحت الاشراف والتعلم من دون اشراف، وعرفنا أن الفرق الرئيسي بينهما أن الاول يتعامل ويتدرب على بيانات معلمة (labelled data) بينما الثاني يتدرب على بيانات غير معلمة (unlabelled data). أم النوع الثالث (SSL) فهو يتعامل مع بيانات خليط (جزء معلم وجزء غير معلم).



شكل رقم 1-6: نوع بيانات التدريب في أنواع التعلم الثلاثة [1]

أهمية SSL

يتم تدريب خوارزميات التعلم شبه الخاضعة للإشراف على مجموعة من البيانات المعلمة وغير المعلمة (labeled and unlabeled data). هذا مفيد لعدة أسباب:

1. غالباً ما تكون عملية وجود كميات هائلة من البيانات المعلمة مكلفة ومحددة للوقت للتعلم تحت الإشراف.
2. وجود الكثير من الملاحظات (too much labeling) يفرض تحيزات بشرية على النموذج (can impose human biases on the model).
3. وهذا يعني أن الكثير من البيانات غير المسموح بها أثناء عملية التدريب تقبل في الواقع إلى تحسين دقة النموذج النهائي مع تقليل الوقت والتكلفة التي يتم إنفاقها.

ولهذه الأسباب، فإن SSL يكون مناسباً في بعض الحالات مثل تصنيف صفحات الويب أو التعرف على الكلام أو التسلسل الجيني. في جميع هذه الحالات، يمكن لعلماء البيانات الوصول إلى كميات كبيرة من البيانات غير المعلمة (unlabeled)، ولكن عملية تخصيص معلومات إشراف فعلي لكل ذلك ستكون مهمة لا يمكن التغلب عليها [2].

من استخدامات SSL

يكون استخدام SSL مفيد في بعض الحالات مثل:

- Webpage classification
- Speech recognition
- Genetic sequencing

مقارنة عملية للثلاث انواع

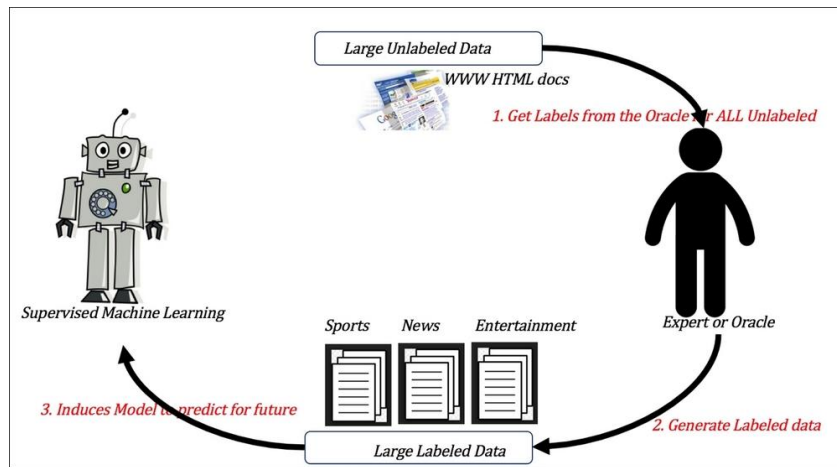
Using classification as an example, let's compare how these three approaches work in practice [2]:

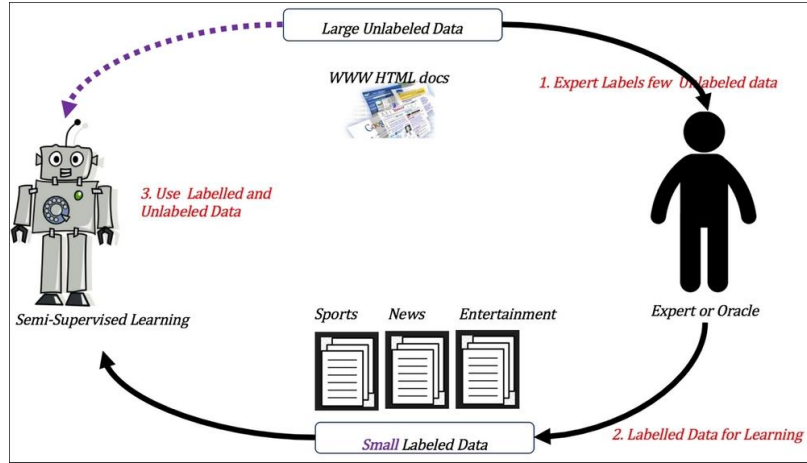
- **Supervised classification:** The algorithm learns to assign labels to types of webpages based on the labels that were inputted by a human during the training process.
- **Unsupervised clustering:** The algorithm looks at inherent similarities between webpages to place them into groups.
- **Semi-supervised classification:** Labeled data is used to help identify *that* there are specific groups of webpage types present in the data and what they *might be*. The algorithm is then trained on unlabeled data to define the boundaries of those webpage types and may even identify new types of webpages that were unspecified in the existing human-inputted labels.

كيف يعمل SSL

تفسير مفهوم SSL بمثال بسيط.

يتعلم SSL من بيانات معلمة وبيانات غير معلمة لتحسين القدرة التنبؤية للنماذج. سنوضح الفكرة في الشكل 2-6. فعندما تتوفر كمية كبيرة من البيانات غير المعلمة (على سبيل المثال ، مستندات HTML على الويب) ، يمكن للخبر تصنيف القليل منها إلى فئات معروفة مثل الرياضة والأخبار. والترفيه ، وما إلى ذلك. ويمكن بعد ذلك لـ SSL استخدام هذه المجموعة الصغيرة من البيانات المصنفة مع مجموعة البيانات الكبيرة غير المعلمة ، لمعرفة النماذج. وبالتالي ، وباستخدام معرفة كل من البيانات المعلمة وغير المعلمة ، يمكن للنموذج تصنيف الوثائق غير المرئية (unseen documents) في المستقبل [3].





الشكل رقم 6-2: عملية تعلم SSL (الأسفل) مع التعليم تحت الإشراف (الأعلى) باستخدام تصنيف مستندات الويب كمثال. والفرق الرئيسي هو مقدار البيانات المعلمة المتاحة للتعلم ، والتي تظهر بـ "small" في حالة SSL [3].

المراجع

1. Jain, S. *Introduction to Pseudo-Labeling : A Semi-Supervised learning technique*. 2017; Available from: [./https://www.analyticsvidhya.com/blog/2017/09/pseudo-labelling-semi-supervised-learning-technique](https://www.analyticsvidhya.com/blog/2017/09/pseudo-labelling-semi-supervised-learning-technique)
2. Castle, N. *What is Semi-Supervised Learning*. 2018; Available from: <https://www.datascience.com/blog/what-is-semi-supervised-learning>
3. *Semi-supervised learning*. Available from: https://www.packtpub.com/mapt/book/big_data_and_business_intelligence/9781785880513/4/ch04lvl1sec34/semi-supervised-learning
4. Sifium. *Types of classification algorithms in Machine Learning*. 2017; Available from: <https://medium.com/@sifium/machine-learning-types-of-classification-9497bd4f2e14>

الباب السابع

التعلم التعزيزي

Reinforcement Learning

الباب السابع

تعلم الآلة التعزيزي

Reinforcement Learning (RL)

نستحدث في هذا الباب عن تعلم الآلة التعزيزي والذي تتعلم فيه الخوارزمية بالتجربة من البيئة. معظم ما في الباب مترجم من [1]. يمكنك الرجوع لربط النسخة الاصلية في المراجع.

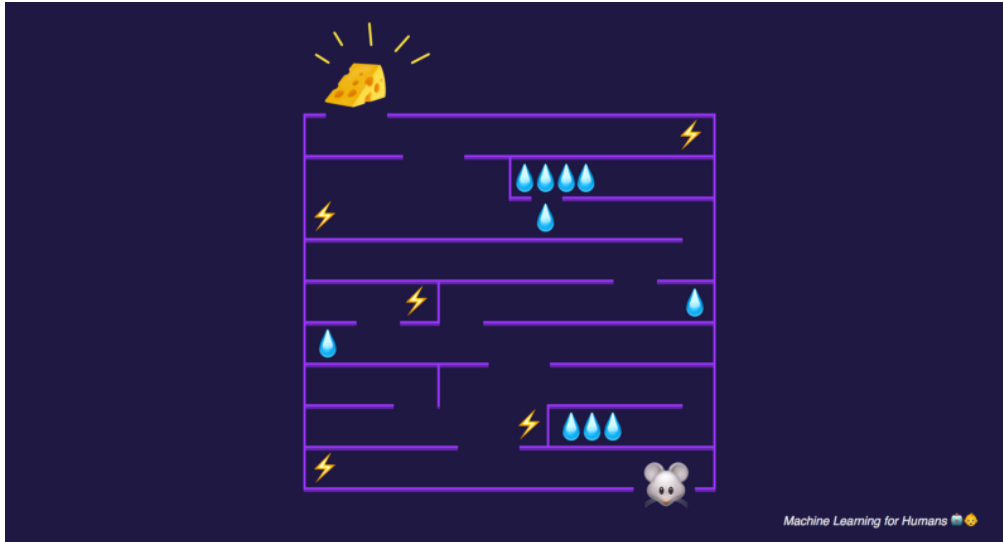
تعريف

في التعلم التعزيزي (RL) لا يوجد مفتاح للإجابة ، ولكن وكيل تعلم التعزيز الخاص بك يتعين عليه أن يقرر كيفية التصرف لأداء مهمته (has to decide how to act to perform its task). في غياب بيانات التدريب على الوكيل التعلم بالتجربة (الخبرة). حيث يقوم بتجميع الأمثلة التدريبية ويعرف ماهو الإجراء الجيد وماهو الإجراء السيئ من خلال التجربة، وأثناء هذه المحاولات يكون هدف الوكيل زيادة المكافأة على المدى الطويل (maximizing long-term reward) [1].

لعبة الفأر الروبوت

إن ايسر طريقة لتوضيح فكرة التعلم التعزيزي هو مقارنته بلعبة ذات هدف واضح ومكافآت على شكل نقاط.

لنفترض أننا نلعب لعبة يسعي فيها فأر إلى الحصول على المكافأة النهائية من الجبن في نهاية المتاهة (إذا احرز 1000 نقطة) ، أو جمع مكافآت الماء على طول الطريق (10 نقاط لكل واحدة). في الوقت نفسه ، يريد الفأر تجنب المواقع التي فيها صدمة كهربائية (فنتفده 100 نقطة (يحصل على -100 نقطة)) [1].



شكل رقم 1-7 : لقطة من لعبة الفأر الروبوت [1]

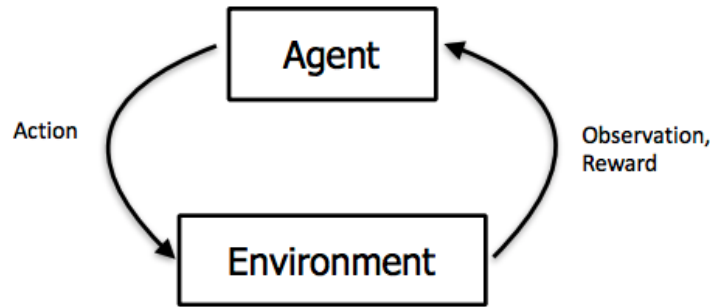
بعد قليل من الاستكشاف (exploration) ، قد يجد الفأر مكان به ثلاثة مصادر مائية متجمعة بالقرب من المدخل ، ويقضي كل وقته في استغلال (exploiting) هذا الاكتشاف من خلال الحصول باستمرار على المكافآت الصغيرة لمصادر المياه هذه وعدم الذهاب إلى داخل المتاهة (maze) لمتابعة (pursue) الجائزة الأكبر.

ولكن كما ترون ، فإن الفأر سيفوت عليه واحة أفضل في المتاهة ، أو المكافأة النهائية للجبن في النهاية!

هذا يحضر تبادل (tradeoff) الاستكشاف / الاستغلال. تتمثل إحدى الاستراتيجيات البسيطة للاستكشاف في أن يتخذ الماوس أفضل إجراء (best action) معروف في معظم الأوقات (على سبيل المثال ، 80٪ من الوقت) ، ولكن أحياناً استكشاف اتجاه جديد تم اختياره عشوائياً على الرغم من أنه قد يكون بعيداً عن مكافأة معروفة.

وتسمى هذه الاستراتيجية استراتيجية الجشع (epsilon-greedy strategy) ، حيث ϵ هي النسبة المتوقعة من الوقت الذي يتخذ فيه الوكيل إجراءً تم اختياره عشوائياً بدلاً من اتخاذ الإجراء الذي يُرجح أن يزيد المكافأة في ضوء ما يعرفه حتى الآن (في هذه الحالة ، 20 ٪). نبدأ عادة بالكثير من الاستكشاف (أي قيمة أعلى لـ ϵ) مع مرور الوقت ، عندما يتعلم الفأر المزيد عن المتاهة وأنها تحقق أفضل مكافأة على المدى الطويل ، سيكون من المنطقي خفض ϵ بشكل مطرد إلى 10٪ أو أقل حتى يستقر (settles) في استغلال ما يعرفه.

من المهم أن تضع في اعتبارك أن المكافأة ليست دائماً فورية، ففي هذا المثال قد يكون هناك ممر طويل في المتاهة يجب عليك المرور خلاله والعديد من نقاط القرار قبل الوصول إلى الجبن.



شكل رقم 2-7 : يراقب الوكيل البيئة ، ويتخذ إجراء للتفاعل معها ويتلقى مكافأة إيجابية أو سلبية . [2]

Markov Decision Processes (MDPs)

يمكن إضفاء الطابع الرسمي على تجوال الفأر عبر المتاهة كعملية قرار ماركوف (Markov Decision Process) ، وهي عملية حددت احتمالات الانتقال من حالة إلى أخرى. سنشرح ذلك بالإشارة إلى مثالنا على الفأر. تشمل MDPs:

..A finite set of states

هذه هي المواقف المحتملة للفأر داخل المتاهة (maze):

مجموعة من الإجراءات (a set of actions) المتاحة في كل حالة. الحركة {back, forward} في الممر (corridor) والحركة

{right, left, back, forward} عند مفترق الطرق (crossroads).

التحولات بين الحالات (Transitions between states). على سبيل المثال ، إذا اتجهت شمالاً في مفترق طرق فستنتهي بموقع جديد.

يمكن أن تكون هذه مجموعة من الاحتمالات التي ترتبط بأكثر من حالة واحدة ممكنة (على سبيل المثال عند استخدامك لهجوم في لعبة بوكيمون (Pokémon) يمكنك إما أن تخطئ، أو تلحق (inflict) بعض الضرر ، أو تلحق ضرراً كافياً لتخطي خصمك (opponent)).

المكافآت المرتبطة بكل عملية انتقال. في مثال الفأر ، فإن معظم المكافآت هي 0 ، لكنها إيجابية إذا وصلت إلى نقطة تحتوي على ماء أو جبن ، وسلبية إذا وصلت إلى نقطة بها صدمة كهربائية (electric shock).

عامل خصم γ بين 0 و 1. هذا يحدد (quantifies) الفرق في الأهمية بين المكافآت الفورية والمكافآت المستقبلية. على سبيل المثال ، إذا كانت γ تساوي 9 ، وهناك مكافأة قدرها 5 بعد 3 خطوات ، فإن القيمة الحالية لهذه المكافأة هي 5×0.9^3 .

Memorylessness: بمجرد معرفة الحالة الحالية ، يمكن مسح تاريخ انتقال الفأر عبر المناهضة ، لأن حالة ماركوف الحالية تحتوي على جميع المعلومات المفيدة من التاريخ (because the current Markov state contains all useful information from the history). بعبارة أخرى : "المستقبل مستقل عن الماضي المعطى للحاضر" (the future is independent of the past given the present).

الآن بعد أن عرفنا ما هو MDP ، يمكننا إضفاء الطابع الرسمي على هدف الفأر (formalize the mouse's objective). نحن نحاول زيادة مجموع المكافآت على المدى الطويل:

$$\sum_{t=0}^{t=\infty} \gamma^t r(x(t), a(t))$$

في المعادلة اعلاه: نحن نجمع في جميع الخطوات الزمنية t . فلنضع y بـ 1 في الوقت الحالي وننسى ذلك. $r(x,a)$ هي دالة المكافأة.

للحالة x والإجراء a (أي الذهاب إلى اليسار عند مفترق الطرق) ، يعطيك المكافأة المرتبطة باتخاذ هذا الإجراء a على الحالة x .

بالعودة إلى المعادلة ، نحاول زيادة مجموع المكافآت المستقبلية من خلال اتخاذ أفضل إجراء في كل حالة. والآن بعد أن أنشأنا مشكلة التعلم التعزيزية وأضفنا الطابع الرسمي على الهدف ، دعنا نستكشف بعض الحلول الممكنة

Q-learning: learning the action-value function

Q-learning هو تقنية تقييم الإجراء الذي يجب اتخاذه بناءً على دالة قيمة إجراء (action-value function)، تحدد قيمة وجودك في حالة معينة، وتتخذ إجراءً معيناً في تلك الحالة.

لدينا دالة Q التي تأخذ حالة واحدة كمدخل وإجراء واحد وإرجاع المكافأة المتوقعة من هذا الإجراء (وجميع الإجراءات التابعة (action-value function)) في تلك الحالة.

قبل أن نستكشف البيئة ، تعطي Q قيمة ثابتة (اعتباطية) (Q gives the same (arbitrary) fixed value).

ولكن بعد ذلك ، بينما نستكشف البيئة أكثر ، تعطي Q تقريب أفضل وأفضل (Q gives us a better and better approximation) لقيمة إجراء a عند الحالة s . سنقوم بتحديث الدالة Q كما تقدمنا (We update our function Q as we go).

المعادلة التالية (Q-learning) توضح كيف نقوم بتحديث قيمة Q بناءً على المكافأة التي نحصل عليها من بيئتنا:

$$Q(s_t, a_t) \leftarrow \underbrace{Q(s_t, a_t)}_{\text{old value}} + \underbrace{\alpha}_{\text{learning rate}} \cdot \left(\underbrace{r_t}_{\text{reward}} + \underbrace{\gamma}_{\text{discount factor}} \cdot \underbrace{\max_a Q(s_{t+1}, a)}_{\text{estimate of optimal future value}} - \underbrace{Q(s_t, a_t)}_{\text{old value}} \right)$$

دعنا نتجاهل عامل الخصم γ عن طريق تعيينه إلى 1 مرة أخرى. أولاً ، ضع في اعتبارك أن Q من المفترض أن تظهر لك مجموع المكافآت من اختيار الإجراء Q وكافة الإجراءات المثلث بعد ذلك.

لننتقل الآن إلى المعادلة من اليسار إلى اليمين. عندما نتخذ إجراء في الحالة st ، نقوم بتحديث قيمنا Q(st,at) بإضافة مصطلح إليها. هذا المصطلح يحتوي على:

- معدل التعلم ألفا (Learning rate alpha): هذه هي الطريقة العدوانية (aggressive) التي نريد أن نكون عليها عند تحديث قيمنا. عندما تكون alpha قريبة من 0 ، فإننا لا نحدث بشكل كبير جدًا (updating very aggressively). عندما تكون alpha قريبة من 1 ، سنستبدل القيمة القديمة بالقيمة المحدثة.
- المكافأة (reward): هي المكافأة التي حصلنا عليها من خلال اتخاذ إجراءات في الحالة at. لذلك نضيف هذه المكافأة إلى تقديرنا القديم.
- نضيف أيضًا المكافأة المستقبلية المقدرة (estimated future reward)، وهي الحد الأقصى للقيمة التي يمكن تحقيقها للمكافأة Q لكل الإجراءات المتاحة على $xt + 1$.
- أخيرًا ، نطرح القيمة القديمة لـ Q للتأكد من أننا فقط نزيد أو نخفض حسب الاختلاف في التقدير (مضروبًا في ألفا (alpha) بالطبع).

الآن وبعد أن قمنا بتقدير القيمة لكل زوج من حالة-إجراء (state-action pair)، يمكننا اختيار الإجراء الذي يجب اتخاذه وفقًا لاستراتيجية اختيار الاجراء (نحن لا نختار بالضرورة الإجراء الذي يؤدي إلى أكثر المكافآت المتوقعة في كل مرة ، على سبيل المثال استراتيجية استكشاف (epsilon-greedy exploration) كنا نتخذ إجراء عشوائي (a random action) بعض النسبة المئوية من الوقت).

في مثال فأر الروبوت ، يمكننا استخدام Q-learning لمعرفة قيمة كل موضع في المناهة وقيمة الإجراءات {right, left, backward, forward} في كل موضع. بعد ذلك ، يمكننا استخدام استراتيجية اختيار الاجراء الخاصة بنا لاختيار ما يفعله الفأر بالفعل في كل خطوة زمنية.

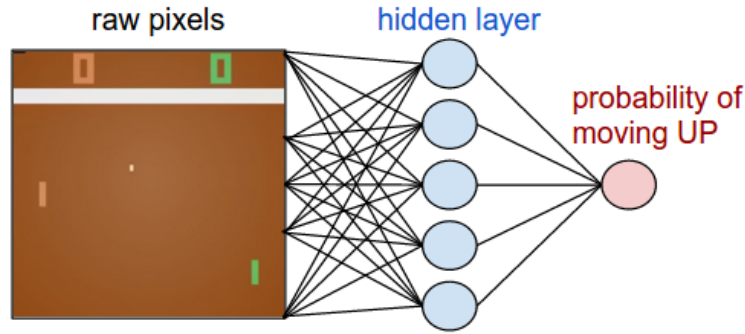
Policy learning: a map from state to action

في نهج Q-learning ، تعلمنا دالة قيمة تقدر قيمة كل زوج اجراء-حالة (state-action pair). إن سياسة التعلم (Policy learning) هي بديل أكثر وضوحًا نتعلم فيه سياسة الدالة (policy function) ، π ، وهي عبارة عن ربط مباشرة من كل حالة إلى أفضل إجراء مماثل في تلك الحالة. فكر في الأمر على أنه سياسة سلوكية (behavioral policy): "عندما ألاحظ الحالة s ، فإن أفضل شيء نفعله هو اتخاذ إجراء a". على سبيل المثال ، يمكن أن تتضمن سياسة السيارة المستقلة (autonomous vehicle's policy) فعليًا شيئًا مثل: "إذا رأيت ضوءًا أصفر وأنا على بعد أكثر من 100 قدم من التقاطع ، فيجب أن أتوقف، وإلا ساستمر في التقدم"

$$a = \pi(s)$$

A **policy** is a map from **state** to **action**.

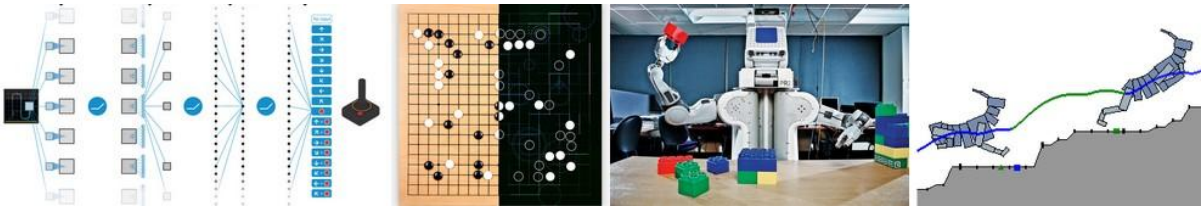
لذلك نحن نتعلم دالة من شأنها زيادة المكافأة المتوقعة. ما الذي نعرفه جيدًا في تعلم الدالة المعقدة؟ الشبكات العصبية العميقة (Deep neural networks). يقدم Andrej Karpathy's Pong من Pixels شرحًا رائعًا حول استخدام التعلم التعزيزي العميق (deep reinforcement learning) لتعلم سياسة Atari game Pong التي تأخذ البكسلات الخام من اللعبة كمدخل (حالة) وتخرج احتمالية لتحريك المضرب (paddle) إلى أعلى أو لأسفل (إجراء (action)) [1].



شكل رقم 7-3.

In a policy gradient network, the agent learns the optimal policy by adjusting its weights through gradient descent based on reward signals from the environment. Image via <http://karpathy.github.io/2016/05/31/rl/>

أمثلة RL .



Examples of RL in the wild. From left to right: Deep Q Learning network playing ATARI, AlphaGo, Berkeley robot stacking Legos, physically-simulated quadruped leaping over terrain [3]

المراجع

1. Maini, V. *Machine Learning for Humans, Part 5: Reinforcement Learning*. 2017; Available from: <https://medium.com/machine-learning-for-humans/reinforcement-learning-6eacf258b265>
2. Available from: John Schulman, P.A. *CS 294: Deep Reinforcement Learning, Fall 2015*. 2015
./rll.berkeley.edu/deeprlcourse-fa15
3. Karpathy, A. *Deep Reinforcement Learning: Pong from Pixels*. 2016; Available from: <http://karpathy.github.io/2016/05/31/rl>

الباب الثامن

تطبيقات تعلم الآلة

الباب الثامن

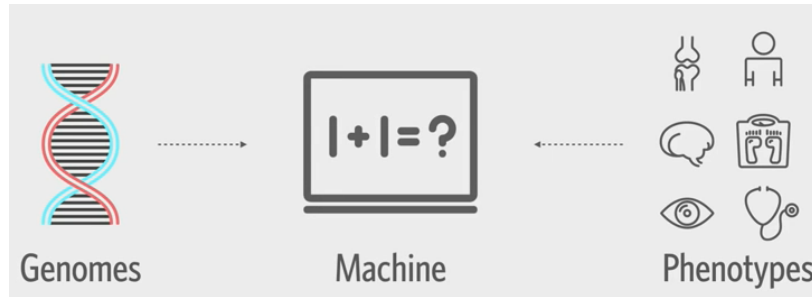
تطبيقات تعلم الآلة

تعلم الآلة الآن موجود في كل مكان، وانت غالبا ماتستخدمه بطريقة أو أخرى، حتى دون ان تعرف ذلك. هنالك الكثير من تطبيقات التعلم الآلي الشائعة. سنذكر في هذا الباب أمثلة لبعض منها.

التنبؤ بشكل الانسان [3]

يتكون الحمض النووي للانسان من ثلاثة مليار زوج من الحروف الاربعة (A,C,T,G). هذه المليارات تحمل الكثير من المعلومات عن الانسان وتعتبر من البيانات الضخمة التي تستفيد منها خوارزميات تعلم الآلة لاستخراج الكثير من المعارف.

هنالك فريق من 40 عالما من علماء البيانات وغيرهم من الباحثين، يسعون لاستخدام تعلم الآلة للتنبؤ بشكل الانسان من خلال حمضه النووي. حيث تم استخدام آلاف التسلسلات من الجينوم البشري، وربطها مع بيانات اخرى عن البشر تشمل (phenotypes, 3D scan, NMR, etc) في قواعد بيانات ضخمة. وجعل تعلم الآلة تتدرب عليها وتربط بينها بعلاقات وتنشئ نماذج تستخدمها في رسم شكل الشخص من خلال حمضه النووي.



تعلم الآلة من العلاقات بين ال genomes وال phenotypes

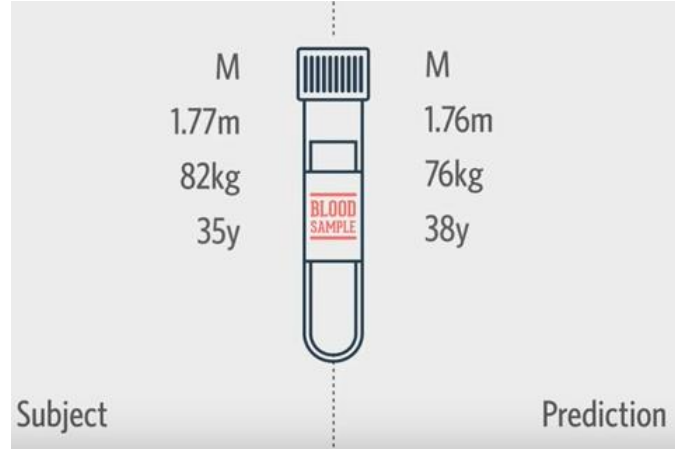
يسعى هذا المشروع لجعل الآلة (ومن خلال ربطها للعلاقات بين الجينوم البشري والصفات البشرية الاخرى التي تم جمعها) من القدرة على التنبؤ بصفات اي شخص من خلال حمضه النووي. مثل الطول، لون العينين، لون الجسد، العمر، شكل الوجه، ... الخ.

هذا التنبؤ كان نتيجة نموذج استطاعت الآلة بناءه من البيانات الضخمة التي تم استخدامها.

وتوصلت الابحاث الى التنبؤ بشكل الوجه من خلال التسلسل الجيني للشخص. وكانت نتيجة التطابق عالية جدا، كما مبين ادناه:



الشبه بين الوجه الحقيقي والوجه الذي تنبأت به الآلة



مقارنة بين بيانات شخص حقيقية وماتنبأت به الآلة.

أمثلة أخرى للتنبؤ

من أهم مجالات تعلم الآلة التنبؤ بالاحداث قبل وقوعها. ويستخدم في ذلك بيانات تاريخية وخبرات سابقة ومعلومات حالية. من الأمثلة التي تم فيها استخدام التنبؤ ما يلي:

- **التنبؤ في الانتخابات:** تم استخدام تعلم الآلة في السياسة بكثرة، مثل استخدام تعلم الآلة في انتخابات الرئاسة الأمريكية حيث تم الاستفادة من بيانات الناخبين وأفكارهم التي تم جمعها من وسائط التواصل الاجتماعي وغيرها من المصادر لمعرفة المرشح المتوقع فوزه في الانتخابات [4].
- **سداد القروض:** تستطيع المؤسسة أو البنك مقدم القرض استخدام تعلم الآلة للتنبؤ بما إذا كان العميل سيسدد القرض أم لا، وذلك بالاستفادة من البيانات المتاحة وتحليلها.
- **التنبؤ بالمبيعات:** في البيع بالتجزئة، في وقت سابق كنا قادرين على الحصول على رؤى مثل تقرير المبيعات في الشهر الماضي / السنة / 5 سنوات (تقارير تاريخية). ولكن الآن وباستخدام تعلم الآلة يمكننا التنبؤ بمبيعات الشهر المقبل أو السنة المقبلة، إلخ. بحيث يمكن أن تتخذ الشركة التجارية القرار المطلوب (المتعلق بالمشتريات والأسهم وغيرها) في الوقت المناسب [2].
- **التنبؤات المرورية:** معظمنا يستخدم حالياً خدمات الملاحة GPS. بينما نفعل ذلك، يتم حفظ مواقعنا الحالية وسرعاتنا في خادم مركزي لإدارة حركة المرور. ثم يتم استخدام هذه البيانات لبناء خريطة لحركة المرور الحالية. في حين أن هذا يساعد في حماية حركة المرور وتحليل الازدحام (preventing the traffic and does congestion analysis)، فإن المشكلة الأساسية هي أن هناك عدداً أقل من السيارات المزودة بنظام تحديد المواقع العالمي. يساعد التعلم الآلي في مثل هذه السيناريوهات على تقدير المناطق التي يمكن العثور فيها على الازدحام على أساس التجارب اليومية [1].
- **التنبؤ بزيادة الطلب على الخدمة:** مثلاً في شبكات النقل عبر الإنترنت، عند حجز سيارة أجرة، يقدر التطبيق سعر الرحلة. ويستخدم في ذلك التعلم الآلي الذي يقوم بتحديد السعر من خلال التنبؤ بزيادة الطلب على السيارة (يرتفع السعر في الاوقات التي يكون الطلب عالي وينخفض مع انخفاض الطلب).

في الأمن والحماية

استخدم تعلم الآلة بكثرة في مجالات الامن والحماية بكثرة. فيما يلي بعض الامثلة لذلك.

كشف الاحتيال عبر الإنترنت (Online Fraud Detection)

يثبت التعلم الآلي قدرته على جعل الفضاء السيبراني مكاناً آمناً وتتبع الاحتيال النقدي على الإنترنت. أحد أمثلة ذلك، هو استخدام Paypal ML للحماية من غسيل الأموال (money laundering). تستخدم الشركة مجموعة من الأدوات التي تساعد على مقارنة ملايين المعاملات التي تتم والتميز بين المعاملات المشروعة أو غير الشرعية التي تتم بين المشتري والبائع [1].

منع الغش

لا يتحمل مقدمو الخدمات المالية أي مسؤولية أكبر من حماية عملائهم ضد أي نشاط احتيالي. يكلف الاحتيال المالي الأميركي بمفرده 50 مليار دولار سنوياً. لم تعد الطرق القديمة للحفاظ على أمان حسابات العملاء جيدة بما فيه الكفاية. مع كل التقدم في أمن البيانات ، هنالك تصعيد من قبل المجرمين في مستوى التحايل. لحماية بيانات العملاء من التهديدات المتطورة بشكل متزايد ، يجب أن تظل المؤسسات والشركات متقدمة على المتسللين. يساعد التعلم الآلي التطبيقات في إحباط المخالفات الأمنية من خلال تفكير المجرمين.

من خلال مقارنة كل معاملة مقابل سجل الحساب ، تكون خوارزميات التعلم الآلي قادرة على تقييم احتمالية أن تكون المعاملة احتيالية. الأنشطة غير العادية، مثل المشتريات من خارج البوالة أو عمليات السحب النقدي الكبيرة ، تجعل النظام يقدم خطوات لتأخير المعاملة حتى يتمكن الإنسان من اتخاذ قرار. في العديد من الحالات ، وفقاً لطبيعة المحاولة ، قد يتم رفض محاولة الشراء أو السحب تلقائياً بواسطة النظام.

بخلاف التعامل البشري ، فإن الخوارزمية قادرة على موازنة تفاصيل المعاملة بسرعة مقابل آلاف نقاط البيانات وتحديد ما إذا كان نشاط المحاولة غير معهود المالك الحساب أم لا. وعلى خلاف برامج الأخرى ، تتعلم برامج التعلم الآلي من كل إجراء يتخذه مالك الحساب ، ومن كل قرار يصدره البرنامج. وبمرور الوقت ، تعدل الخوارزميات نفسها استجابة للعادات المتغيرة من جانب مالك الحساب [5].

إدرك قيمة التعلم الآلي ، كل من Amazon ، و Microsoft ، و IBM ، و Google ، حيث إضافة هذه الشركات قدرات التعلم الآلي المتكاملة في واجهات التطوير المبنية على السحابة (cloud-based developer interfaces).

لكي يكون التعلم الآلي فعالاً ، يجب أن يكون قادراً على الوصول بسرعة إلى كميات كبيرة من البيانات واستيعابها.

عندما يصبح المجرمون أكثر تقدماً في استراتيجياتهم ، لن تتمكن سوى أنظمة الكمبيوتر التي يمكنها الوصول إلى البيانات الكبيرة والقدرة على التفكير والتعلم من إيقافها.

إدارة المخاطر

لا ينبغي أن يكون من المفاجئ أن تكون تكنولوجيا التعلم الآلي حلياً قوياً في السعي إلى إدارة أفضل للمخاطر. في حين تنبأ تطبيقات البرمجيات التقليدية بالجدار الائتمانية على أساس المعلومات الثابتة من تطبيقات القروض والتقارير المالية ، فإن تكنولوجيا التعلم الآلي يمكن أن تذهب إلى أبعد من ذلك وتحلل أيضاً الوضع المالي لمقدم الطلب حيث قد يتم تعديله حسب اتجاهات السوق الحالية وحتى الأخبار ذات الصلة.

من خلال تطبيق التحليل التنبئي على كميات هائلة من البيانات في الوقت الفعلي ، يمكن لتكنولوجيا التعلم الآلي اكتشاف المستثمرين الخارجين عن القانون الذين يعملون في انسجام عبر حسابات متعددة - وهو أمر قد يكون مستحيلاً تقريباً بالنسبة لمدير الاستثمار البشري.

الكفاءة هي فائدة أخرى من التعلم الآلي. بافتراض قدر كبير من العبء لرصد الحسابات ، تمكن أنظمة التعلم الآلي مديري الاستثمار من التركيز على مهام أكثر إنتاجية ، مثل خدمة العملاء [5].

مراقبة الفيديوها (Videos Surveillance)

إن مراقبة العديد من كاميرات الفيديو بواسطة شخص واحد تعتبر مهمة صعبة ومملة. يمكن تدريب أجهزة الكمبيوتر للقيام بهذا العمل المنطقي. نظام المراقبة بالفيديو يستخدم فيه تعلم الآلة الذي يجعل من الممكن كشف الجريمة قبل وقوعها. حيث يتم تتبع أي سلوك غير عادي لأي شخص ، مثل الوقوف بلا حراك لفترة طويلة ، أو التعثر (stumbling) ، أو النوم على المقاعد (napping on benches) وما إلى ذلك. ويمكن للنظام إعطاء تنبيه للموجودين ، مما يمكن أن يساعد في النهاية على تجنب الحوادث المؤسفة. وعندما يتم الإبلاغ عن هذه الأنشطة وتصحيحها ، فإنها تساعد على تحسين خدمات المراقبة [1].

تنقية البريد الإلكتروني والبرامج الضارة (Email Spam and Malware Filtering)

هناك عدد من طرق لتصفية الرسائل غير المرغوب فيها التي يستخدمها عملاء البريد الإلكتروني. للتأكد من تحديث عوامل تصفية الرسائل غير المرغوب فيها باستمرار (spam filters) ، يتم تشغيلها بواسطة التعلم الآلي. عندما يتم إجراء تصفية الرسائل غير المرغوب فيها بناء على أحكام وقوانين معينة (rule-based spam filtering) ، فإنه يفشل في تتبع أحدث الحيل التي اعتمدها مرسلو الرسائل غير المرغوب فيها (tricks adopted by spammers). Multi Layer Perceptron ، 4.5 Decision Tree Induction ، تعتبر بعض تقنيات تصفية البريد المزعج التي يتم تشغيلها بواسطة تعلم الآلة [1].

يتم اكتشاف أكثر من 325000 من البرامج الضارة (malwares) كل يوم ، وكل جزء من الكود هو 98-90 ٪ مشابه لإصداراتها السابقة. تدرك برامج أمان النظام (The system security programs) التي يتم تشغيلها بواسطة التعلم الآلي نمط التشفير (coding pattern). لذلك ، فإنها تكشف عن برامج ضارة (malware) جديدة مع تباين بنسبة 2-10 ٪ بسهولة وتوفر الحماية ضدها [1].

في التجارة وخدمة العملاء

استخدم تعلم الآلة في المجال التجاري وتقديم خدمات العملاء ، من الأمثلة ما يلي:

توصيات المنتج (Product Recommendations)

إذا اشترت أحد المنتجات عبر الإنترنت ، ستأتيك رسائل البريد الإلكتروني بقتراحات التسوق. أو ربما تلاحظ أن موقع الويب للتسوق أو التطبيق يوصيك ببعض العناصر التي تتطابق بطريقة ما مع ذوقك. بالتأكيد ، هذا يحسن تجربة التسوق ، حيث يتم هذا العمل من خلال التعلم الآلي الذي يبني ما يقدم لك على أساس سلوكك مع موقع الويب أو التطبيق ، أو من خلال عمليات الشراء السابقة ، أو العناصر المحببة أو التي تمت إضافتها إلى سلة التسوق ، أو تفضيلات العلامة التجارية وما إلى ذلك ، يتم تقديم توصيات المنتج.

توقعات الاستثمار

كانت خدمات التداول بمساعدة الكمبيوتر موجودة منذ بعض الوقت. وهي تسمح للمستثمرين بأن يكون لديهم طلبية عندما يصل السهم إلى سعر محدد مسبقًا ، ويبيع عند انخفاض السعر إلى ما دون حد معين. من خلال أتمتة الوظائف ، تجعل هذه المنصات التداول أسهل بالنسبة للمستثمر الكبير والصغير ، على حد سواء. في حين يمكنهم حتى تقديم توصيات تستند إلى التحليل الآلي لاتجاهات السوق ، إلا أن لديهم قيودًا.

في السنوات الأخيرة ، ابتعدت صناديق التحوط بشكل متزايد عن طرق التحليل التنبؤية التقليدية واعتمدت خوارزميات تعلم الآلة للتنبؤ باتجاهات التمويل. باستخدام التعلم الآلي ، يأمل مديرو الصناديق في تحديد التغيرات في السوق في وقت أبكر مما هو ممكن مع نماذج الاستثمار التقليدية.

تؤخذ المؤسسات الكبرى على محمل الجد في إمكانات تكنولوجيا التعلم الآلي لتعطيل الصناعة المصرفية الاستثمارية. يقوم كل من JPMorgan و Bank of America و Morgan Stanley بتطوير مستشارين استثماريين مؤتمنين ، مدعومين بتكنولوجيا التعلم الآلي. الشركات الحكيمة في مجال التكنولوجيا سوف تحذوها [5].

دعم العملاء عبر الإنترنت (Online Customer Support)

يوفر عدد من مواقع الويب حاليًا خيارًا للدردشة مع ممثل دعم العملاء أثناء التنقل داخل الموقع. ومع ذلك ، ليس لكل موقع ويب مسؤول تنفيذي مباشر للإجابة على استفساراتك. في معظم الحالات ، يمكنك التحدث إلى chatbot. هذه البرامج تميل إلى استخراج المعلومات من الموقع وتقديمها إلى العملاء. في هذه الأثناء ، تتقدم وسائل الدردشة (chatbots) مع الوقت. فهم يميلون إلى فهم استعلامات المستخدم بشكل أفضل وتقديم إجابات أفضل ، وهذا ممكن بسبب خوارزميات التعلم الآلي [1].

خدمة الزبائن

لا تزال خدمة العملاء الضعيفة واحدة من الشكاوى الرئيسية بين المستهلكين، تركزت الشكاوى على خدمة العملاء البطيئة ، ولكن مع الاستخدام العالمي لدعم الهاتف الآلي ، يشعر العملاء بالإحباط لعدم تمكنهم من التحدث إلى إنسان. بالنسبة لشركة الخدمات المالية المبتكرة الراغبة في الاستثمار في تكنولوجيا التعلم الآلي ، فإن هذا لا يمثل مشكلة بقدر ما يمثل فرصة.

وتشمل مزايا نظم الدعم الآلي توجيه العملاء إلى قسم الصحيح، ومنحهم خيار حل المشاكل البسيطة باستخدام واجهة الآلي، والحفاظ على العملاء من الحاجة إلى الانتظار لشخص ما للرد على الهاتف - وكل ذلك دون التفاعل البشري. وتستفيد الشركة من عدم دفع رواتب الموظفين الذين سيتولون هذه المهام ، ويستفيد العملاء (المقترضين) من خلال التعامل مع مشكلتهم مع سرعة أجهزة الكمبيوتر الحديثة. من الناحية النظرية ، على أي حال.

في الواقع ، لا يستفيد العديد من العملاء من خدمة العملاء التلقائية عندما تكون مشكلتهم غير شائعة ولا يتم تمثيلها بخيار رقمي. وبالرغم من كونه تحسينًا ، إلا أن التعرف على الصوت غالبًا ما يفشل في استيعاب الطلب غير المعتاد ، مع "أنا آسف لأنني لا أفهمك ؛ يرجى الانتظار للحصول على المساعدة" كونها استجابة الأسهم للتعامل مع مشكلة فريدة من نوعها.

الحل ، كما هو منصوص عليه من قبل تكنولوجيا التعلم الآلي ، لا يحل محل أنظمة دعم العملاء الآلية ، ولكن لجعلها أفضل. القوة الهائلة لتكنولوجيا التعلم الآلي للوصول إلى البيانات، التعرف على الأنماط، وتفسير السلوك يعني أن التكنولوجيا يمكن استخدامها لإنشاء أنظمة دعم العملاء الآلية التي تحاكي وكيل

البشري، مع القدرة على فهم والاستجابة للمخاوف شائعة. من خلال جعل بوابات دعم العملاء عبر الإنترنت والهاتف أكثر تشبهاً للإنسان ، يمكن للمؤسسات المالية تقديم دعم فعال يقلل من ضرر العملاء.

بالإضافة إلى ذلك ، يمكن أن تساعد تقنية التعلم الآلي العملاء في تحديد أفضل للمنتجات من خلال تقييم أنشطة الحساب السابقة مقابل البيانات الحالية التي يقدمها العميل ومن أي مكان آخر. يمكن تقديم توصيات المنتج أو الخدمة مباشرة إلى العميل ، أو من خلال المستشار المالي. والنتيجة هي عميل مستنير أفضل لا يضيع وقته بسبب العروض غير اللائقة.

واحدة من أهم العائدات للشركات التي تستثمر في خدمة العملاء في التعلم الآلي هي القدرة على فهم احتياجات كل عميل على نحو أفضل [5].

خدمات اخرى

المساعد الرقمي

لا توجد شركة ، أو سوق تكنولوجيا أو أي سوق ، يعمل بدون إدارة سليمة. ولكي تكون الإدارة "سليمة" ، يجب أن تعمل بفعالية وكفاءة. يمكن ان تساعد تكنولوجيا التعلم الآلي المدراء التنفيذيين والمديرين على أداء وظائفهم بسهولة أكبر من أي وقت مضى.

يمكن ان يتم ذلك من خلال المساعد الرقمي ، حيث تمتلك كل من Google و Apple و Facebook و Microsoft إصدارًا خاصًا بها يدعم ذلك [5].

بعض الأمثلة الشائعة للمساعد الرقمي في قدرته على الرد على اسئلة مثل:

- "ما هو جدول مواعيدي اليوم؟"

- "ما هي رحلات الطيران من ألمانيا إلى لندن"

و للإجابة يبحث المساعد الشخصي عن المعلومات ، أو يذكر طلبات البحث ذات الصلة ، أو يرسل أمرًا إلى موارد أخرى (مثل تطبيقات الهاتف) لجمع المعلومات.

يمكنك حتى توجيه المساعدين للقيام بمهام معينة مثل:

- "ظبط المنبه على الساعة 6 صباحًا في اليوم التالي"

- "ذكرني بزيارة مكتب الفيزا بعد غد."

يعد التعلم الآلي جزءًا مهمًا من هؤلاء المساعدين الرقميين ، حيث يقومون بجمع المعلومات وتنقيحها على أساس مشاركتك السابقة معهم. في وقت لاحق ، يتم استخدام هذه المجموعة من البيانات لتقديم النتائج المصممة لتفضيلاتك [1].

خدمات وسائل التواصل الاجتماعي

تستخدم منصات وسائل التواصل الاجتماعي التعلم الآلي لإضفاء طابع شخصي على صفحتك (عرض الأخبار التي تهتمك، عرض الإعلانات التي تناسبك، إلخ). فهناك الكثير من الأشياء التي تستخدمها وتحبها في حساباتك على الشبكات التواصل الاجتماعية دون أن تدرك أن هذه الميزات الرائعة جلبتها لك تطبيقات التعلم الآلي. مثل: .

- الأشخاص الذين قد تعرفهم: التعلم الآلي يعمل على مفهوم بسيط: التفاهم مع الخبرات. يلاحظ Facebook باستمرار الأصدقاء الذين تتصل بهم ، أو الملفات الشخصية التي تزورها كثيرًا ، أو اهتماماتك ، أو مكان عملك ، أو مجموعة تشاركها مع شخص ما إلخ. على أساس التعلم المستمر ، يتم اقتراح قائمة بمستخدمي فيسبوك الذين يمكنك أن تصادقهم.
- التعرف على الوجه: يمكنك تحميل صورة لك مع صديق وفيسبوك يتعرف على الفور على هذا الصديق. يتحقق الفيسبوك من الوضعيات والتوقعات في الصورة ، ويلاحظ الميزات الفريدة ، ثم يطابقها مع الأشخاص الموجودين في قائمة أصدقائك. العملية برمتها في الخلفية معقدة وتعتني بعامل الدقة ولكن تبدو لك في الواجحة الامامية بسيطة.
- الدبابيس المتشابهة (Similar Pins): التعلم الآلي هو العنصر الأساسي في الرؤية بالحاسب (وهو تقنية لاستخراج المعلومات المفيدة من الصور ومقاطع الفيديو). يستخدم Pinterest الرؤية بالحاسب لتحديد كائنات (أو دبابيس) في الصور والتوصية بكائنات (أو دبابيس) مشابهة وفقًا لذلك [1].

تنقية نتيجة محرك البحث (Search Engine Result Refining)

تستخدم Google ومحركات البحث الأخرى التعلم الآلي لتحسين نتائج البحث نيابةً عنك. في كل مرة تنفذ فيها بحثًا ، تحافظ الخوارزميات الموجودة في الخلفية على مراقبة طريقة ردك على النتائج. إذا فتحت النتائج الأولى (top results) وبقيت على صفحة الويب لفترة طويلة ، سيفترض محرك البحث أن النتائج التي عرضها عليك تتوافق مع طلبك. وبالمثل ، إذا وصلت إلى الصفحة الثانية أو الثالثة من نتائج البحث ، لكنك لم تفتح أيًا من النتائج ، يقدر محرك البحث أن النتائج المعروضة لم تتطابق مع ما تريد. بهذه الطريقة ، تعمل خوارزميات التعلم الآلي التي تعمل في الخلفية على تحسين نتائج البحث [1].

التعرف على الصور

أحد أكثر الاستخدامات الشائعة لتعلم الآلة هو التعرف على الصور. هناك العديد من الحالات حيث يمكنك تصنيف الكائن كصورة رقمية. بالنسبة للصور الرقمية ، تصف القياسات مخرجات كل بيكسل في الصورة. في حالة وجود صورة بالأبيض والأسود ، تعمل كثافة كل بكسل كقياس واحد. إذا كانت الصورة بالأبيض والأسود تحتوي على $N * N$ pixels ، فإن إجمالي عدد البكسل (القياس) هو N^2 في الصورة الملونة ، كل بكسل يعتبر بمثابة 3 قياسات (مكونات اللون الرئيسي) أي RGB . لذلك $N * N$ صورة ملونة هناك $3 * N^2$ قياسات للتعرف على الحروف يمكننا تقسيم جزء من الكتابة إلى صور أصغر ، كل منها يحتوي على حرف واحد. (قد تتكون الفئات من 26 حرفًا من الأبجدية الإنجليزية ، والأرقام 10 ، وبعض الأحرف الخاص) [2].

التشخيص الطبي (Medical Diagnosis)

يوفر تعلم الآلة الطرق والتقنيات والأدوات التي يمكن أن تساعد في حل المشاكل التشخيصية والتكهن (diagnostic and prognostic) في مجموعة متنوعة من المجالات الطبية. يتم استخدام تعلم الآلة في بعض التحليلات مثل التنبؤ بتطور المرض لاستخراج المعرفة الطبية للبحوث ، وللتخطيط للعلاج والدعم ، ولإدارة العامة للمرضى. كما يستخدم تعلم الآلة لتحليل البيانات تفسيرها [2].

استخراج المعلومات (Information Extraction)

استخراج المعلومات هو تطبيق آخر للتعلم الآلي. حيث يمكن استخراج المعلومات المنظمة من البيانات غير المهيكلة. على سبيل المثال ، صفحات الويب والمقالات والمدونات وتقارير الأعمال ورسائل البريد الإلكتروني. تحتفظ قاعدة البيانات العلائقية بالإنتاج الناتج عن استخراج المعلومات. تأخذ عملية الاستخراج المدخلات كمجموعة من الوثائق وتنتج بيانات منظمة. هذا الناتج في شكل ملخص مثل ورقة Excel و جدول في قاعدة بيانات علائقية [2].

المراجع

1. Daffodil. 9 Applications of Machine Learning from Day-to-Day Life. 2017; Available from: <https://medium.com/app-affairs/9-applications-of-machine-learning-from-day-to-day-life-112a47a429d0>
2. Sharma, S. Top 9 Machine Learning Applications in Real World. 2017; Available from: <https://www.datasciencecentral.com/profiles/blogs/top-9-machine-learning-applications-in-real-world>
3. Sabatini, R. How to read the genome and build a human being (video). 2016; Available from: <https://www.youtube.com/watch?v=s6rJLXq1Re0>
4. Eihab Osman, A.O. التنبؤات الانتخابية باستخدام تنقيب البيانات. 2016.
5. Eisenberg, A. 7 Ways Fintechs Use Machine Learning to Outsmart the Competition. Available from: <https://igniteoutsourcing.com/publications/machine-learning-in-finance>