Lecture 8

Supervised Learning

Algorithms for Supervised Learning

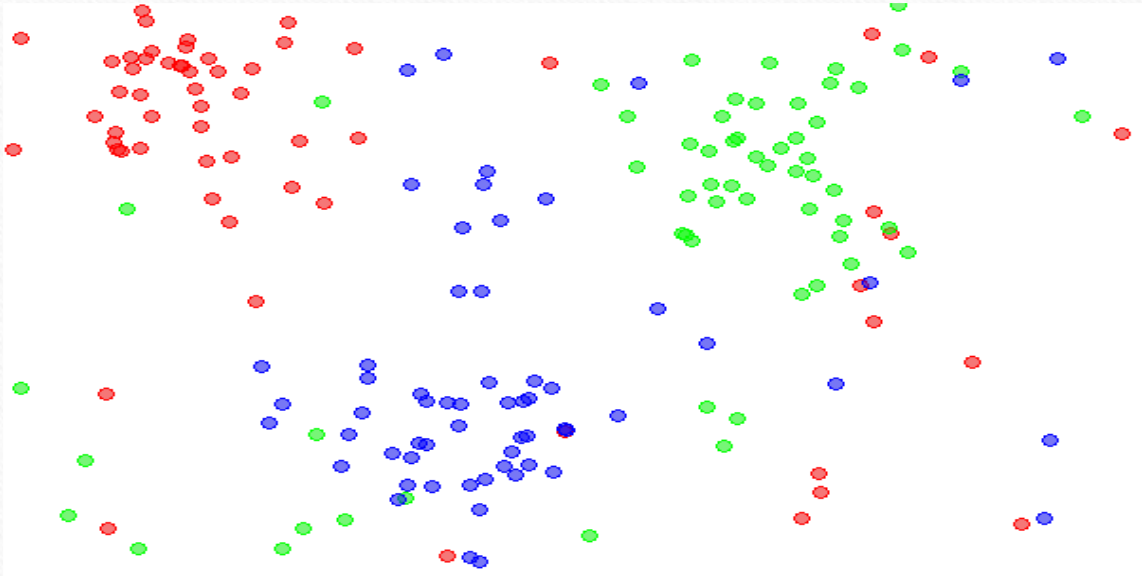
هناكالعديد من الخوارزميات المتاحة للتعلم تحت الإشراف. بعض الخوارزميات
المستخدمةعلى نطاق واسع للتعلم تحت الإشراف موضحة أدناه:

- There are several algorithms available for supervised learning. Some of the widely used algorithms of supervised learning are as shown below:
- k-Nearest Neighbours
- Decision Trees
- Naive Bayes
- Logistic Regression
- Support Vector Machines

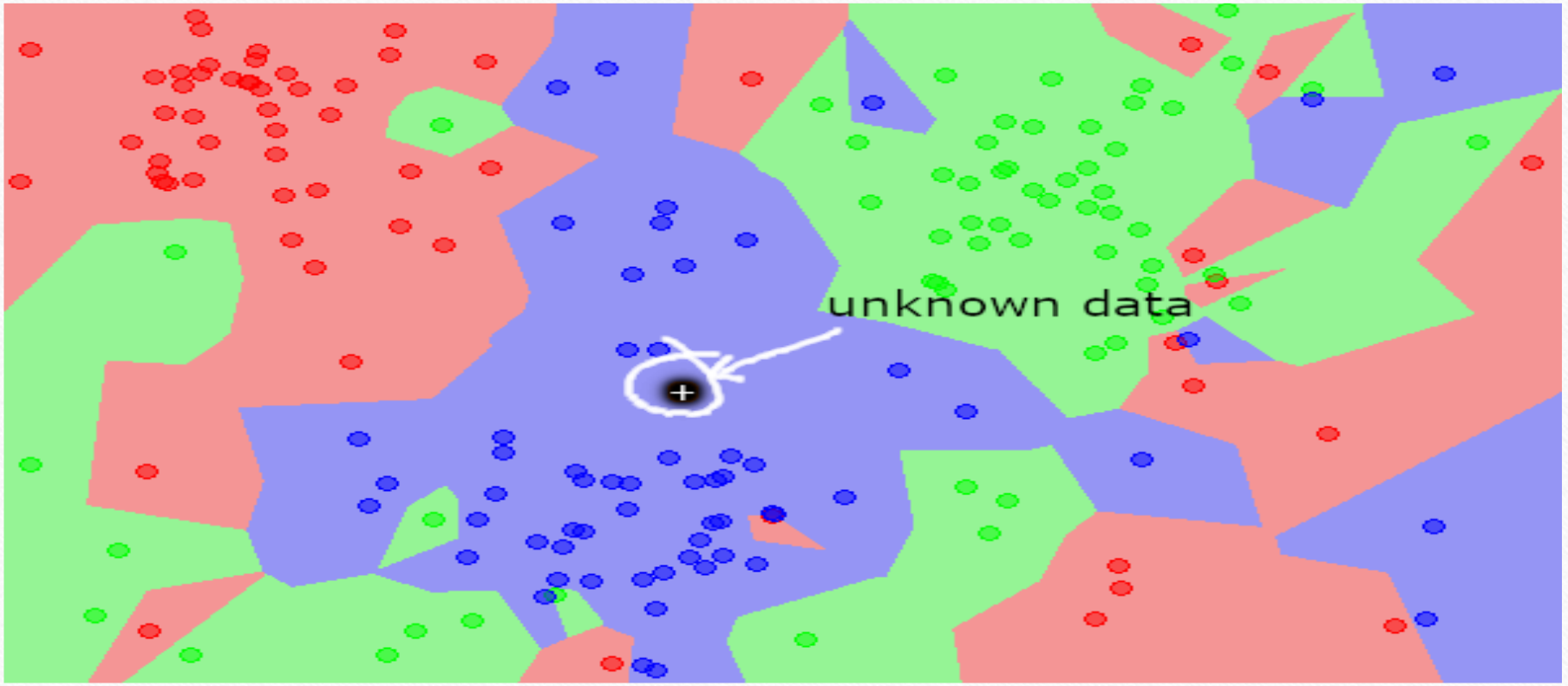
تعدk-Nearest Neighbours ، والتي تسمى ببساطة kNN ، تقنية إحصائية يمكن استخدامهالحل مشاكل التصنيف والانحدار. دعونا نناقش حالة تصنيف كائن غير معروف باستخدامkNN. ضع في اعتبارك توزيع الكائنات كما هو موضح في الصورة أدناه

k-Nearest Neighbours

The k-Nearest Neighbours, which is simply called kNN is a statistical technique that can be used for solving for classification and regression problems. Let us discuss the case of classifying an unknown object using kNN. Consider the distribution of objects as shown in the image given below

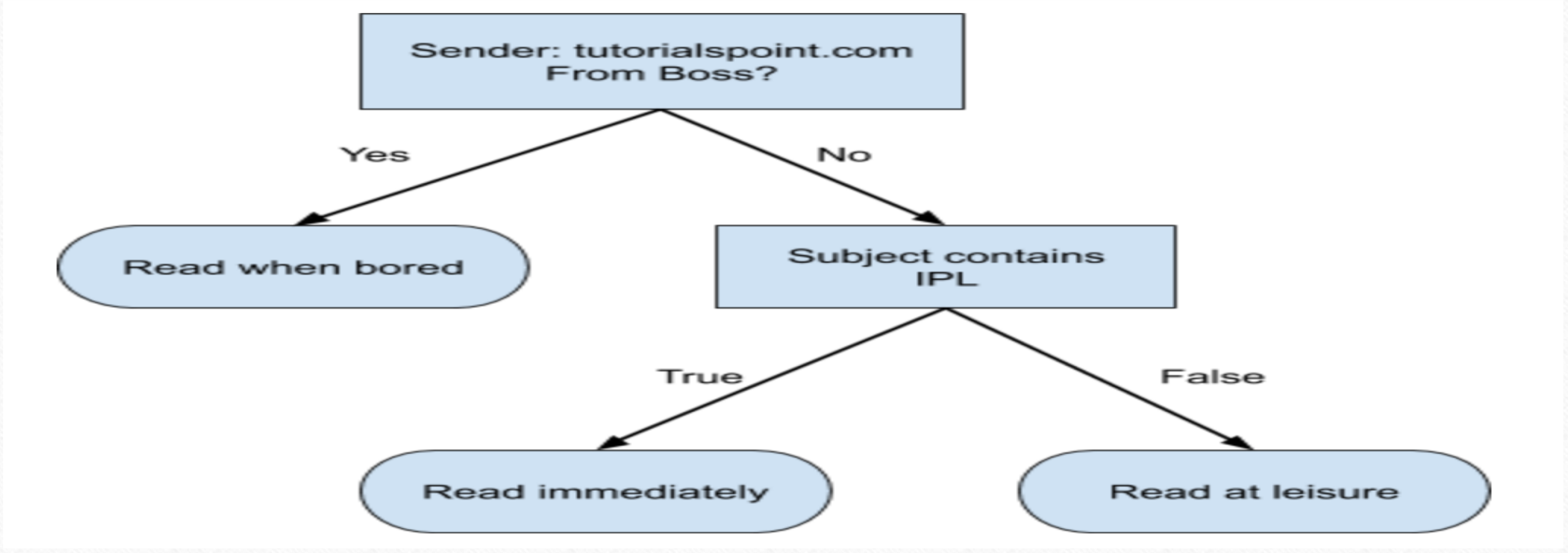


Now, consider a new unknown object that you want to classify as red, green or blue. This is depicted in the figure below.



Decision Trees

A simple decision tree in a flowchart format is shown below:



In this scenario, you are trying to classify an incoming email to decide when to read it.

مثال عليها زي شنو تصنيف بريد إلكتروني وارد لتحديد وقت قراءته

Naive Bayes

يستخدم Naive Bayes لإنشاء المصنفات.

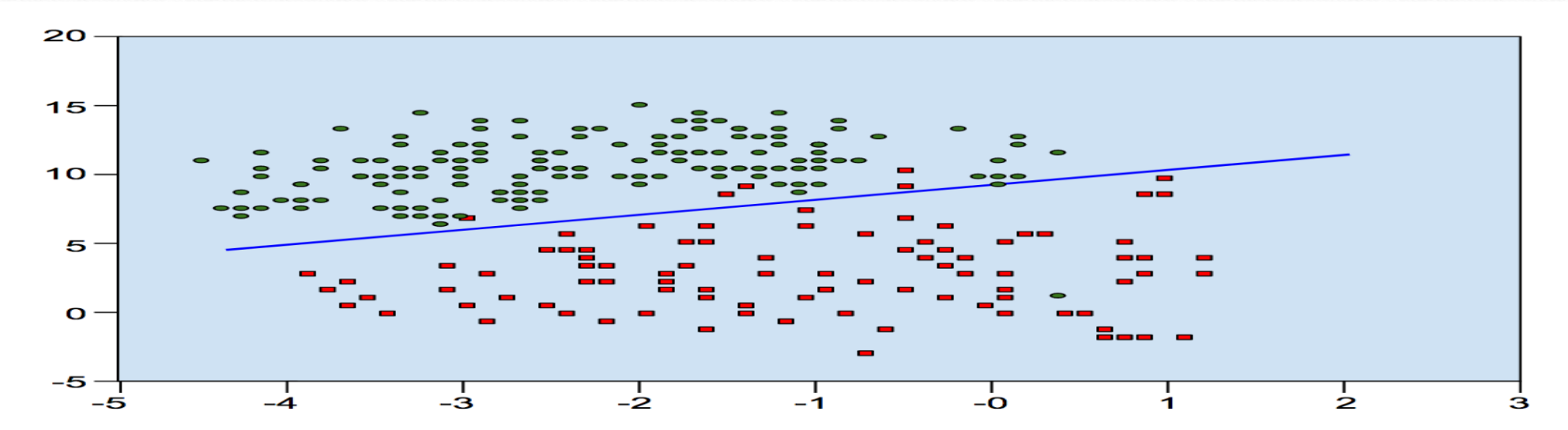
Naive Bayes is used for creating classifiers. Suppose you want to sort out (classify) fruits of different kinds from a fruit basket. You may use features such as color, size and shape of a fruit, For example, any fruit that is red in color, is round in shape and is about 10 cm in diameter may be considered as Apple. So to train the model, you would use these features and test the probability that a given feature matches the desired constraints. The probabilities of different features are then combined to arrive at a probability that a given fruit is an Apple. Naive Bayes generally requires a small number of training data for classification.

، لنفترض أنك تريد فرز (تصنيف) أنواع مختلفة من الفاكهةمن سلة فواكه. يمكنك استخدام ميزات مثل لون وحجم وشكل الفاكهة ، على سبيل المثال ،أي فاكهة حمراء اللون ، تكون مستديرة الشكل ويبلغ قطرها حوالي 10 سم يمكن اعتبارهاتفاحة. لذلك لتدريب النموذج ، يمكنك استخدام هذه الميزات واختبار احتمال تطابق ميزةمعينة مع القيود المطلوبة. يتم بعد ذلك دمج احتمالات الميزات المختلفة للوصول إلى احتمالأن تكون فاكهة معينة تفاحة. يتطلب Naive Bayes عموماً عدداً صغيراً من بيانات التدريبمن أجل التصنيف.

Logistic Regression

انظرإلى الرسم البياني التالي. يوضح توزيع نقاط البيانات في المستوى XY.

Look at the following diagram. It shows the distribution of data points in XY plane.



From the diagram, we can visually inspect the separation of red dots from green dots. You may draw a boundary line to separate out these dots. Now, to classify a new data point, you will just need to determine on which side of the line the point lies.

منالرسم التخطيطي ، يمكننا أن نفحص بصرياً فصل النقاط الحمراء عن النقاط الخضراء. يمكنك رسمخط حدودي للفصل بين هذه النقاط. الآن ، لتصنيف نقطة بيانات جديدة ، ما عليك سوى تحديدأي جانب من الخط تقع النقطة.

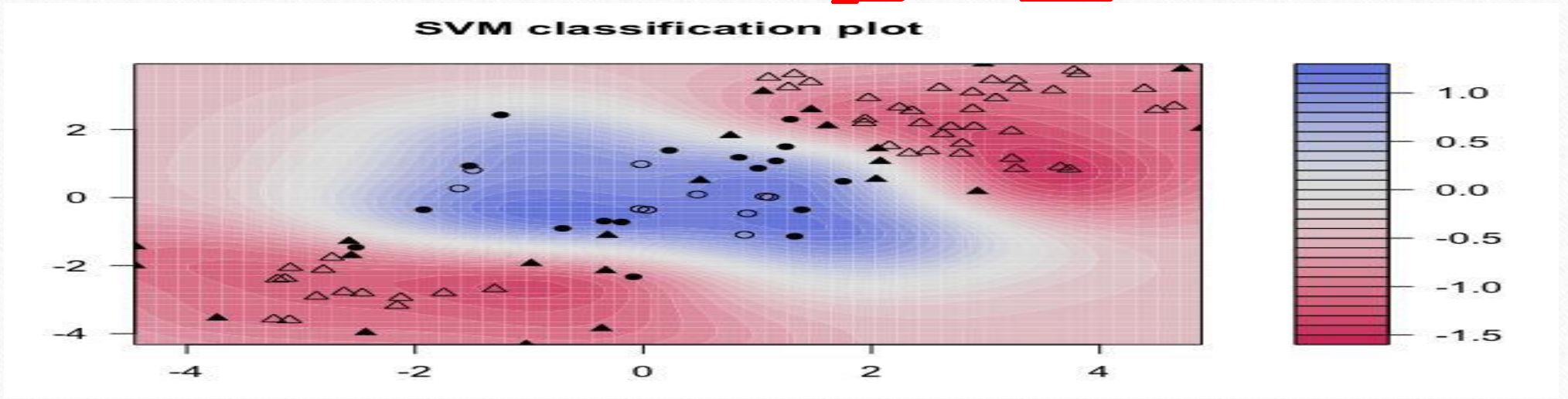
Support Vector Machines

انظرإلى التوزيع التالي للبيانات. هنا لا يمكن فصل فئات البيانات الثلاثة خطياً.

Look at the following distribution of data. Here the three classes of data cannot be linearly separated. The boundary curves are non-linear. In such a case, finding the equation of the curve becomes a complex job.

المنحنيات الحدودية غير خطية في الحالة هذه العثور على معادلة المنحنى مهمة معقدة

The Support Vector Machines (SVM) comes handy in determining the separation boundaries in such situations. مفيدة في تحديد حدود الفصل في مثل هذه المواقف



Unsupervised Learning

Algorithms for Unsupervised Learning

تجمعهو نوع من التعلم غير الخاضع للإشراف والذي يشكل تلقائياً مجموعات من الأشياءالمتشابهة. إنه مثل التصنيف التلقائي. يمكنك تجميع أي شيء تقريباً ، وكلماكانت العناصر أكثر تشابهاً في المجموعة ، كانت المجموعات أفضل.

Clustering is a type of unsupervised learning that automatically forms clusters of similar things. It is like automatic classification. You can cluster almost anything, and the more similar the items are in the cluster, the better the clusters are.

الفرقالرئيسي من**تصنيف**هل تعرف ما تبحث عنه في التصنيف. في حين أن هذا ليسهو الحال في التجميع.

The key difference from **classification** is that in classification you know what you are looking for. While that is not the case in clustering.

Clustering is sometimes called unsupervised classification because it produces the same result as classification does but without having predefined classes.

يطلقعلى التجميع أحياناً التصنيف غير الخاضع للإشراف لأنه ينتج نفس النتيجة التييُنتج عنها التصنيف ولكن دون وجود فئات محددة مسبقاً.

أنواع مختلفة من التجميع يمكن استخدامها:

There are different types of clustering you can utilize:

Exclusive (partitioning)

In this clustering method, Data are grouped in such a way that one data can belong to one cluster only.

حصري(تجزئة)

Example: K-means

في طريقة التجميع هذه ، يتم تجميع البيانات بطريقة يمكن لبيانات واحدة أن تنتمي إلى مجموعة واحدة فقط.

Agglomerative **تكتلي**

In this clustering technique, every data is a cluster. The iterative unions between the two nearest clusters reduce the number of clusters.

Example: Hierarchical clustering

في تقنية التجميع هذه ، كل البيانات عبارة عن كتلة. تقلل الاتحادات التكرارية بين أقرب مجموعتين من عدد المجموعات.

تداخل

fuzzy sets

في هذه التقنية ، يتم استخدام مجموعات ضبابية لتجميع البيانات. قد تنتمي كل نقطة إلى مجموعتين أو أكثر مع درجات منفصلة من العضوية.

هنا ،سيتم ربط البيانات بقيمة عضوية مناسبة. مثال: Fuzzy C-Means

---

Overlapping

In this technique, fuzzy sets is used to cluster data. Each point may belong to two or more clusters with separate degrees of membership.

Here, data will be associated with an appropriate membership value. Example: Fuzzy C-Means

clusters

احتمالية

Probabilistic تستخدم هذه التقنية توزيع الاحتمالات لإنشاء الكتل أو المجموعات

This technique uses probability distribution to create the clusters

Deep learning

التعلمالعميق هو نوع من التعلم الآلي والذكاء الاصطناعي (AI) الذي يقلد الطريقة التي يكتسببها البشر أنواعاً معينة من المعرفة. يعد التعلم العميق عنصراً مهماً في علم البيانات ،والذي يتضمن الإحصاء والنمذجة التنبؤية.

Deep learning

Deep learning is a type of machine learning and artificial intelligence (AI) that imitates the way humans gain certain types of knowledge. Deep learning is an important element of data science, which includes statistics and predictive modeling.

كـيف يعمل

How deep learning works

Computer programs that use deep learning go through much the same process as the toddler learning to identify the dog. Each algorithm in the hierarchy applies a nonlinear transformation to its input and uses what it learns to create a statistical model as output. Iterations continue until the output has reached an acceptable level of accuracy. The number of processing layers through which data must pass is what inspired the label ***deep***.

تمربرامج الكمبيوتر التي تستخدم التعلم العميق بنفس العملية التي يتعلم بها الطفل الصغير للتعرفعلى الكلب. تطبق كل خوارزمية في التسلسل الهرمي تحويلاً غير خطي على مدخلاتها وتستخدمما تتعلمه لإنشاء نموذج إحصائي كمخرج. تستمر التكرارات حتى يصل الإخراج إلى مستوىمقبول من الدقة. إن عدد طبقات المعالجة التي يجب أن تمر البيانات من خلالها هو ماألهم التسميةعميق. deep

Types of Algorithms used in Deep Learning

Here is the list of top 10 most popular deep learning algorithms:

1. Convolutional Neural Networks (CNNs)
2. Long Short Term Memory Networks (LSTMs)
3. Recurrent Neural Networks (RNNs)
4. Generative Adversarial Networks (GANs)
5. Radial Basis Function Networks (RBFNs)
6. Multilayer Perceptrons (MLPs)
7. Self Organizing Maps (SOMs)
8. Deep Belief Networks (DBNs)
9. Restricted Boltzmann Machines(RBMs)
10. Autoencoders

الشبكات العصبية الالتفافية

1. Convolutional Neural Networks (CNNs)

تتكون من طبقات متعددة وتستخدم بشكل أساسي لمعالجة الصور واكتشاف الكائنات

CNN's, also known as ConvNets, consist of multiple layers and are mainly used for image processing and object detection. Yann LeCun developed the first CNN in 1988 when it was called LeNet. It was used for recognizing characters like ZIP codes and digits.

أول شبكة CNN تم استخدامه للتعرف على الأحرف مثل الرموز البريدية والأرقام

CNN's are widely used to identify satellite images, process medical images, forecast time series, and detect anomalies.

تستُخدمشبكات CNN على نطاق واسع لتحديد صور الأقمار الصناعية ومعالجة الصور الطبية وتوقعالسلاسل الزمنية واكتشاف الحالات الشاذة.

شبكات الذاكرة طويلة المدى

2. Long Short Term Memory Networks (LSTMs)

يمكنها تعلم وحفظ التبعيات طويلة المدى , استدعاء المعلومات السابقة لفترات طويلة هو السلوك الافتراضي

LSTMs are a type of Recurrent Neural Network (RNN) that can learn and memorize long-term dependencies. Recalling past information for long periods is the default behavior.

LSTMs retain information over time. They are useful in time-series prediction because they remember previous inputs. LSTMs have a chain-like structure where four interacting layers communicate in a unique way. Besides time-series predictions, LSTMs are typically used for speech recognition, music composition, and pharmaceutical development.

تحتفظLSTMs بالمعلومات بمرور الوقت. إنها مفيدة في تنبؤ السلاسل الزمنية لأنها تتذكرالمدخلات السابقة. تمتلك LSTMs بنية تشبه السلسلة حيث تتواصل أربع طبقاتمتفاعلة بطريقة فريدة. إلى جانب تنبؤات السلاسل الزمنية ، تستُخدم LSTM عادةً للتعرف على الكلام وتكوين الموسيقى والتطوير الصيدلاني.

3. Recurrent Neural Networks (RNNs)

تحتويRNN على اتصالات تشكل دورات موجهة ، والتي تسمح بتغذية المخرجاتمن LSTM كمدخلات إلى المرحلة الحالية.

RNNs have connections that form directed cycles, which allow the outputs from the LSTM to be fed as inputs to the current phase.

The output from the LSTM becomes an input to the current phase and can memorize previous inputs due to its internal memory. RNNs are commonly used for image captioning, time-series analysis, natural-language processing, handwriting recognition, and machine translation.

يصبحالإخراج من LSTM مدخلاً إلى المرحلة الحالية ويمكنه حفظ المدخلات السابقةبسبب ذاكرته الداخلية. تستُخدم RNNs بشكل شائع لتسميات الصور ، وتحليلالسلاسل الزمنية ، ومعالجة اللغة الطبيعية ، والتعرف على خط اليد ، والترجمةالآلية.

Reinforcement learning

What is reinforcement learning?

Reinforcement learning is a machine learning training method based on rewarding desired behaviors and/or punishing undesired ones. In general, a reinforcement learning agent is able to perceive and interpret its environment, take actions and learn through trial and error.

التعلمالمعزز هو طريقة تدريب على التعلم الآلي تعتمد على مكافأة السلوكيات المرغوبةو / أو معاقبة السلوكيات غير المرغوب فيها. بشكل عام ، يكون عامل التعلمالمعزز قادراً على إدراك وتفسير بيئته ، واتخاذ الإجراءات والتعلم من خلال التجربةوالخطأ.

في التعلم المعزز ، يبتكر المطورون طريقة لمكافأة السلوكيات المرغوبة ومعاقبة السلوكيات السلبية. تحدد هذه الطريقة قيماً إيجابية للإجراءات المرغوبة لتشجيع العامل والقيم السلبية على السلوكيات غير المرغوب فيها. يبرمج هذا الوكيل على السعي وراء المكافأة الشاملة طويلة الأجل والحد الأقصى لتحقيق الحل الأمثل.

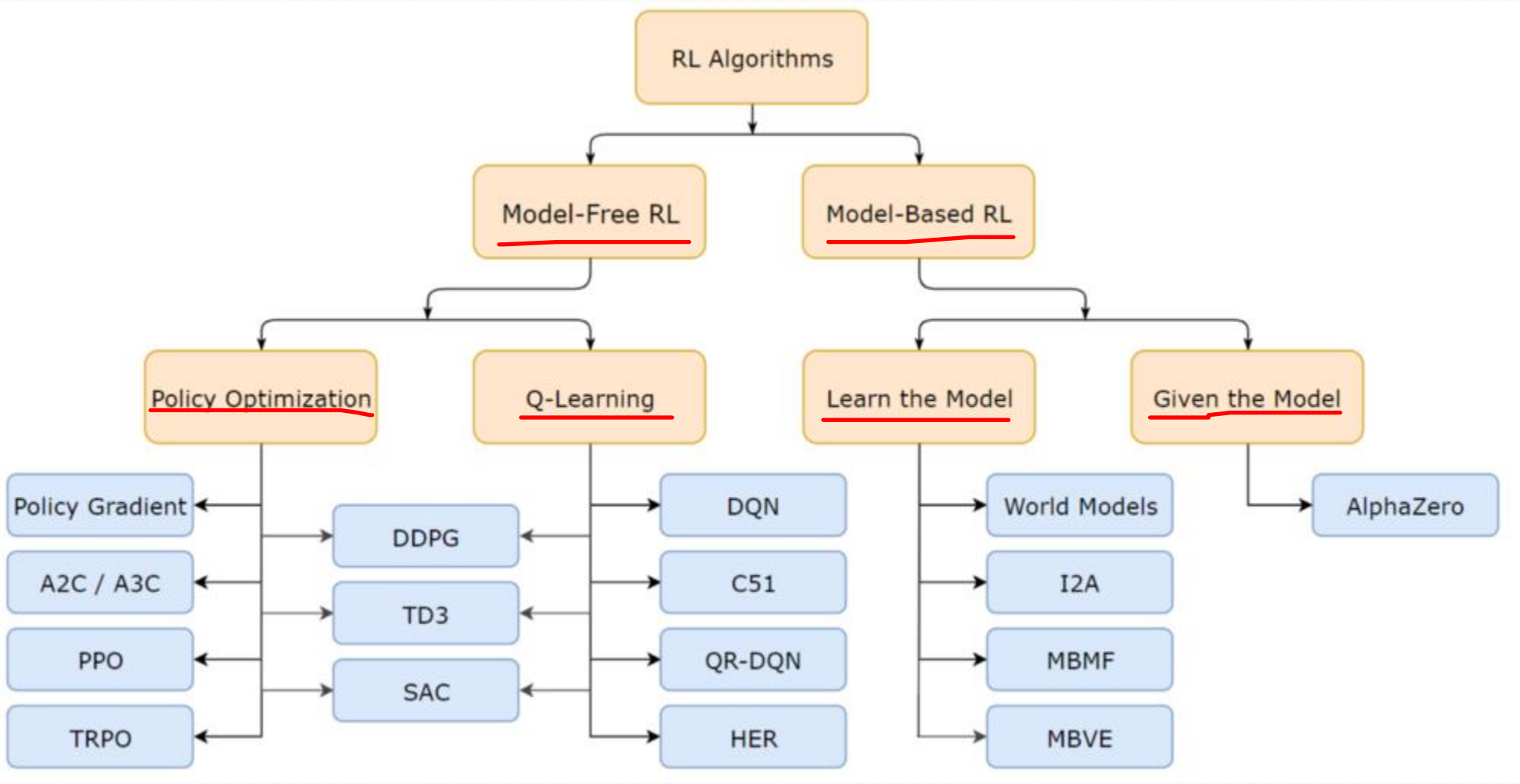
كـيف يعمل؟

How does reinforcement learning work?

In reinforcement learning, developers devise a method of rewarding desired behaviors and punishing negative behaviors. This method assigns positive values to the desired actions to encourage the agent and negative values to undesired behaviors. This programs the agent to seek long-term and maximum overall reward to achieve an optimal solution.

These long-term goals help prevent the agent from stalling on lesser goals. With time, the agent learns to avoid the negative and seek the positive. This learning method has been adopted in artificial intelligence (AI) as a way of directing unsupervised machine learning through rewards and penalties.

تساعدهذه الأهداف طويلة المدى في منع الوكيل من المماطلة في الأهداف الأقل.مع مرور الوقت ، يتعلم الوكيل أن يتجنب السلبيات ويبحث عن الإيجابي. تماعتماد طريقة التعلم هذه في الذكاء الاصطناعي (AI) كطريقة لتوجيه التعلم الآلي غير الخاضع للإشراف من خلال المكافآت والعقوبات.


RL Algorithms
Model-Free RL
Model-Based RL
Policy Optimization
Q-Learning
Learn the Model
Given the Model
Policy Gradient
A2C / A3C
PPO
TRPO
DDPG
TD3
SAC
DQN
C51
QR-DQN
HER
World Models
I2A
MBMF
MBVE
AlphaZero

الحالات والمكافآت والإجراءات

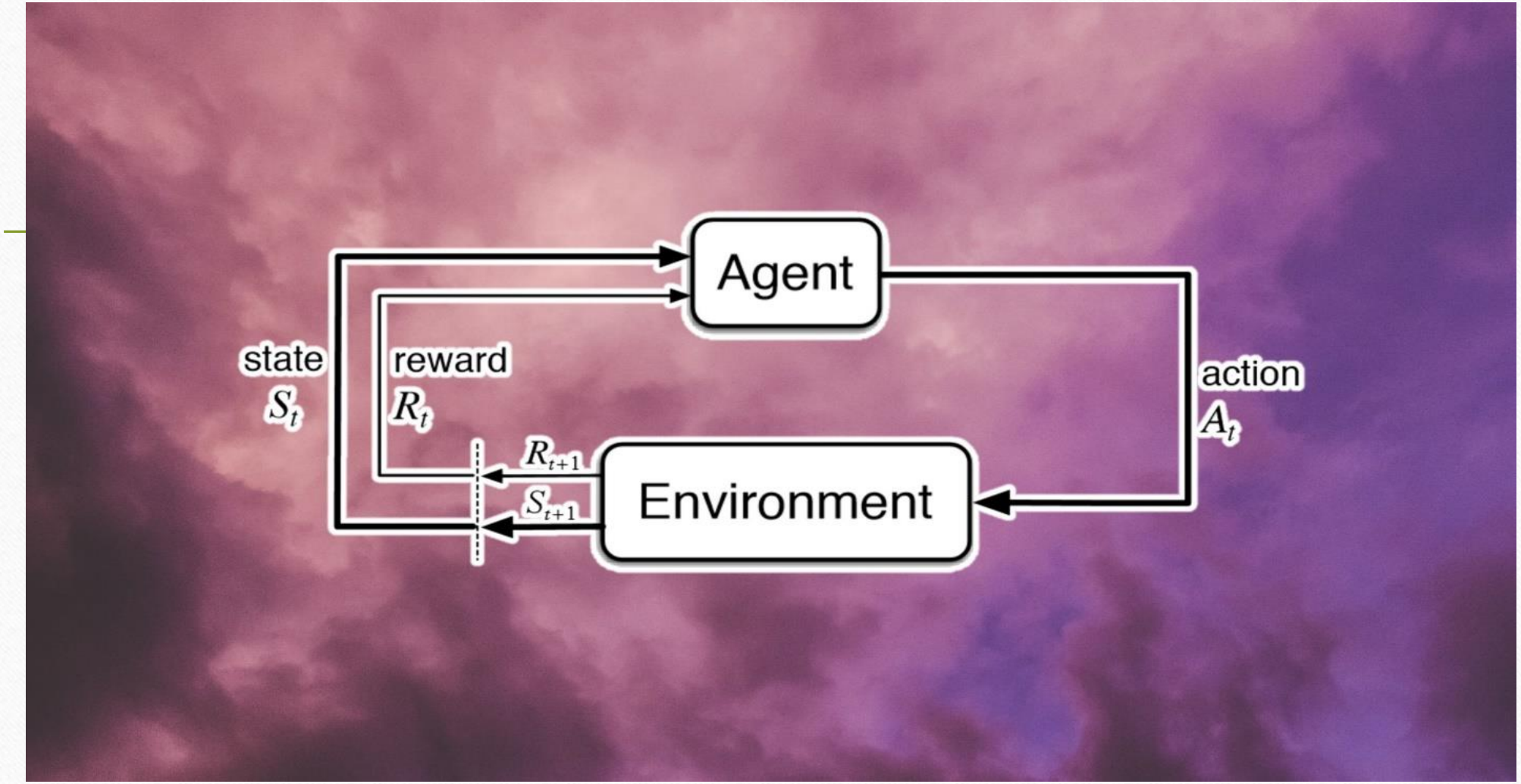
States, rewards, and actions

يعني فيه عميل وبيئة

توفر البيئة معلومات حول الحالة للنظام

At the heart of every reinforcement learning problem are ***an agent*** *and* ***an environment.*** The environment provides information about ***the state*** of the system. The agent observes these states and interacts with the environment by taking **actions.** Actions can be discrete (e.g., flipping a switch) or continuous (e.g., turning a knob). These actions cause the environment to transition to a ***new state***. And based on whether the new state is relevant to the goal of the system, the agent receives a ***reward*** (the reward can also be zero or negative if it moves the agent away from its goal).

النظام.يلاحظ الوكيل هذه الحالات ويتفاعل مع البيئة من خلال أخذ **الأجراءات.**
يمكنأن تكون الإجراءات منفصلة (على سبيل المثال ، قلب مفتاح) أو مستمرة (علىسبيل المثال ، تدوير المقبض). تؤدي هذه الإجراءات إلى انتقال البيئة إلى حالة جديدة
ملف دولة جديدة. واستناداً إلى ما إذا كانت الحالة الجديدة ذات صلة بهدف النظام ،يتلقى الوكيل ***أجائزة***(يمكن أن تكون المكافأة أيضاً صفرية أو سلبية إذا أبعدتالوكيل عن هدفه).


Agent
state
S_t
reward
R_t
action
A_t
R_{t+1}
S_{t+1}
Environment

يستمر في التكرار خلال الدورات حتى يصل إلى الحالة المطلوبة أو إنتهاء صلاحية الحد الأقصى لعدد الخطوات

كل دورة

Every ***cycle*** of state-action-reward is called **a *step***. The reinforcement learning system continues to iterate through cycles until it reaches the desired state or a maximum number of steps are expired. This series of steps is called an ***episode***. At the beginning of each episode, the environment is set to an initial state and the agent's reward is reset to zero.

هذه السلسلة من الخطوات تسمى حلقة

في بداية كل حلقة يتم تعيين البيئة على الحالة الأولية ويتم إعادة تعيين مكافأة الوكيل إلى الصفر

The goal of reinforcement learning is to train the agent to take actions that maximize its rewards. The agent's action-making function is called a ***policy***. An agent usually requires many episodes to learn a good policy. For simpler problems, a few hundred episodes might be enough for the agent to learn a decent policy. For more complex problems, the agent might need millions of episodes of training.

الهدفمن التعلم المعزز هو تدريب الوكيل على اتخاذ الإجراءات التي تزيد من مكافآته. تسمى وظيفةعمل الوكيل [illegible] عادة ما يتطلب الوكيل العديد من الحلقات لتعلم سياسة جيدة. بالنسبةللمشكلات الأبسط ، قد تكون بضع مئات من الحلقات كافية للوكيل لتعلم سياسة لائقة.للمشاكل الأكثر تعقيداً ، قد يحتاج الوكيل إلى ملايين حلقات التدريب.

تسمى السياسة

policy

Reinforcement learning applications

شطرنج

Chess

Here, the environment is the chessboard and the state of the environment is the location of chess pieces on the board. The RL agent can be one of the players (alternatively, both players can be RL agents separately training in the same environment). Each game of chess is an episode. The episode starts at an initial state, with black and white pieces lined on the edges of the board. At each step, the agent observes the board (the state) and moves one of its pieces (takes an action), which transitions the environment to a new state. The agent receives a reward for reaching the checkmate state and zero rewards otherwise. One of the key challenges of chess is that the agent doesn't receive any rewards before it checkmates the opponent, which makes it hard to learn.

هنا ،البيئة هي رقعة الشطرنج وحالة البيئة هي موقع قطع الشطرنج على السبورة. يمكن أنيكون وكيل RL أحد اللاعبين (بدلاً من ذلك ، يمكن أن يكون كلا اللاعبين من وكلاء RL يتدربانبشكل منفصل في نفس البيئة). كل لعبة شطرنج عبارة عن حلقة. تبدأ الحلقة في حالتهاالأولية ، بقطع سوداء وبيضاء مبطنة على حواف السبورة. في كل خطوة ، يلاحظ الوكيل اللوحة(الحالة) وينقل إحدى قطعها (يتخذ إجراء) ، مما يؤدي إلى نقل البيئة إلى حالة جديدة. يتلقىالوكيل مكافأة مقابل وصوله إلى حالة كش ملك ولا مكافآت بخلاف ذلك. أحد التحدياتالرئيسية في لعبة الشطرنج هو أن العميل لا يتلقى أي مكافآت قبل أن يقوم بكش زملائهعلى الخصم ، مما يجعل من الصعب التعلم.

السياراتذاتية القيادة:في القيادة الذاتية ، الوكيل هو السيارة ، والبيئة هي العالم الذيتتنقل فيه السيارة. يراقب وكيل RL حالة البيئة من خلال الكاميرات والليدار وأجهزةالاستشعار الأخرى. يمكن للوكيل اتخاذ إجراءات ملاحية مثل التسريع أو الضغطعلى الفرامل أو الانعطاف يساراً أو يميناً أو عدم القيام بأي شيء. يكافأ وكيل RL للبقاء على الطريق ، وتجنب الاصطدامات ، والامتثال لأنظمة القيادة ، والبقاء في المسارالصحيح.

Self-driving cars: In autonomous driving, the agent is the car, and the environment is the world that the car is navigating. The RL agent observes the state of the environment through cameras, lidars, and other sensors. The agent can take navigation actions such as accelerate, hit the brake, turn left or right, or do nothing. The RL agent is rewarded for staying on the road, avoiding collisions, conforming to driving regulations, and staying on course.

Deep reinforcement learning

يعدالتعلم المعزز العميق أحد أكثر فروع الذكاء الاصطناعي إثارة للاهتمام. إنه وراء بعض الإنجازات الأكثر بروزاً لمجتمع الذكاء الاصطناعي ، بما في ذلك التغلب على أبطال البشر في ألعاب الطاولة وألعاب الفيديو ، والسيارات ذاتية القيادة ، والروبوتات ،وتصميم أجهزة الذكاء الاصطناعي.

يستفيدالتعلم المعزز العميق من قدرة التعلم للشبكات العصبية العميقة لمعالجة المشكلات التي كانت معقدة للغاية بالنسبة لتقنيات RL الكلاسيكية. يعد التعلم المعززالعميق أكثر تعقيداً من الفروع الأخرى للتعلم الآلي.

Deep reinforcement learning is one of the most interesting branches of artificial intelligence. It is behind some of the most remarkable achievements of the AI community, including beating human champions at board and video games, self-driving cars, robotics, and AI hardware design.

Deep reinforcement learning leverages the learning capacity of deep neural networks to tackle problems that were too complex for classic RL techniques. Deep reinforcement learning is much more complicated than the other branches of machine learning.

في التعلم المعزز العميق ، يمكنك تنفيذ جميع الخوارزميات المذكورة أعلاه للتعلم المعززبأي طريقة تريدها. على سبيل المثال ، ينُشئ Qlearning ، وهو نوع كلاسيكيمن خوارزمية التعلم المعزز ، جدولاً لقيم الحالة - الإجراء - المكافأة حيث يتفاعلالوكيل مع البيئة. تعمل هذه الأساليب بشكل جيد عندما تتعامل مع بيئة بسيطةللغاية حيث يكون عدد الحالات والإجراءات صغيراً جداً.

In Deep reinforcement learning you can implement all the above-mentioned algorithms for reinforcement learning in any way you want. For example, Q-learning, a classic type of reinforcement learning algorithm, creates a table of state-action-reward values as the agent interacts with the environment. Such methods work fine when you're dealing with a very simple environment where the number of states and actions are very small.

ولكنعندما تتعامل مع بيئة معقدة ، حيث يمكن أن يصل العدد المشترك من الإجراءاتوالحالات إلى أعداد هائلة ، أو يمكن أن يكون لها حالات غير محدودة فعلياً ،يصبح تقييم كل زوج ممكن من الإجراءات أمراً مستحيلاً.

فيهذه الحالات ستحتاج إلى وظيفة تقريبية يمكنها التعرف على السياسات المثلىبناءً على بيانات محدودة. وهذا ما**الشبكات العصبية الاصطناعية**يفعل.

ANN

But when you're dealing with a complex environment, where the combined number of actions and states can reach huge numbers, or can have virtually limitless states, evaluating every possible state-action pair becomes impossible.

In these cases, you'll need an approximation function that can learn optimal policies based on limited data. And this is what **artificial neural networks** do.

بالنظرإلى البنية الصحيحة ووظيفة التحسين ، يمكن للشبكة العصبية العميقة أن تتعلمالسياسة المثلى دون المرور بجميع الحالات الممكنة للنظام.**وكلاء التعلم التعزيزيالعميق**لا تزال بحاجة إلى كميات هائلة من البيانات (على سبيل المثال ، آلافالساعات من اللعب في Dota و StarCraft) ، لكن يمكنهم معالجة المشكلات التيكان من المستحيل حلها باستخدام أنظمة التعلم المعززة الكلاسيكية.

Given the right architecture and optimization function, a deep neural network can learn an optimal policy without going through all the possible states of a system. **Deep reinforcement learning agents** still need huge amounts of data (e.g., thousands of hours of gameplay in Dota and StarCraft), but they can tackle problems that were impossible to solve with classic reinforcement learning systems.

على سبيل المثال ، يمكن لنموذج RL العميق استخدام الشبكات العصبية التلافيفية لاستخراج معلومات الحالة من البيانات المرئية مثل موجزات الكاميرا ورسومات ألعاب الفيديو. ويمكن للشبكات العصبية المتكررة استخلاص معلومات مفيدة من تسلسل الإطارات ، مثل المكان الذي تتجه فيه الكرة أو إذا كانت السيارة متوقفة أو تتحرك. يمكن أن تساعد هذه القدرة التعليمية المعقدة وكلاء RL على فهم البيئات الأكثر تعقيداً وتعيين حالاتهم إلى الإجراءات.

For example, a deep RL model can use convolutional neural networks to extract state information from visual data such as camera feeds and video game graphics. And recurrent neural networks can extract useful information from sequences of frames, such as where a ball is headed or if a car is parked or moving. This complex learning capacity can help RL agents to understand more complex environments and map their states to actions.

التعلمالمعزز العميق يمكن مقارنته بـ**التعلم الآلي الخاضع للإشراف**. يولد النموذج إجراءات ،وبناءً على التعليقات الواردة من البيئة ، يقوم بتعديل معلماته. ومع ذلك ، فإنالتعلم المعزز العميق له أيضاً بعض التحديات الفريدة التي تجعله مختلفاً عن التعلمالتقليدي الخاضع للإشراف.

علىعكسمشاكل التعلم الخاضعة للإشراف ، حيث يحتوي النموذج على مجموعة منالبيانات المصنفة ، يكون لوكيل RL فقط الوصول إلى نتائج تجاربه الخاصة. قد يكونقادراً على تعلم السياسة المثلى بناءً على الخبرات التي يجمعها عبر حلقات تدريبيةمختلفة. لكنها قد تفقد أيضاً العديد من المسارات المثلى الأخرى التي كان من الممكنأن تؤدى إلى سياسات أفضل.

Deep reinforcement learning is comparable to **supervised machine learning**. The model generates actions, and based on the feedback from the environment, it adjusts its parameters. However, deep reinforcement learning also has a few unique challenges that make it different from traditional supervised learning.

Unlike supervised learning problems, where the model has a set of labeled data, the RL agent only has access to the outcome of its own experiences. It might be able to learn an optimal policy based on the experiences it gathers across different training episodes. But it might also miss many other optimal trajectories that could have led to better policies.

إجراءات الحالة

يحتاج التعلم المعزز أيضاً إلى تقييم مسارات أزواج الإجراءات الحكومية ، والتي يصعب تعلمها أكثر من مشكلات التعلم الخاضعة للإشراف حيث يتم إقران كل مثال تدريبي بنتائج المتوقعة.

يزيد هذا التعقيد الإضافي من متطلبات البيانات لنماذج التعلم المعزز العميق. ولكن على عكس التعلم الخاضع للإشراف ، حيث يمكن تنسيق بيانات التدريب وإعدادها مسبقاً ، تجمع نماذج التعلم المعزز العميق بياناتها أثناء التدريب. في بعض أنواع خوارزميات RL ، يجب التخلص من البيانات التي تم جمعها في حلقة ما بعد ذلك ولا يمكن استخدامها لزيادة تسريع عملية ضبط النموذج في الحلقات المستقبلية.

Reinforcement learning also needs to evaluate trajectories of state-action pairs, which is much harder to learn than supervised learning problems where every training example is paired with its expected outcome.

This added complexity increases the data requirements of deep reinforcement learning models. But unlike supervised learning, where training data can be curated and prepared in advance, deep reinforcement learning models gather their data during training. In some types of RL algorithms, the data gathered in an episode must be discarded afterward and can't be used to further speed up the model tuning process in future episodes.

قراية مشّي حالك):