Techniques for Predictive Modeling

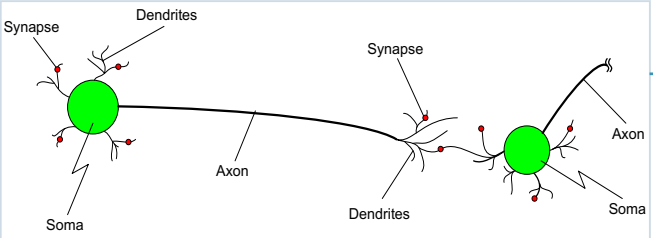
Neural Network Concepts

- **Neural networks (NN):** a brain metaphor for information processing
- Neural computing
- Artificial neural network (ANN)
- Many uses for ANN for
  - pattern recognition, forecasting, prediction, and classification
- Many application areas
  - finance, marketing, manufacturing, operations, information systems, and so on

مفهوم الشبكة العصبية

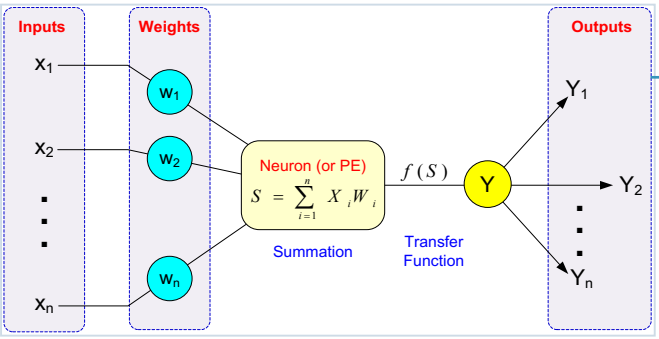
- **الشبكات العصبية (NN) :** استعارة دماغية لمعالجة المعلومات
- الحوسبة العصبية
- الشبكة العصبية الاصطناعية (ANN)
- العديد من الاستخدامات لـ ANN لـ
  - التعرف على الأنماط والتوقع والتنبؤ والتصنيف
- العديد من مجالات التطبيق
  - التمويل والتسويق والتصنيع والعمليات ونظم المعلومات ، وغيرها

Biological Neural Networks → الشبكات العصبية البيولوجية



- Two interconnected brain cells (neurons)
  اثنين من خلايا الدماغ المترابطة (الخلايا العصبية)

Processing Information in ANN → معالجة المعلومات في الشبكة العصبية الإصطناعية



- A single neuron (processing element – PE) with inputs and outputs
  خلية عصبيه وحده (عنصر المعالجة – PE) مع المدخلات و المخرجات

Biology Analogy → التناظر او التشابه علم الاحياء

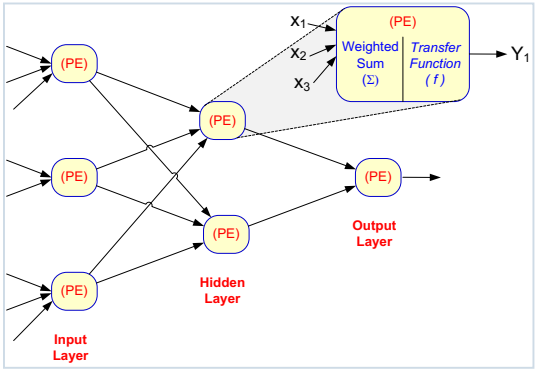
Biological – البيلوجي	Vs.	Artificial – الاصطناعي
Soma		Node
Dendrites - التشعبات		Input
Axon – محور عصبي		Output
Synapse – تشابك عصبي		Weight – له وزن
Slow – بطيء		Fast – سريع
Many neurons (10^9) – العديد من الخلايا العصبية		Few neurons (~100s) – عدد قليل من الخلايا العصبية

Elements of ANN

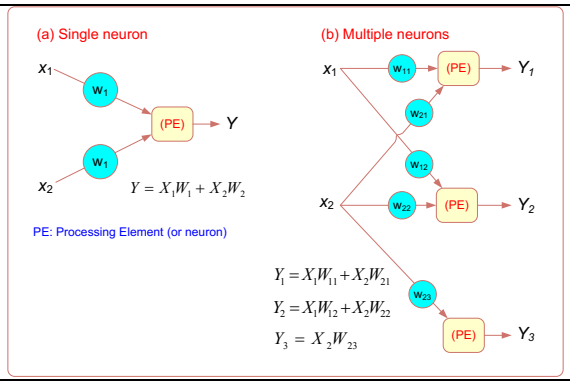
- Processing element (PE)
- Network architecture
  - Hidden layers
  - Parallel processing
- Network information processing
  - Inputs
  - Outputs
  - Connection weights
  - Summation function

عناصر ANN (الشبكة العصبية الاصطناعية)

- عنصر المعالجة (PE)
- بنية الشبكات
  - طبقات مخفية
  - المعالجة المتوازية
- معالجة معلومات الشبكة
  - المدخلات
  - المخرجات
  - أوزان التوصيل
  - دالة التلخيص



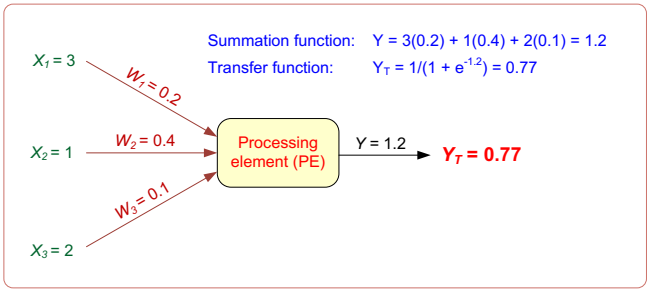
- Neural Network with One Hidden Layer
- الشبكة العصبية مع طبقة واحدة مخفية



- Summation Function for a Single Neuron (a), and Several Neurons (b)
- دالة التلخيص لخلية عصبية واحدة (a) ، و "العديد من الخلايا العصبية (ب)

Elements of ANN (Cont.)

- Transformation (Transfer) Function → دالة التحويل او النقل
  - Linear function → الدالة الخطية
  - Sigmoid (logical activation) function [0 1] → الدالة اللوجستية تكون للتنشيط المنطقي
  - Tangent Hyperbolic function [-1 1] → دالة المماس الزائدي او تسمى قوانين الدوال الزائدية



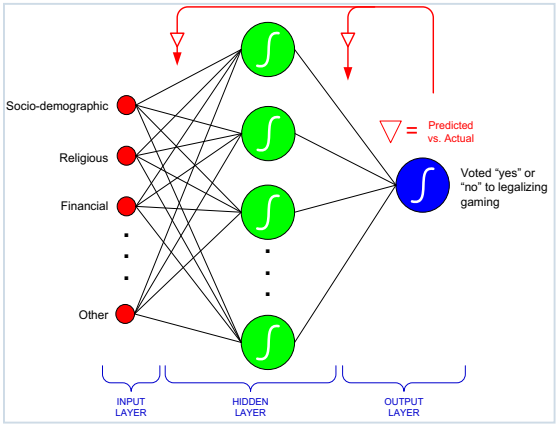
Neural Network Architectures

- Architecture of a neural network is driven by the task it is intended to address
  - Classification, regression, clustering, general optimization, association, ….
- **Most popular architecture:** Feedforward, multi-layered perceptron with backpropagation learning algorithm
  - Used for both classification and regression type problems
- **Others** – Recurrent, self-organizing feature maps, Hopfield networks, …

معمارية الشبكات العصبية

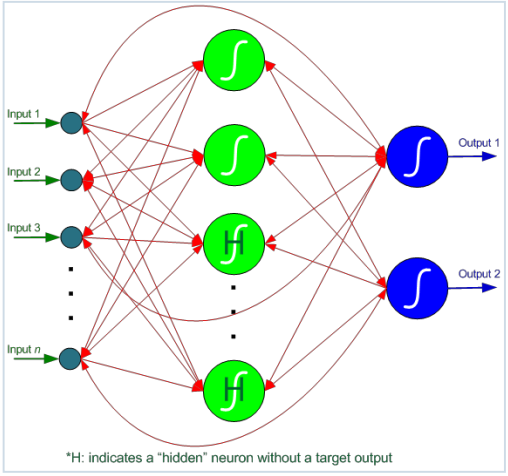
- البنية للشبكة العصبية تقاد بواسطة المهمة المقصودة بها العنوان
  - التصنيف ، التراجع ، التجميع ، التحسين العام ، الارتباط ، ...
- **البنيات الأكثر شعبية :** (هنا يذكر لي البنيات الي يتم استخدامها في طرق تعلم الشبكات العصبونية لتناقل المعلومات)
  - تستخدم لكل من مشاكل التصنيف والتراجع

Neural Network Architectures
(Feed-Forward Neural Networks)

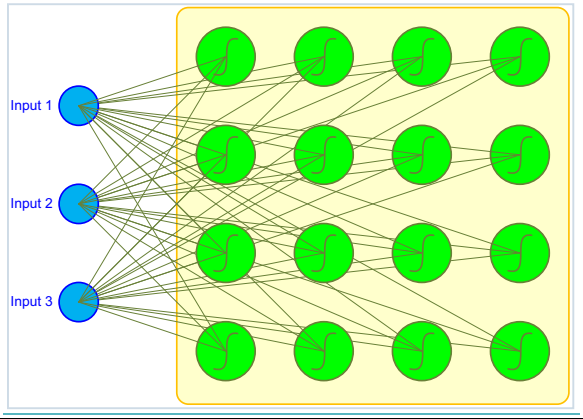


- Feed-forward MLP with 1 Hidden Layer

Neural Network Architectures
(Recurrent Neural Networks)



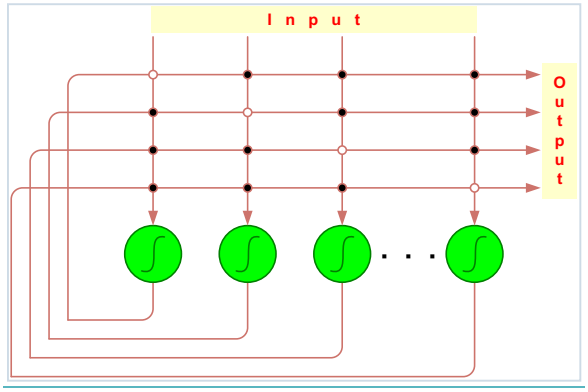
Other Popular ANN Paradigms
Self-Organizing Maps (SOM)



- First introduced by the Finnish Professor Teuvo Kohonen
- Applies to clustering type problems

- قدم لأول مرة من قبل الأستاذ الفنلندي Teuvo Kohonen
- ينطبق على نوع المشاكل العنقودية

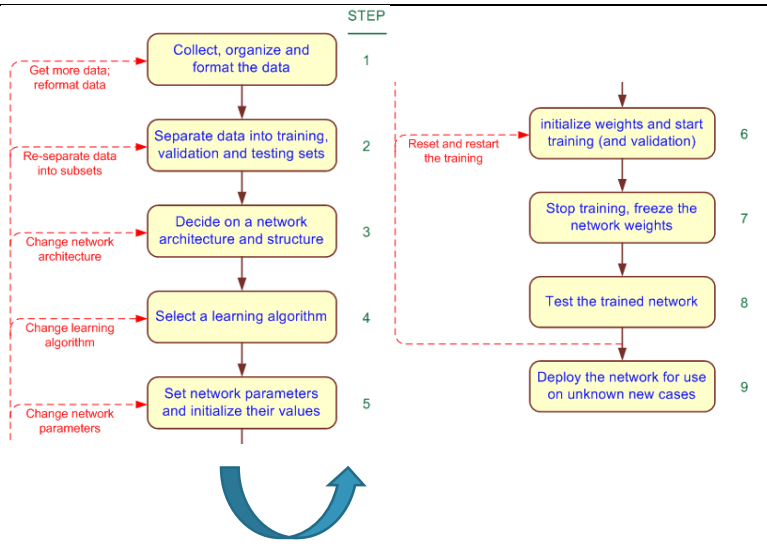
Other Popular ANN Paradigms
Hopfield Networks



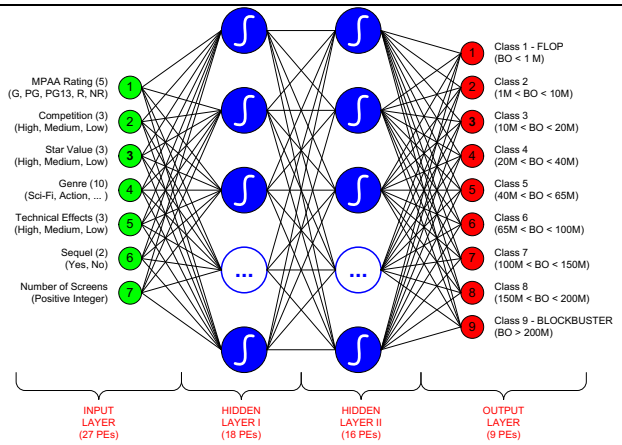
- First introduced by John Hopfield
- Highly interconnected neurons
- Applies to solving complex computational problems (e.g., optimization problems

- قدم لأول مرة من قبل John Hopfield
- الخلايا العصبية مترابطة للغاية
- تنطبق على حل المشكلات الحسابية المعقدة (مثل مشكلات التحسين)

Development Process of an ANN



An MLP ANN Structure for the Box-Office Prediction Problem



Testing a Trained ANN Model

- Data is split into three parts
  - Training (~60%)
  - Validation (~20%)
  - Testing (~20%)
- *k*-fold cross validation
  - Less bias
  - Time consuming

اختبار نموذج ANN المدرّب

- يتم تقسيم البيانات إلى ثلاثة أجزاء
  - التدريب (~ 60٪)
  - التحقق من الصحة (~ 20٪)
  - اختبار (~ 20٪)
- التحقق عبر k-fold
  - أقل تحيز
  - استهلاك الوقت

AN Learning Process A Supervised Learning Process

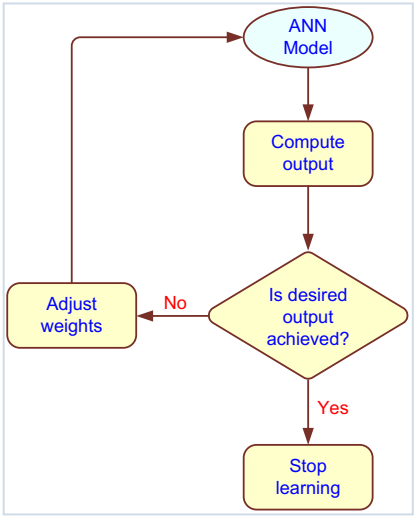
Three-step process:

1. Compute temporary outputs.
2. Compare outputs with desired targets.
3. Adjust the weights and repeat the process.

عملية التعلم AN عملية تتم بالإشراف

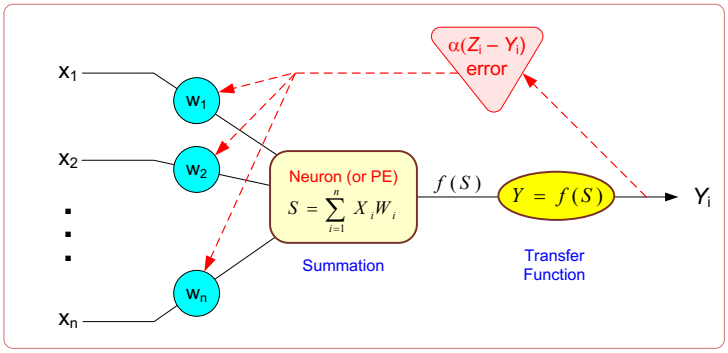
عملية من ثلاث خطوات:

1. حساب النواتج المؤقتة.
2. مقارنة المخرجات مع الأهداف المرجوة (المطلوبه).
3. ضبط الأوزان وتكرار العملية.



Backpropagation Learning

- Backpropagation of Error for a Single Neuron
- من الخطأ ان تكون لخلية عصبيه واحدة



- The learning algorithm procedure
  1. Initialize weights with random values and set other network parameters
  2. Read in the inputs and the desired outputs
  3. Compute the actual output (by working forward through the layers)
  4. Compute the error (difference between the actual and desired output)
  5. Change the weights by working backward through the hidden layers
  6. Repeat steps 2-5 until weights stabilize

- نهج خوارزمية التعلم
  1. قم بتهيئة الأوزان مع القيم العشوائية وقم بتعيين عوامل الشبكة الأخرى
  2. قراءة في المدخلات والمخرجات المطلوبة
  3. حساب الناتج الفعلي (من خلال العمل للأمام عبر الطبقات)
  4. حساب الخطأ (الفرق بين المخرجات الفعلية والمرغوبة)
  5. تغيير الأوزان من خلال العمل إلى الوراء من خلال الطبقات المخفية
  6. كرر الخطوات من 2 إلى 5 حتى تستقر الأوزان

Illuminating The Black Box Sensitivity Analysis on ANN

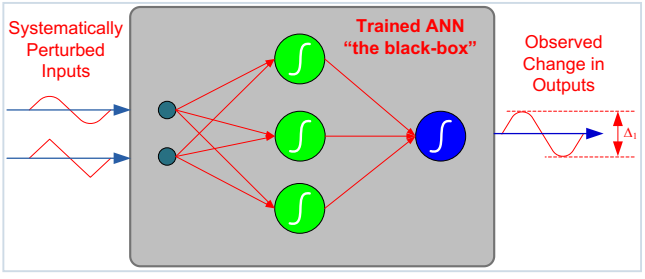
- A common criticism for ANN: The lack of transparency/explainability
- The black-box syndrome!
- Answer: sensitivity analysis
  - Conducted on a trained ANN
  - The inputs are perturbed while the relative change on the output is measured/recorded
  - Results illustrate the relative importance of input variables

إلقاء الضوء على الصندوق الأسود تحليل الحساسية على ANN

- انتقاد شائع لـ ANN: غياب الشفافية / التفسير
- متلازمة الصندوق الأسود!
- الجواب: تحليل الحساسية
  - أجريت على مدربين ANN
  - تشوش المدخلات بينما يتم قياس / تسجيل التغيير النسبي على المخرج
  - توضح النتائج أهمية النسبية لمتغيرات المدخلات

Sensitivity Analysis on ANN Models

- Sensitivity analysis reveals the most important injury severity factors in traffic accidents
- يكشف تحليل الحساسية عن أهم عوامل شدة الخطورة في حوادث المرور



Support Vector Machines (SVM)

- SVM are among the most popular machine-learning techniques.
- SVM belong to the family of generalized linear models… (capable of representing non-linear relationships in a linear fashion).
- SVM achieve a classification or regression decision based on the value of the linear combination of input features.
- Because of their architectural similarities, SVM are also closely associated with ANN.

الات ناقلات الدعم (SVM)

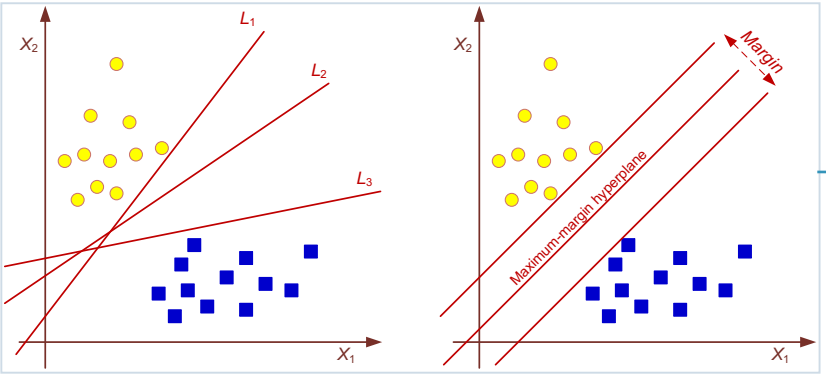
- SVM هي من بين تقنيات التعلم الآلي الأكثر شعبية.
- تنتمي SVM إلى عائلة النماذج الخطية المعممة ... (القادرة على تمثيل العلاقات غير الخطية بطريقة خطية).
- يحقق SVM قرار تصنيف أو تراجع يستند إلى قيمة التركيبة الخطية لميزات الإدخال.
- وبسبب التشابه المعماري او البنية ، ترتبط SVM أيضًا ارتباطًا وثيقًا بـ ANN.

- Goal of SVM: to generate mathematical functions that map input variables to desired outputs for classification or regression type prediction problems.
  - First, SVM uses nonlinear kernel functions to transform non-linear relationships among the variables into linearly separable feature spaces.
  - Then, the maximum-margin hyperplanes are constructed to optimally separate different classes from each other based on the training dataset.
- SVM has solid mathematical foundation!

- الهدف من SVM: لإنشاء دوال حسابية تقوم بتعيين متغيرات الإدخال إلى المخرجات المرغوبة من أجل تصنيف أو مشاكل التنبؤ بنوع التراجع.
  - أولاً ، يستخدم SVM دالة kernel غير الخطية لتحويل العلاقات غير الخطية بين المتغيرات إلى مسافات ميزة قابلة للفصل خطيًا.
  - بعد ذلك ، يتم إنشاء تعامد ذات حد أقصى للهامش لفصل الطبقات المختلفة عن بعضها البعض بشكلٍ مثالي استنادًا إلى مجموعة بيانات التدريب.
- SVM له أساس رياضي متين!

- A **hyperplane** is a geometric concept used to describe the separation surface between different classes of things.
  - In SVM, two parallel hyperplanes are constructed on each side of the separation space with the aim of maximizing the distance between them.
- **A kernel function** in SVM uses the kernel trick (a method for using a linear classifier algorithm to solve a nonlinear problem)
  - The most commonly used kernel function is the radial basis function (RBF).

- **و التعامد (الجبر خطي)** هو مفهوم هندسي يستخدم لوصف سطح الانفصال بين فئات مختلفة من الأشياء.
  - في SVM ، يتم إنشاء تعامديين فرعيين متوازيين على كل جانب من مساحة الفصل بهدف زيادة المسافة بينهما.
- تستخدم **دالة kernel** في SVM خدعة kernel (طريقة لاستخدام خوارزمية تصنيف خطي لحل مشكلة غير خطية)
  - دالة kernel الأكثر استخدامًا هي دالة الأساس الشعاعي (RBF)



- Many linear classifiers (hyperplanes) may separate the data
- قد يفصل العديد من المصنفات الخطية (hyperplanes) البيانات .

How Does an SVM Work?

- Following a machine-learning process, an SVM learns from the historic cases.
- The Process of Building SVM

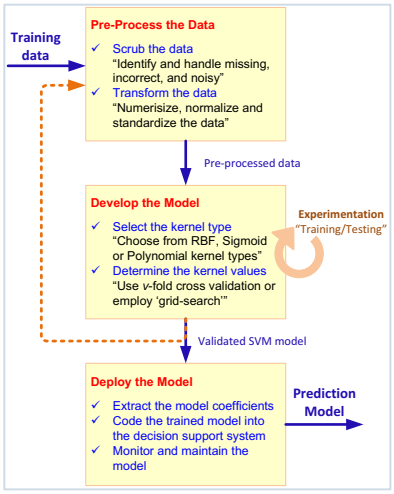
1. Preprocess the data
   - Scrub and transform the data.
2. Develop the model.
   - Select the kernel type (RBF is often a natural choice).
   - Determine the kernel parameters for the selected kernel type.
   - If the results are satisfactory, finalize the model; otherwise change the kernel type and/or kernel parameters to achieve the desired accuracy level.
3. Extract and deploy the model.

كيف تعمل SVM ؟

- بعد عملية تعلم الآلة ، يتعلم SVM من الحالات التاريخية.
- عملية بناء SVM

1. المعالجة المسبقة للبيانات
   - تنظيف وتحويل البيانات.
2. تطوير النموذج.
   - حدد نوع kernel (RBF غالبًا ما يكون اختيارًا طبيعيًا).
   - حدد عوامل kernel لنوع kernel المحدد.
   - إذا كانت النتائج مرضية ، قم بوضع الصيغة النهائية للنموذج ؛ خلاف ذلك تغيير نوع kernel و / أو عوامل kernel لتحقيق مستوى الدقة المطلوبة.
3. استخراج ونشر النموذج.

The Process of Building an SVM



SVM Applications

- SVMs are the most widely used kernel-learning algorithms for wide range of classification and regression problems
- SVMs represent the state-of-the-art by virtue of their excellent generalization performance, superior prediction power, ease of use, and rigorous theoretical foundation
- Most comparative studies show its superiority in both regression and classification type prediction problems.
- SVM versus ANN?

تطبيقات SVM

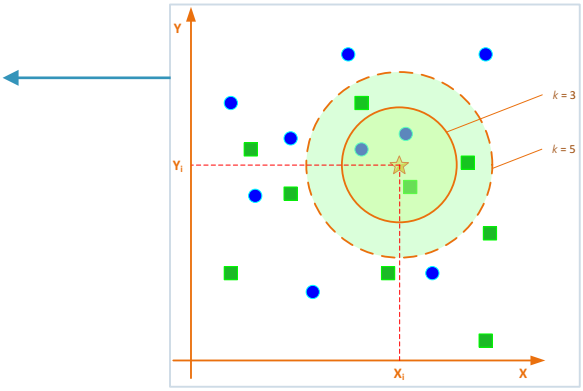
- SVMs هي أكثر خوارزميات تعلم kernel المستخدمة على نطاق واسع لمجموعة واسعة من مشاكل التصنيف والتراجع
- تمثل SVMs أحدث ما توصلت إليه التكنولوجيا بفضل أداءها الممتاز في التعميم ، وقوة التنبؤ الفائقة ، وسهولة الاستخدام ، والأساس النظري الصارم
- تظهر معظم الدراسات المقارنة تفوقها في كل من مشاكل التنبؤ بالتراجع والتصنيف.

k-Nearest Neighbor Method (k-NN)

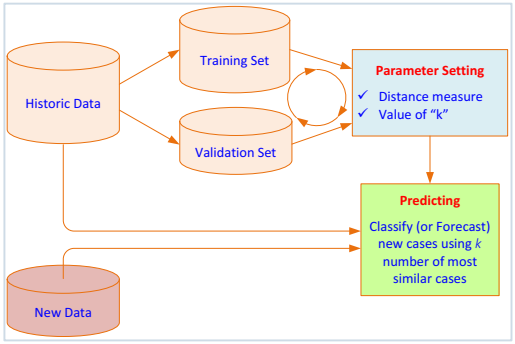
- ANNs and SVMs → time-demanding, computationally intensive iterative derivations
- *k*-NN is a simplistic and logical prediction method, that produces very competitive results
- *k*-NN is a prediction method for classification as well as regression types (similar to ANN & SVM)
- *k*-NN is a type of instance-based learning (or lazy learning) – most of the work takes place at the time of prediction (not at modeling)
- *k* : the number of neighbors used

- ANNs و SVMs ← التي تتطلب وقت ، اشتقاق متكرر مكثف حسابياً
- k-NN هي طريقة تنبؤ مبسطة ومنطقية ، تنتج نتائج تنافسية للغاية
- k-NN هي طريقة تنبؤ للتصنيف بالإضافة إلى أنواع التراجع المشابهة لـ (ANN & SVM)
- k-NN هو نوع من التعلم القائم على التثقيف (أو التعلم البطيء) - معظم العمل يحدث في وقت التنبؤ (وليس في النمذجة)
- k :عدد المتجاور المستخدم

- The answer depends on the value of *k*
- الجواب يعتمد على قيمة k



The Process of *k*-NN Method



k-NN Model Parameter

1. Similarity Measure: The Distance Metric → مقاييس التشابه : مقاييس المسافة

Minkowski distance

$$d(i,j) = \sqrt[q]{(|x_{i1} - x_{j1}|^q + |x_{i2} - x_{j2}|^q + \ldots + |x_{ip} - x_{jp}|^q)}$$

If $q = 1$, then d is called Manhattan distance

$$d(i,j) = \sqrt{|x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \ldots + |x_{ip} - x_{jp}|}$$

If $q = 2$, then d is called Euclidean distance

$$d(i,j) = \sqrt{(|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \ldots + |x_{ip} - x_{jp}|^2)}$$

   - Numeric versus nominal values? → ؟ القيم العددية مقابل القيم الأسمية

2. Number of Neighbors (the value of *k*)
   - The best value depends on the data
   - Larger values reduce the effect of noise but also make boundaries between classes less distinct
   - An "optimal" value can be found heuristically

- Cross Validation is often used to determine the best value for k and the distance measure

2 . عدد الجيران (هي قيمة k)
   - أفضل قيمة تعتمد على البيانات
   - القيم الأكبر تقلل من تأثير الضوضاء ولكنها أيضًا تجعل الحدود بين الطبقات أقل وضوحًا
   - يمكن العثور على قيمة "مثالية" بشكل تجريبياً
- وكثيرا ما يستخدم التحقق من صحة لتحديد أفضل قيمة ل k ومقياس المسافة