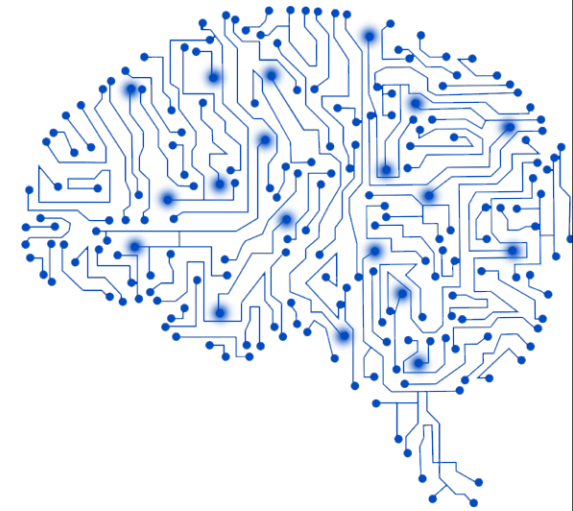


Python Programming Machine Learning Lecture 01

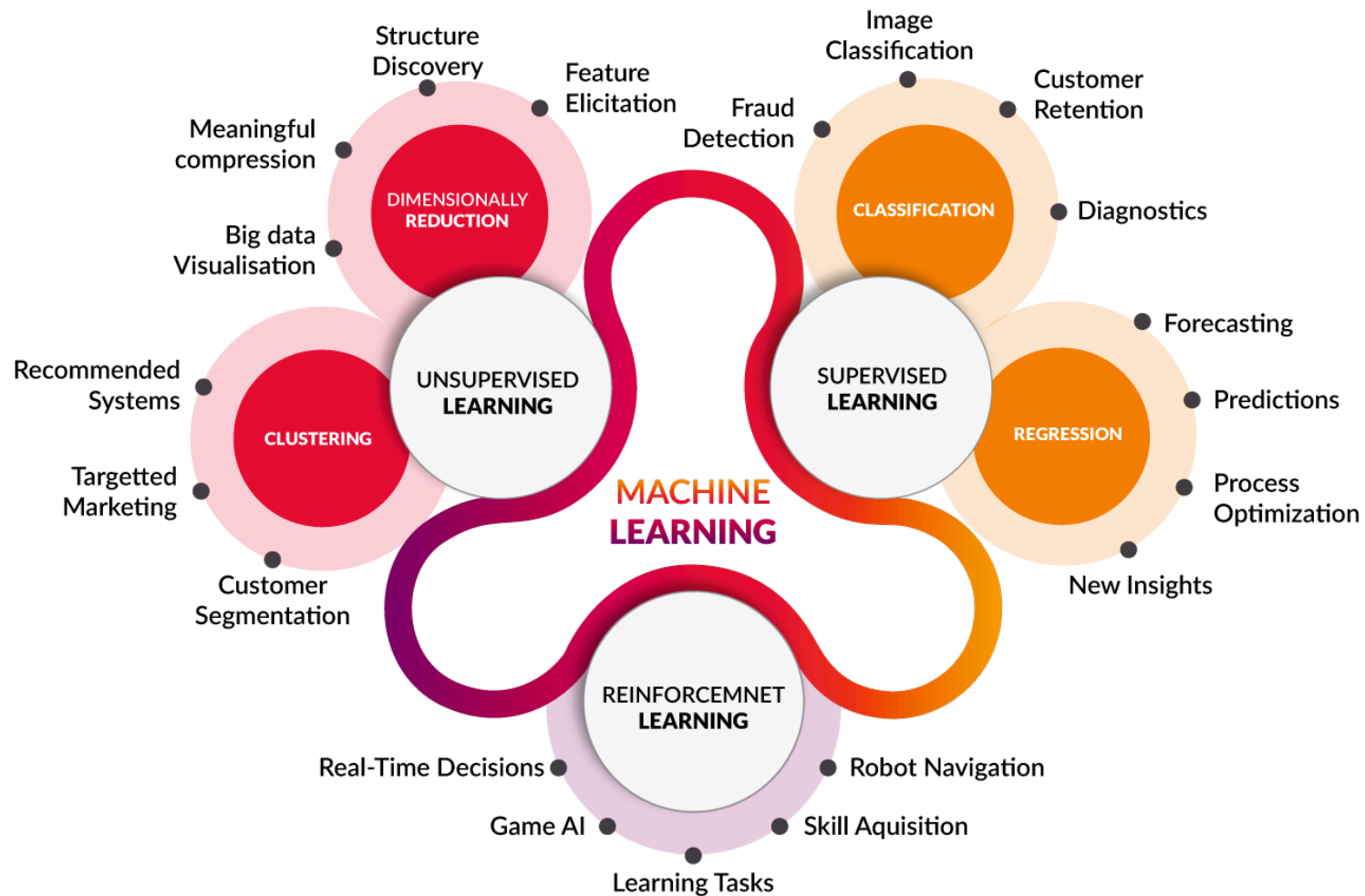
Min-Kuan Chang
GICE, EECS



機器學習簡介

機器學習簡介

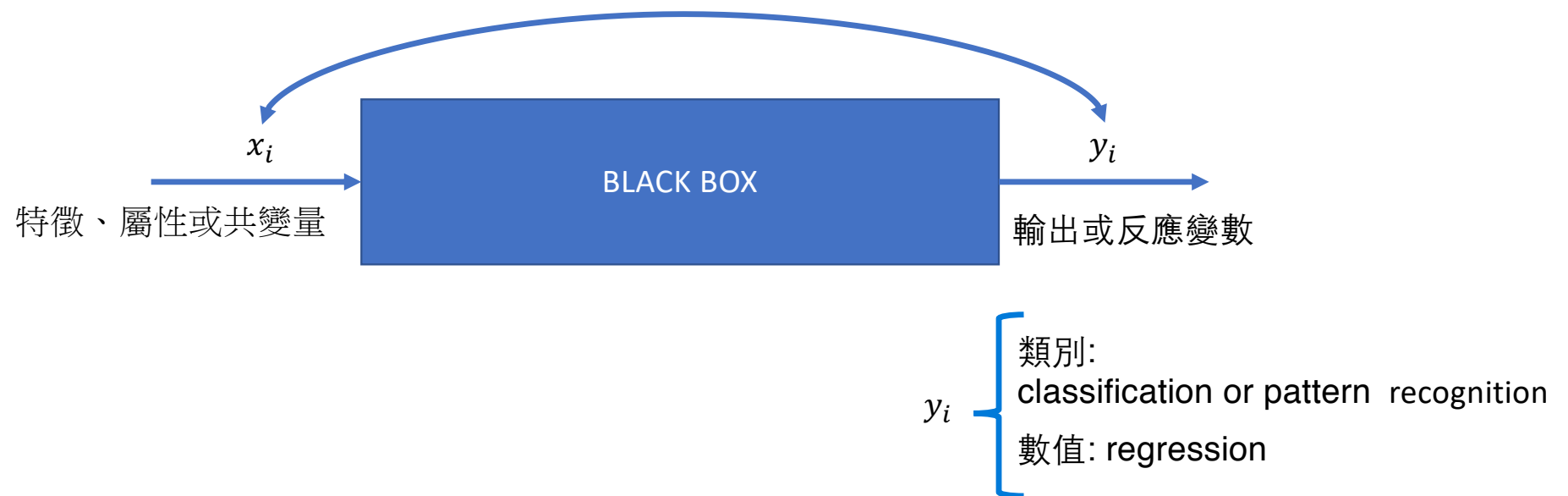
- 大數據時代的到來
 - 美國零售商Walmart 每小時須處理一百萬的交易且該公司的資料庫的容量約可儲存 2.5 petabytes (2.5×10^{15}) 的資料
 - 自動化資料分析的需求日益增加
 - 機器學習開始展露頭角



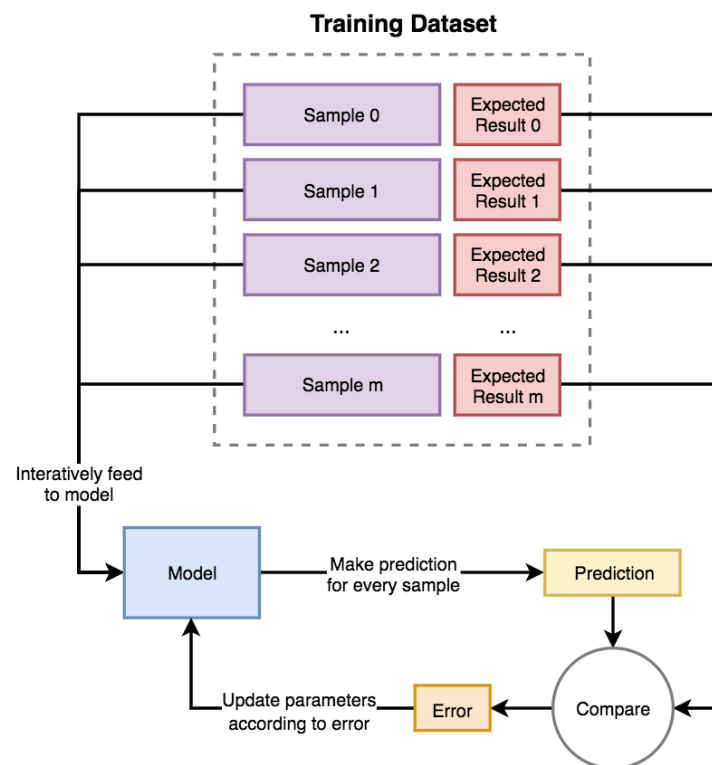
<https://towardsdatascience.com/coding-deep-learning-for-beginners-types-of-machine-learning-b9e651e1ed9d>

機器學習簡介

- 監督式學習

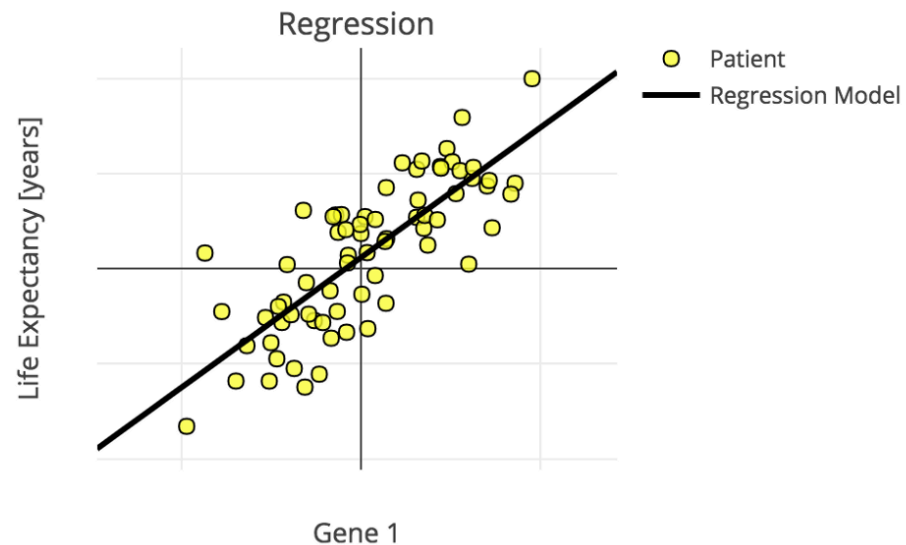
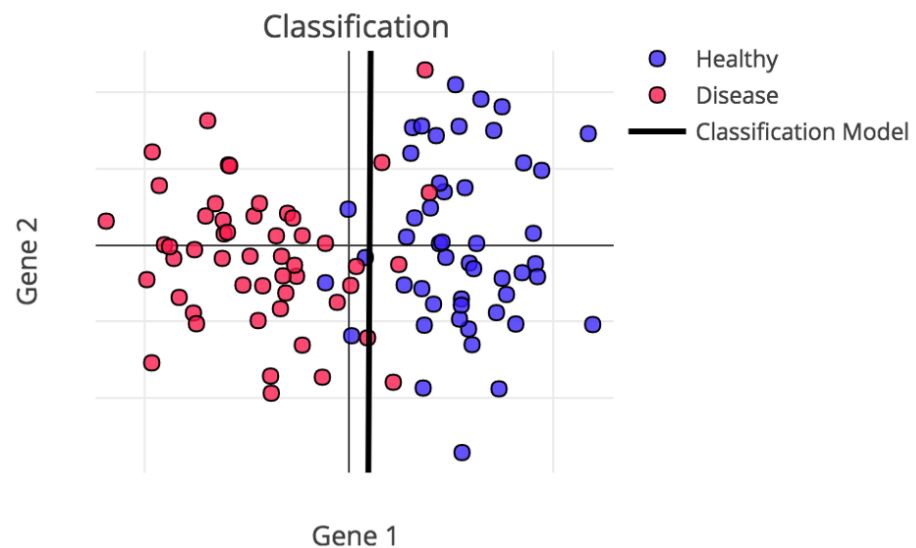


機器學習簡介



<https://towardsdatascience.com/coding-deep-learning-for-beginners-types-of-machine-learning-b9e651e1ed9d>

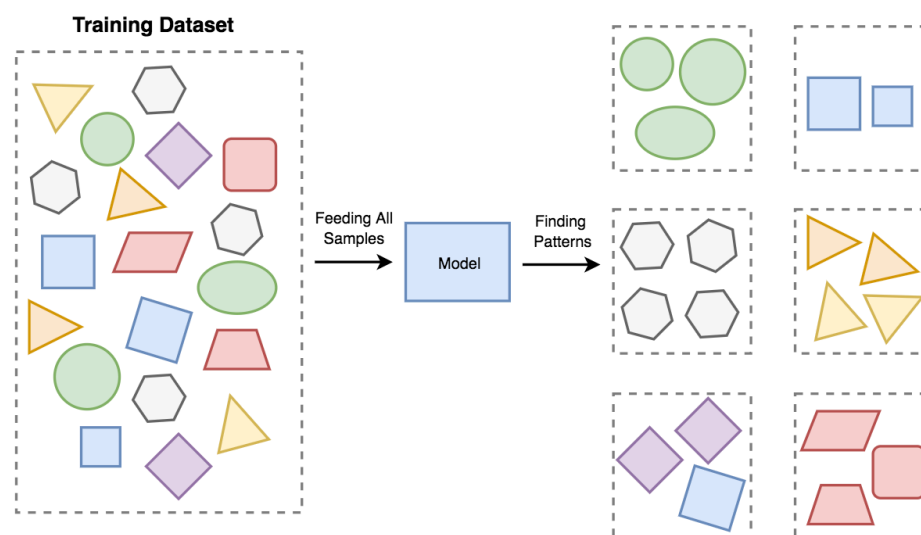
機器學習簡介



<https://towardsdatascience.com/coding-deep-learning-for-beginners-types-of-machine-learning-b9e651e1ed9d>

機器學習簡介

- 非監督式學習



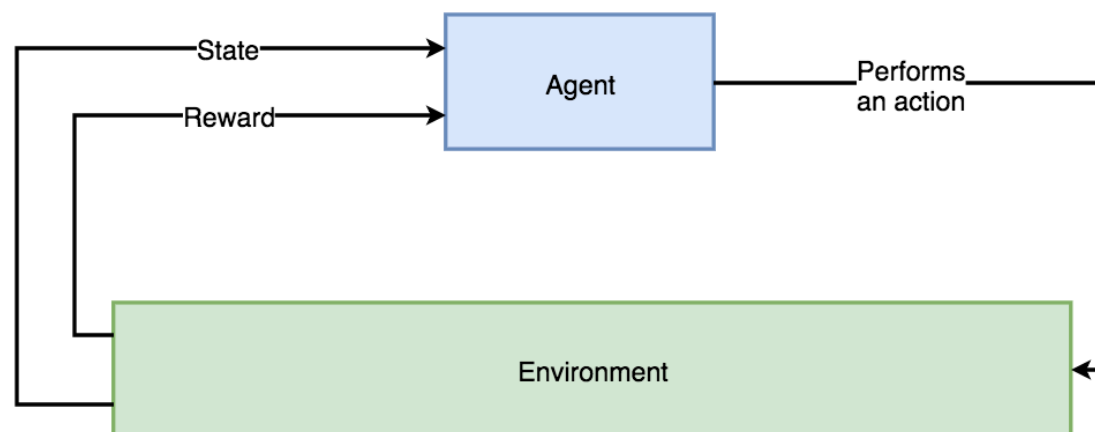
<https://towardsdatascience.com/coding-deep-learning-for-beginners-types-of-machine-learning-b9e651e1ed9d>

機器學習簡介

- 常見的非監督式學習方法:
 - 圖形辨識與資料分群
 - 將類似的資料歸類在一起，進而將資料分成許多不同的類別
 - 這些類別通常稱為群
 - 例如: 顧客分群、將訊號中的雜訊去除
 - 減少資料維度
 - 資料的維度定義為將資料完整呈現所需的特徵數目
 - 資料的降維是將資料中的特徵進行壓縮，保留較為重要的特徵而捨去不重要的特徵

機器學習簡介

- 強化學習



機器學習的一些基本概念

- 參數模型 vs 非參數模型
 - 參數模型
 - 有固定數量的參數
 - 優點：快速使用
 - 缺點：對於資料的本質做了過強的假設
 - 非參數模型
 - 參數的數量會隨著訓練資料的增長而增加
 - 較有彈性但是容易因為資料量過大而造成計算上的負擔
 - 例如：K nearest neighbor (KNN) 分類器

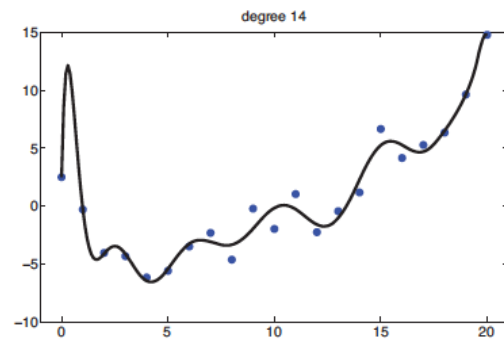
機器學習的一些基本概念

- 推論的偏差
 - 或稱為學習偏差
 - 學習者對於未曾蒙面的資料所做的統計上的假設，並且基於這樣的假設來預測結果
 - Example:
 - 奧卡姆剃刀法則(Occam's razor)
 - 用最簡單的假設來面對問題
 - 當有很多不同的假設可以用來解決當前的問題，選擇最少假設的那個方法
 - https://en.wikipedia.org/wiki/Occam%27s_razor
 - <http://inductivebias.com/Blog/what-is-inductive-bias/>
 - https://en.wikipedia.org/wiki/Inductive_bias

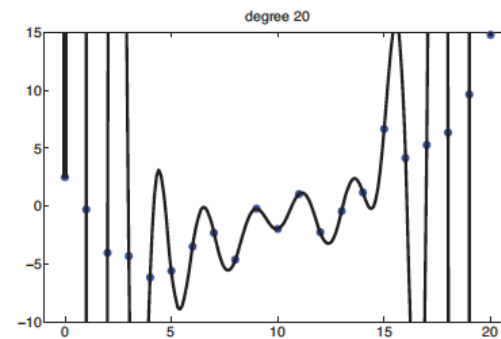
機器學習的一些基本概念

- 過適 (Overfitting)

- 當我們使用高度有彈性的模型時，我們得要注意不讓學習的結果太過貼近資料的樣子
- 我們應該避免將學習的模型來貼近任一個可能發生的偏差，因為這些偏差有可能是來自於干擾的雜訊

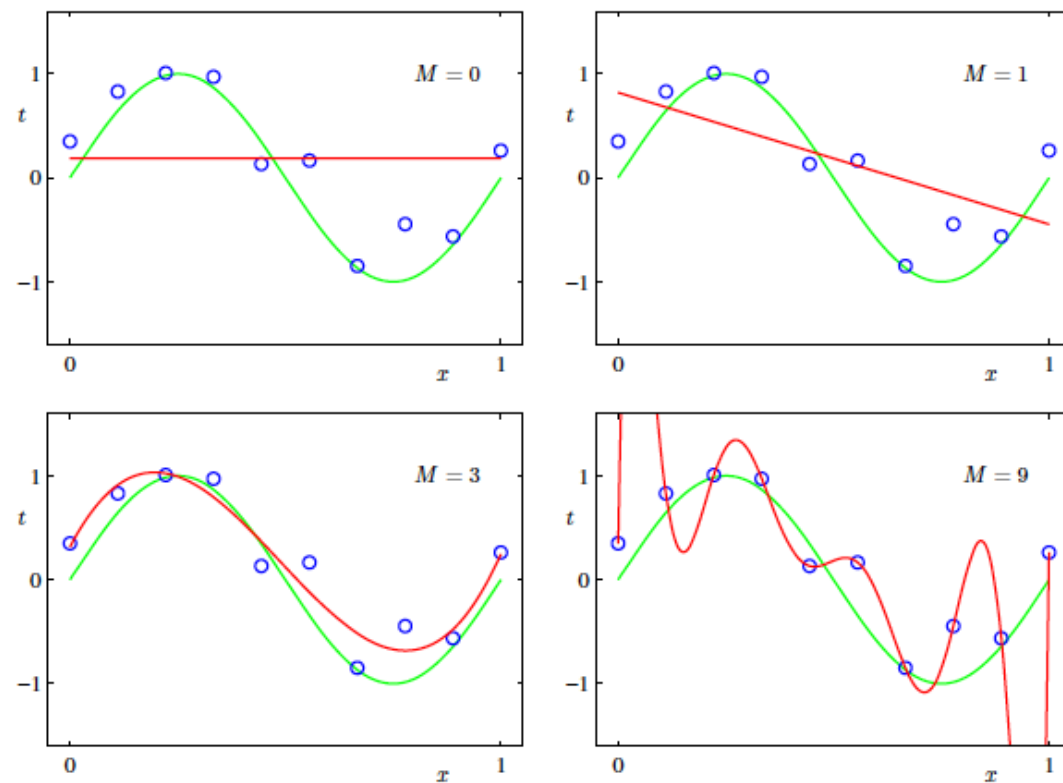


(a)

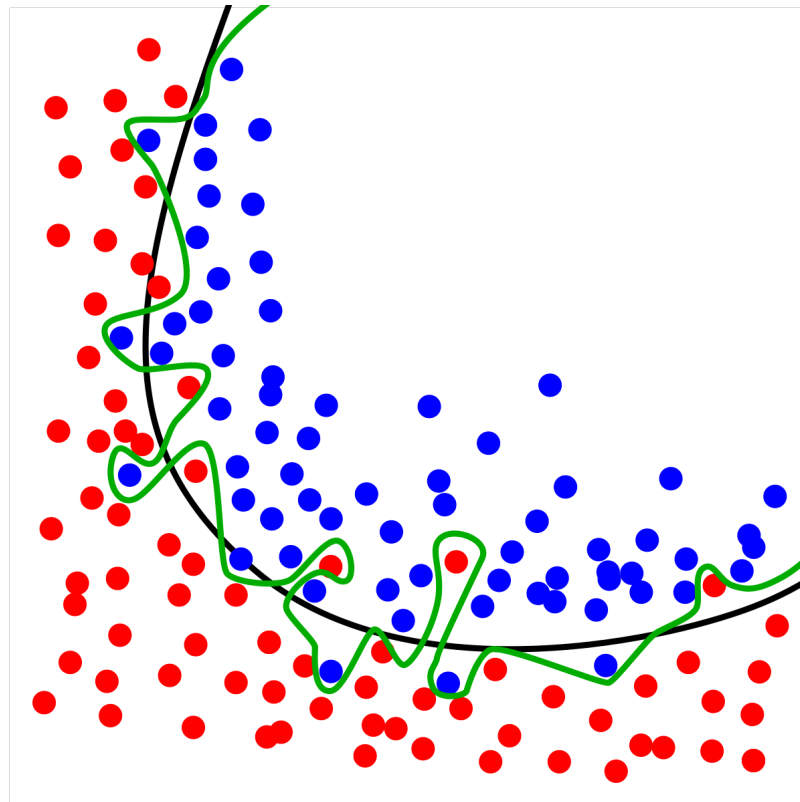


(b)

機器學習的一些基本概念



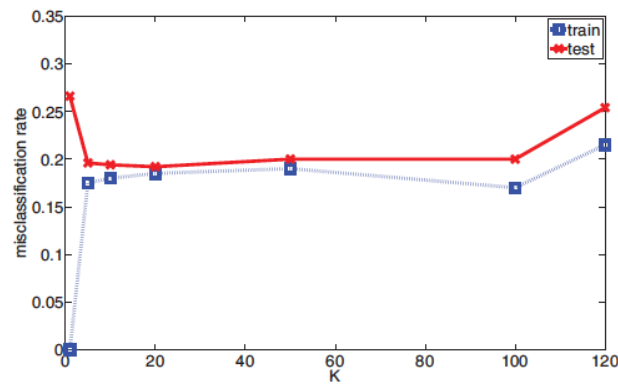
機器學習的一些基本概念



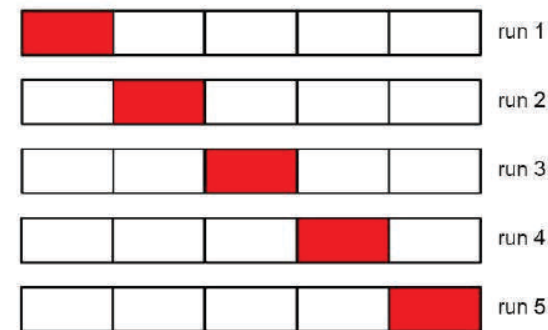
<https://www.apixio.com/engineering/introduction-to-machine-learning/>

機器學習的一些基本概念

- 如何避免過適 (How to prevent overfitting)
 - 交叉驗證(Cross-validation)

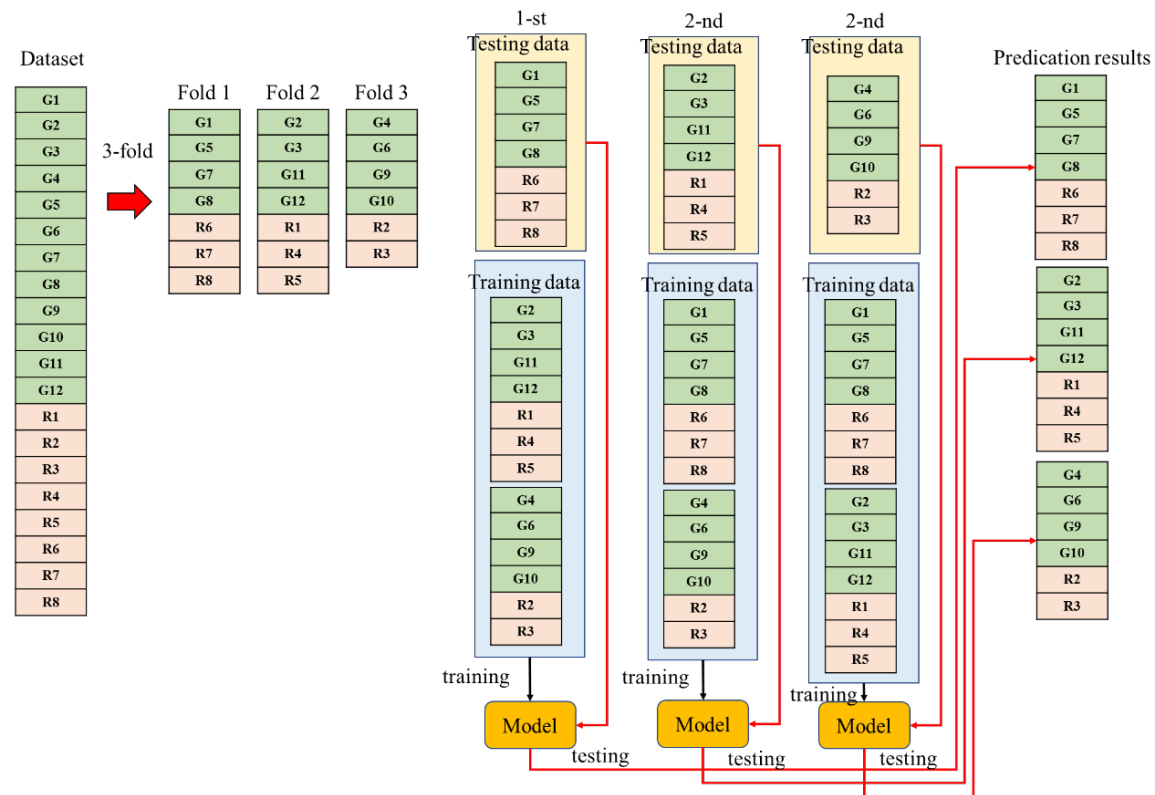


(a)



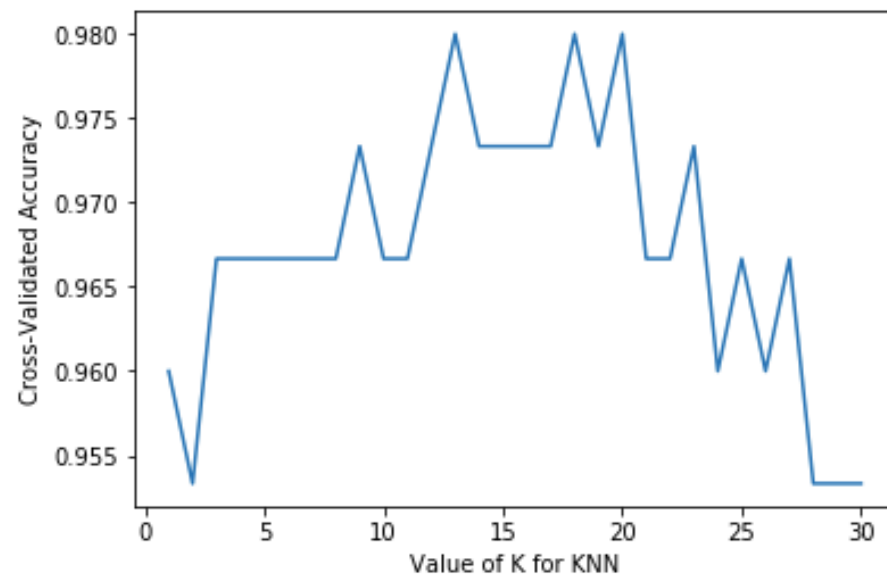
(b)

機器學習的一些基本概念



<https://medium.com/@chih.sheng.huang821/%E4%BA%A4%E5%8F%89%E9%A9%97%E8%AD%89-cross-validation-cv-3b2c714b18db>

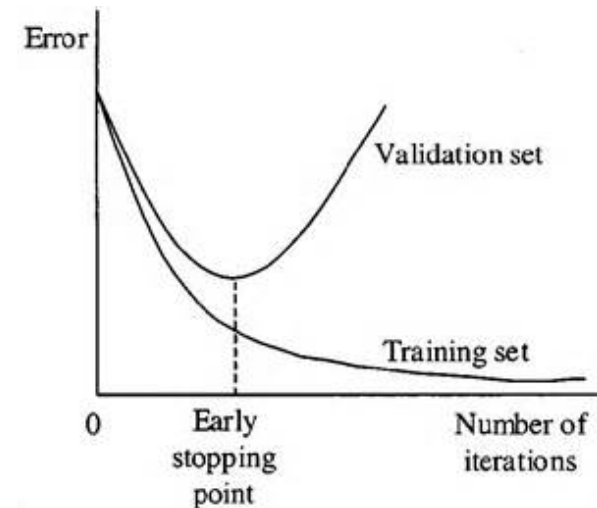
機器學習的一些基本概念



<https://ithelp.ithome.com.tw/articles/10197461>

機器學習的一些基本概念

- 如何避免過適 (How to prevent overfitting)
 - 使用更多訓練資料
 - 移除某部分特徵
 - 提早結束(Early stopping)
 - 正規化(Regularization)
 - 集成(Ensembling)



<https://elitedatascience.com/overfitting-in-machine-learning>



In “Nature” 27 January 2016:

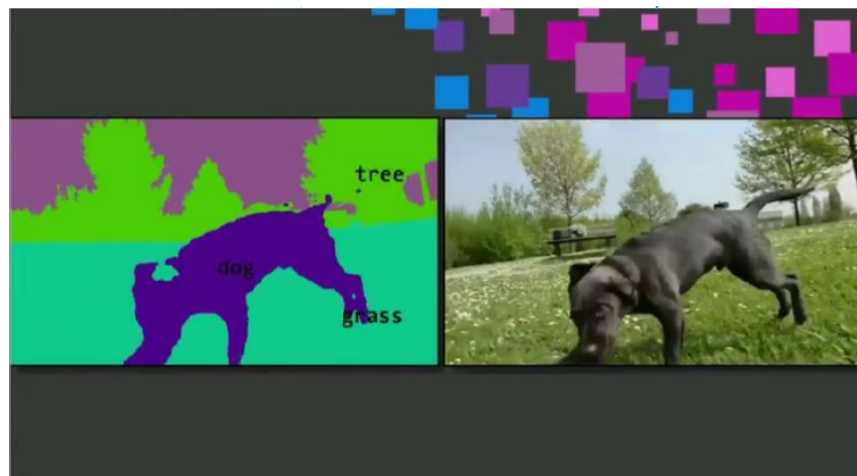
- “DeepMind’s program AlphaGo beat Fan Hui, the European Go champion, five times out of five in tournament conditions...”
- “AlphaGo was not preprogrammed to play Go: rather, it learned using a general-purpose algorithm that allowed it to interpret the game’s patterns.”
- “...AlphaGo program applied **deep learning** in neural networks (convolutional NN) — brain-inspired programs in which connections between layers of simulated neurons are strengthened through examples and experience.”

機器學習的例子



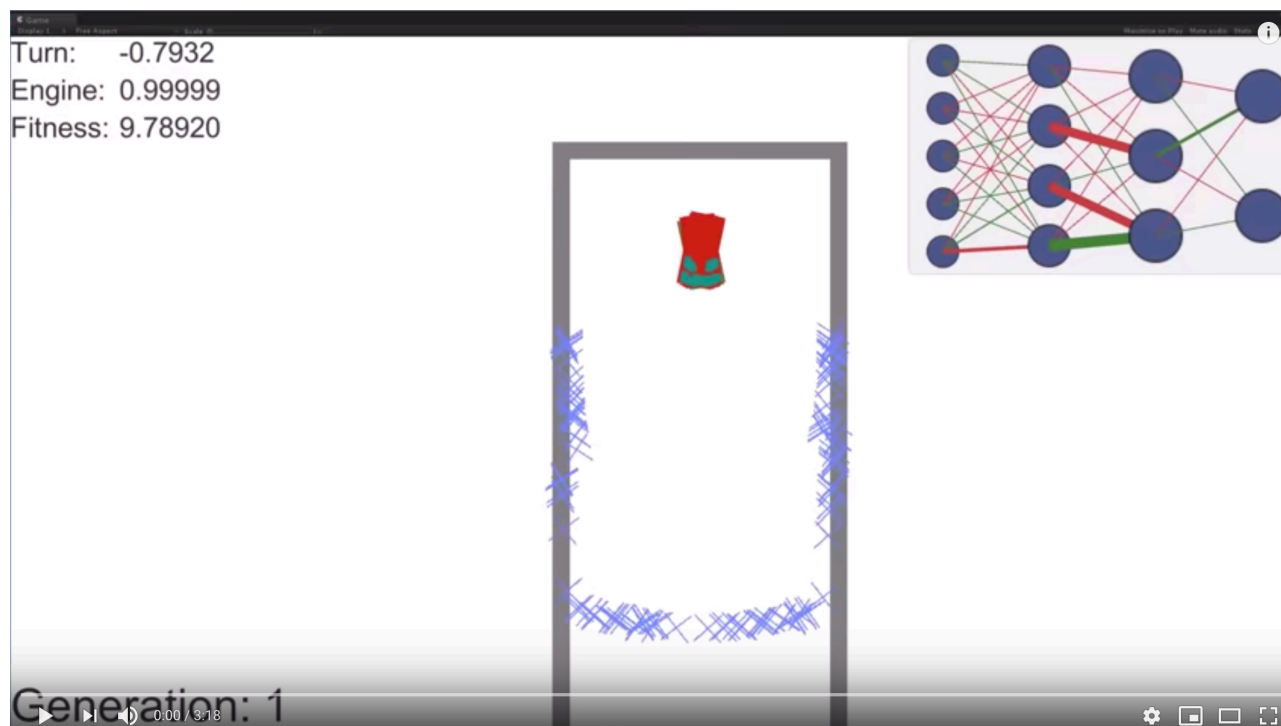
<https://www.youtube.com/watch?v=V1eYniJ0Rnk>

機器學習的例子



https://www.youtube.com/watch?v=CxanE_W46ts

機器學習的例子



<https://www.youtube.com/watch?v=Aut32pR5PQA>

Q&A