**機器學習[Machine learning** (**ML**)**]**

**機器學習是一門人工智慧的科學，該領域的主要研究物件是如何在經驗學習中改善具體演算法的效能。**

**機器學習是對能通過 經驗 自動改進 的 電腦演算法 的研究。**

**機器學習是用 資料 或 以往的經驗，以此最佳化電腦程式的效能標準。**

**A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E**

**機器學習類型:**

|  |  |  |
| --- | --- | --- |
| **類型** | **說明** | **代表性演算法** |
| **監督式學習**  **Supervised learning** | **由訓練資料中學到或建立一個模式(函數/learning model)，並依此模式推測新的實例。**  **訓練資料(labeled Training data)是由輸入物件（通常是向量）和預期輸出所組成。**  **函數的輸出可以是一個連續的值(迴歸分析regression)，或是預測一個分類標籤(分類classification)** | **迴歸分析regression**  **Linear regression**  **Logistic regression**  **分類演算法classification**  **決策樹Decision trees**  **支援向量機(SVM)**  **Support vector machine** |
| **非監督式學習**  **Unsupervised learning** | **機器研究輸入的資料，多數是未標記(no labeled)與非結構化的資料，並開始使用所有相關且可存取的資料來識別模式和關聯性。**  **非監督式學習是在模仿人類如何觀察世界。我們運用直覺和經驗將事情分類，而隨著經歷更多體驗，分類和識別的能力會越來越精確；對於機器而言，「經驗」則是輸入和可用的資料量。** | **Clustering叢集演算法**  **k-Means…..**  **維度縮減Dimensionality reduction**  **主成分分析PCA**  **自動編碼器Autoencoder**  **生成對抗網路(GAN)** |
| **半監督式學習**  **Semi-supervised learning** | **Combines a small amount of labeled data with a large amount of unlabeled data during training.**  **Semi-supervised learning falls between unsupervised learning (with no labeled training data) and supervised learning (with only labeled training data)** | **transductive support vector machine(TSVM)**  **部分特殊型Deep Learning演算法** |
| **Reinforcement learning** | **an area of machine learning concerned with how intelligent agents ought to take actions in an environment in order to maximize the notion of cumulative reward.** | **蒙特卡洛學習 Monte-Carlo Learning**  **Temporal-Difference Learning**  **SARSA**  **Q-Learning** |

**分類演算法效能評估指標;** **混淆矩陣(confusion matrix)**

**【資料來源】https://en.wikipedia.org/wiki/Confusion\_matrix**

如何辨別機器學習模型的好壞？==> Confusion Matrix

**TP(True Positive): 正確預測成功的正樣本 TN(True Negative): 正確預測成功的負樣本**

**FP(False Positive): 錯誤預測成正樣本(實際上為負樣本) 第一型錯誤 (Type 1 Error)**

**FN(False Negative): 錯誤預測成負樣本(或者說沒能預測出來的正樣本) 第二型錯誤(Type 2 Error)**

**在統計學上FP被還被稱為第一型錯誤(Type 1 Error)，FN被稱為第二型錯誤(Type 2 Error)**

|  |  |
| --- | --- |
|  | **全部樣本數 = P + N**  **condition positive (P) P = TP +FN**  **the number of real positive cases in the data**  **condition negative (N) N = FP +TN**  **the number of real negative cases in the data**  **正確預測 = TP + TN 錯誤預測 = FP + FN** |
|  |  |