**Task #08 作業 01【簡答題】機器學習基本觀念**

1. 機器學習可以分成「監督式學習」與「非監督式學習」，請分別舉一個應用的例子。

監督式學習是指在訓練過程中，資料樣本標有正確的標籤或結果，機器學習模型學習從輸入數據中將輸入和輸出關係建立起來的學習方式。

例如給定大量已經標註好的影像(如各種動物圖像)，讓模型透過label過的影像學習各個動物的特徵後，下次可給新的一張動物影像讓模型進行分類成某一動物。

非監督式學習是指在訓練過程中，資料樣本沒有標籤或結果，機器學習模型通過尋找資料的結構和相似性來學習。

例如給定大量未標註的影像，讓模型自行從中進行分群的動作。

2. 線性回歸  
（Linear Regression）能夠利用一個條線代表一組資料的分佈，請問線性回歸是如何找出代表資料點的線？

Linear Regression會利用找尋模型的最佳參數，使得預測值與實際值得誤差最小，例如兩個參數的數學模型y= ax+b，模型可以透過訓練資料找到最佳狀況的a及b，使得測試資料x得出的預測值y可以與實際值最接近。

3. 通常越複雜的模型越有可能出現過擬合（Overfitting）的狀況，請問什麼是 Overfitting 呢？實際上又該怎麼避免或解決？

Overfitting是指模型在訓練資料上表現得很好，在測試資料上表現很差的狀況，通常這種狀況發生代表可能模型太複雜，或是訓練資料量太少，讓模型無法學到資料的特徵。

解決方法可以嘗試簡化模型的參數數量(ex dropout方法)、或是增加數據量(收集更多數據或是利用data augmentation)，也可以利用early stop方法提早讓模型停止學習，避免模型把訓練資料全部用背起來的方式。

4. 承上題，要如何從模型的結果中觀察到可能出現過擬合（Overfitting）的狀況？

可以利用訓練資料及驗證validation資料的loss結果觀察，當training data的loss有持續下降而validation的誤差不再下降或反而上升，則代表有overfitting的發生。

5. PCA 模型能夠將高維度的資料集轉換成比較低的維度，請問什麼情況下該這麼做？

當數據有大量的特徵時，可利用PCA降維，去除不重要的特徵，以便於模型減少計算複雜度或是無法學習導致overfitting的狀況。