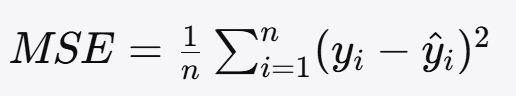
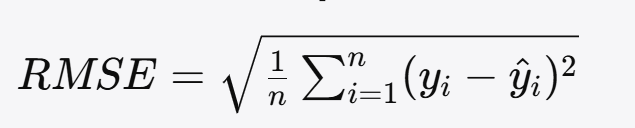
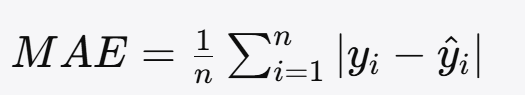
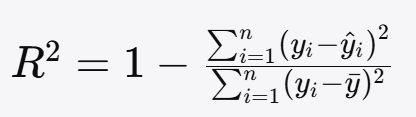
MOAI Theoretical Part

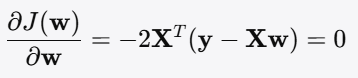
Jimmy Choi, May, 2025

Part I. Machine Learning / Scikit-learn

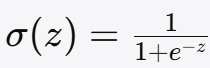
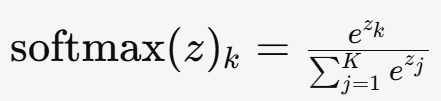
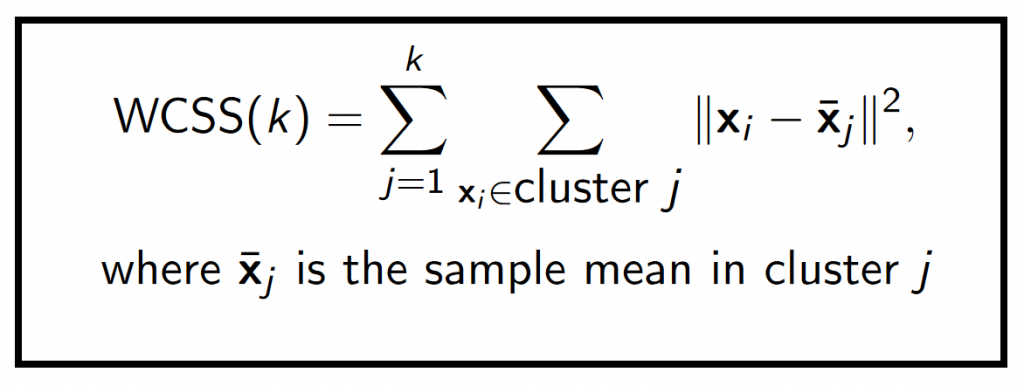
1. Supervised learning: Linear regression綫性回歸
   1. 主要目的：尋找最佳擬合數據的直綫（一元簡單綫性回歸）或者超平面（多元綫性回歸）
   2. 模型表示：y = wx + b（簡單綫性回歸）或者y=∑ni=1 wixi + b（多元綫性回歸）
   3. Loss function損失函數：
      1. Mean Squared Error, MSE 
      2. Root Mean Squared Error, RMSE 
      3. Mean Absolute Error, MAE 
      4. R-squared, R^2 
   4. Optimization Techniques
      1. Least Squares Method最小二乘法
         1. 使用MSE作爲loss function，完全不需要进行迭代，只需要一次计算就可以得到最优的权重向量 w
         2. 計算方法：

將公式用矩陣形式表示，即 y = Xw + ϵ，其中y表示真實值向量，X表示輸入的特徵矩陣，w表示權重矩陣，ϵ表示所有樣品的誤差向量

將損失函數用矩陣表示，即J(w) = (y - Xw)T(y - Xw)

爲了最小化J(w)，需要找到使J(w)對w的導數為0的w，即，解得，就是最小二乘法的解析解(Analytical Solution)，也即閉合形式解(Closed-form Solution)

* + 1. Gradient Descent梯度下降：見Part II

1. Supervised Learning: Logistic regression邏輯回歸
   1. 主要思想：用于分类问题，通过 Sigmoid 函数将线性回归的输出映射到 0 和 1 之间，表示属于某个类别的概率。
   2. 二分類
      1. 使用Sigmoid函數作爲激活函數：，其中z = wTx + b
      2. Decision Threshold決策閾值：通常設定爲0.5，大於0.5屬於類別1，小於0.5屬於類別0
      3. Loss function損失函數：Binary Cross-Entropy二元交叉熵
      4. 優化方法：梯度下降
   3. 多分類：
      1. 使用softmax函數作爲激活函數，將k個綫性組合映射到k個概率值，這些概率值的和為1，選擇概率最高的類別作爲預測結果
      2. Loss function損失函數：Categorical Cross-Entropy多類別交叉熵
      3. 優化方法：梯度下降
2. Supervised Learning: KNN（K nearest neighbor algorithm) K近鄰算法
   1. 主要思想：找到距離目標樣本最近的 K 个鄰居，然後根據這些鄰居的類別來預測目標樣本的類別
   2. 算法步驟：
      1. 選擇k的大小：過小發生overfitting，過大發生underfitting
      2. 計算距離：Euclidean Distance, Manhattan Distance, Minkowski Distance, Cosine Similarity（一般使用前兩種方法）
      3. 尋找k個最近的鄰居，選擇頻率最高的類別分類
3. Supervised Learning: Decision Tree決策樹
   1. 主要思想：通過一系列的決策規則，將數據集逐步劃分成更小的子集，直到每個子集中的樣本都屬於同一類別（或具有相似的數值）。其中，每个节点表示一个特征，每个分支表示该特征的一个取值，每个叶节点表示一个类别（或一个数值）。
   2. 算法步驟：
      1. 計算數據集的Gini impurity：，其中K為類別數，pk是數據集中第k類樣本的所占比例
      2. 計算每個特徵劃分后的weighted Gini impurity:,A表示一種特徵，有V種可能的取值，將D劃分成V個子集，計算每個子集的weighted Gini impurity和。
      3. 就計算每個特徵的信息增益：，選擇信息增益最大的為當前節點的劃分特徵
      4. 重複上述步驟，直到滿足下述條件之一：
         1. 當前節點的所有數據屬於同一類別；
         2. 沒有更多的特徵可用於分割；
         3. 達到預定義的樹的最大深度；
         4. 當前節點的數據量小於預定義的閾值。
4. Supervised Learning: Random Forest隨機森林
   1. 主要思想：構建多個決策樹，組合他們的預測。“隨機”體現在隨機選擇樣本和隨機選擇特徵（并不選擇所有特徵）上。
   2. 算法步驟：
      1. Bootstrap抽樣：從原始訓練集中隨機有放回地抽取 N個樣本，生成一個新的訓練集。 重複此步驟T次，生成T個訓練集。
      2. 訓練決策樹：隨機選擇m個特徵，通常m=M^(1/2)，M是原始特徵的數量，再依照決策樹的方法訓練決策樹
      3. 預測：classification取最多預測結果，regression取平均值
5. Unsupervised Learning: K-means
   1. 主要思想：將數據集無監督地分成k個不同的cluster，使得每個不同的樣本都屬於離它最近的centroid對應的cluster。
   2. 算法步驟：
      1. 選擇一個k，并且隨機選擇k個樣本/k個點作爲初始的centroid。
      2. 對於每個樣本，計算與k個centroid的距離，并且分配到最近的centroid對應的cluster。
      3. 重新計算每一個cluster的centroid。
      4. 重複上述步驟，直到滿足下述條件之一：
         1. Centroid不再變化；
         2. Cluster的分配不再變化；
         3. 達到預定義的最大迭代次數。
   3. 損失函數：Within-Cluster Sum of Squares, WCSS（簇内平方和）其中xj就是第j個cluster的centriod。用Euclidean distance。
   4. K值的選擇：（Elbow Method）繪製k與WSCC的關係圖，選擇WSCC下降變緩的“肘部”。
6. Unsupervised Learning: PCA (Principle Component Analysis)主成分分析
7. Model Evaluation模型分析指標
8. Overfitting & Underfitting過擬合和欠擬合

Part II. Deep Learning and Neutral Network / PyTorch

1. Gradient Descent梯度下降
2. BP (Backpropagation)反向傳播
3. Activation function激活函數
4. Loss function損失函數
5. MLP (Multilayer perceptron)多層感知機
6. SGD (Stochastic Gradient Descent)隨機梯度下降
7. Adam
8. 學習率

Part III. Computer Vision / PyTorch

1. CNN (Convolution Neutral Network)卷積神經網絡
2. Image Classification圖像分類

Part IV. NLP (Natural Language Processing)自然語言處理

1. Transformer and Attention