**機器學習導論 期中報告**

**Part A.**

1. 在一次線性回歸中，只有一個係數，而截距會自動存放至 intercept\_ 中，所以lr.coef\_[0] 就是迴歸係數θ1，在二次線性回歸中截距則會被存放在 lr.coef\_[0] 的位置，所以迴歸係數θ1會在lr.coef\_[1]。
2. 在Elastic Net 中的損失函數可以看出 alpha和 l1\_ratio分別是設定 Ridge 跟 LASSO 的值，根據題目給的函數的係數為1/3 為LASSO 的 1/4 為 Ridge，所以alpha和 l1\_ratio設定為 1/4 (0.25)與 1/3(0.33)
3. 在建立決策樹時，無須先對數執行特徵做正規化，因為資料數值縮放不影響決策樹節點位置，對樹模型的結構不造成影響。

**Part B.**

1. 隨機梯度下降法

   2. i = 2

i = 3

i = 4

i = 5

* 1. 會產生不同的模型，因為不同順序帶入值，每次計算都會產生不同的值，最後算出也會有不同的模型出來。

Part 3.

1. 比較決策樹 ID3, C4.5 CART

|  |  |  |  |
| --- | --- | --- | --- |
| **演算法** | **資料屬性** | **分割規則** | **修剪樹規則** |
| ID3 | 離散型 | Information Gain | Gain Ratio Rate |
| C4.5 | 離散型 | Gain Ratio | Gain Ratio Rate |
| CART | 離散與連續型 | Gini index | Entire Error Rate |

1. 什麼是過擬合?如何降低?

過擬合是指訓練的模型對驗證資料集的Performance很差，但是對訓練資料集的Performance卻很好。通常會從損失曲線圖來觀察是否有過擬合現象。

可以透過縮減模型大小、加入權重正規化，在神經網路的話可以使用dropout。

1. 比較梯度下降法和隨機梯度下降法

GD是一次用**全部訓練集**的數據去計算損失函數的梯度，每一次下降就更新一次參數。

SGD是一次跑**一個樣本**然後算出一次梯度平均後就更新一次，那這個樣本是隨機抽取的。

1. Precision 與 Recall很難同時兼得，題目中表示Precision 很高，代表 Recall 可能偏低，而 Recall 的意義在於在所有的目標中總共找出了多少個的比例，因此才會出現使用者卻還是常常反應找不到他們想要物品的問題。
2. 將樣本數提高，因為樣本數提高時，在建立模型時有更好的資料能夠去建立分類模型，再拿測試資料去測試時，才能夠把 TP Rate 提升，FP Rate 降低，此時的，折線就會越趨向平滑。
3. (1) Logistic Regression，可以正確判斷，但需要兩個或以上的羅吉斯回歸才可判斷，依照此數據無法用一條回歸線正確分類成兩類，但可以先用一條回歸線將全部先分成左右或是上下，再用另一條回歸線正確切割分類，即可獲得正確分類。

(2) Decision Tree，可以正確判斷，因為只要判斷兩次就能知道是屬於O 或是 X，比如判斷 x軸是否大於0.5，若有再判斷y 值是否有大於0.5，若有，即可分類為O，反之為X，但若有新的資料產生界在於0.5左右時，決策樹判斷可能會失準。