ML\_HW1

s1041501 黃威竣



以上為我的Python code

使用的平台為Anaconda Jupyter

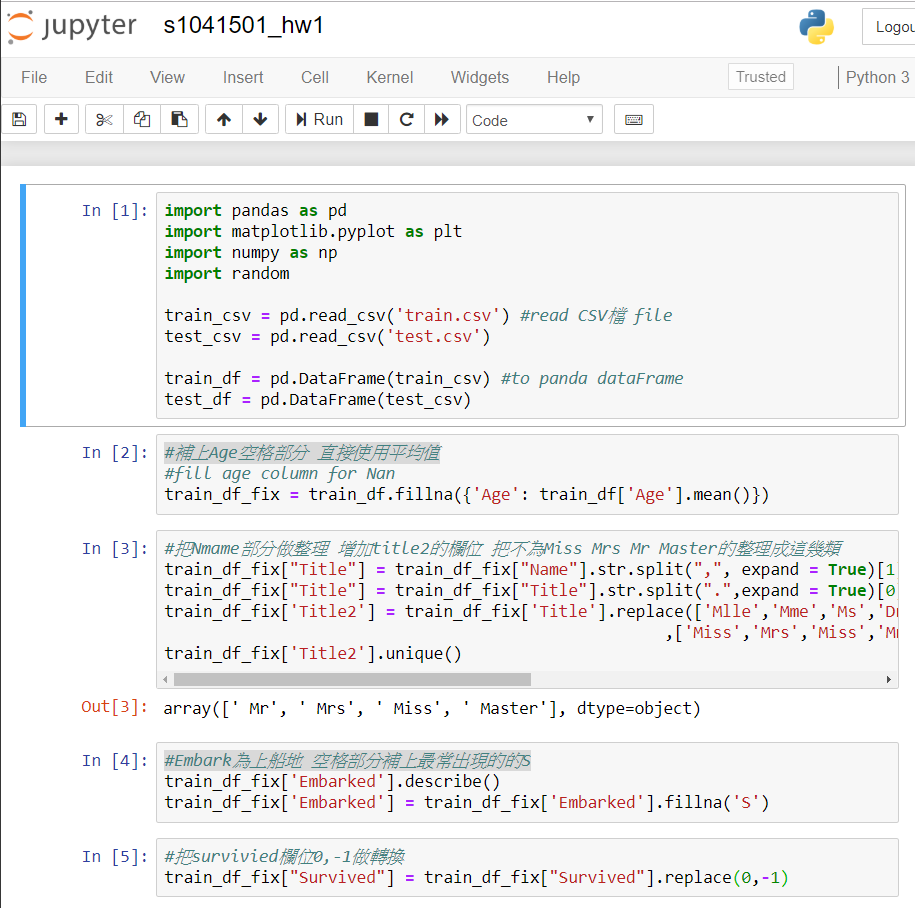
可以使用Jypyter看code與解釋

使用到的套件有 numpy, pandas, matplotlib.pyplot(觀察數據圖表才用到)

#read CSV檔 file #to panda dataFrame

#補上Age空格部分 直接使用平均值

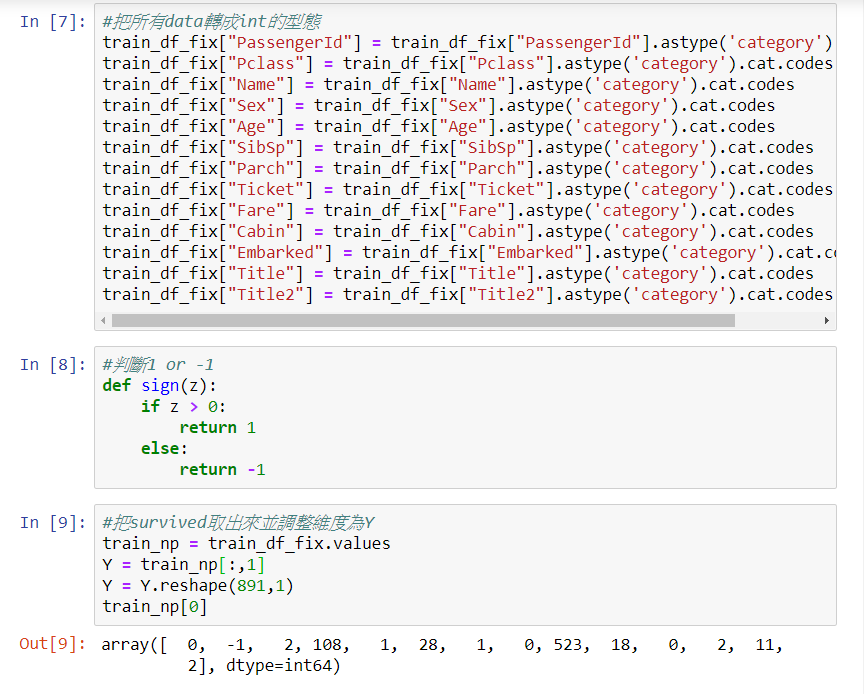
#把Name部分做整理 增加title2的欄位 把不為Miss Mrs Mr Master的整理成這幾類



#Embark為上船地 空格部分補上最常出現的的S

#把survivied欄位0,-1做轉換

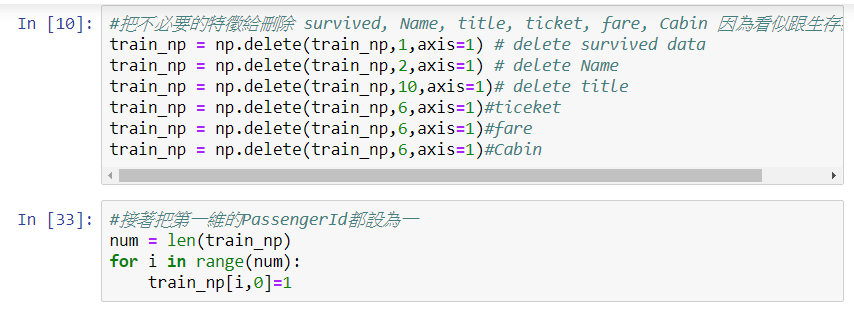
#把所有data轉成int的型態



#把survived取出來並調整維度為Y

#把不必要的特徵給刪除 survived, Name, title, ticket, fare, Cabin 後三者因為看似跟生存結果無關,而Name,survived,title都用不到

#接著把第一維的PassengerId都設為一





max\_update = 25000 #執行次數 #測試後25000就找得差不多了

alpha = 0.05 #權重調整成長係數 #測試後感覺上0.1以下調整的都差不多

wt = np.random.rand(8) #使用為隨機向量

###Pocket\_PLA 基本就是因為大多數的資料都不是線性可分 使用greedy的方式與原本PLA方法一樣再暴力找出最佳權重的方式

#很多種用隨機的方式挑 這種方式感覺最快最準

#隨機挑一筆來看上次的Wt是否正確 若不正確就修改



#處理 Test.csv 與 train.csv方式相同

#補上Age空格部分 直接使用平均值

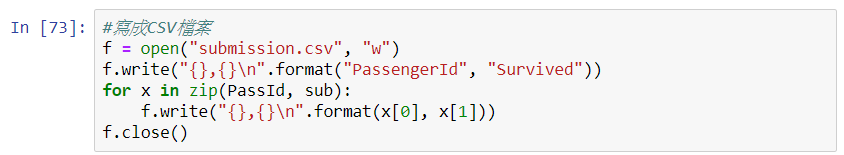
#PassengerId 第一維全設為1

#把不必要的特徵刪除



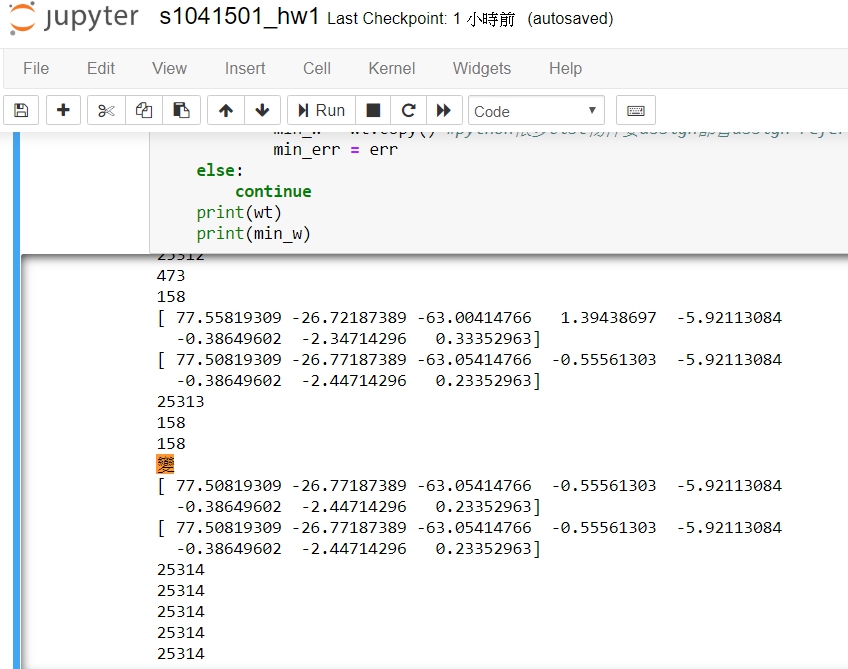
#計算出Test sequence

#寫成CSV檔案

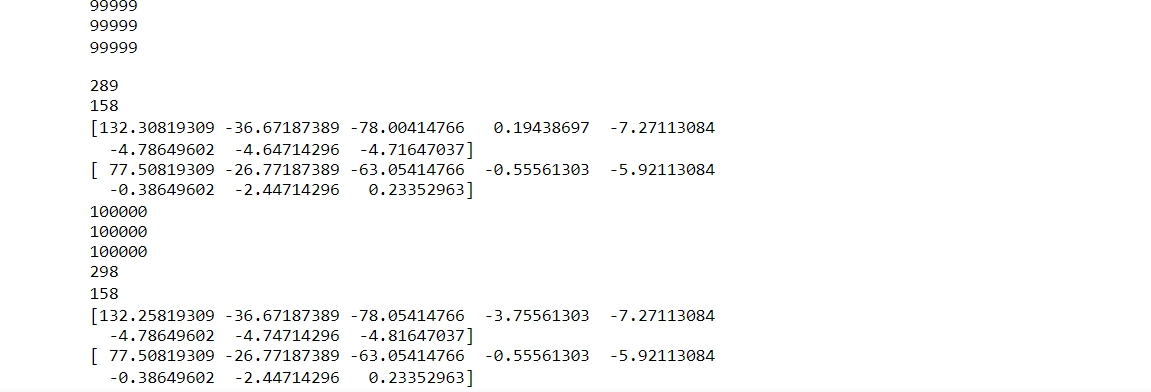


以下為執行圖

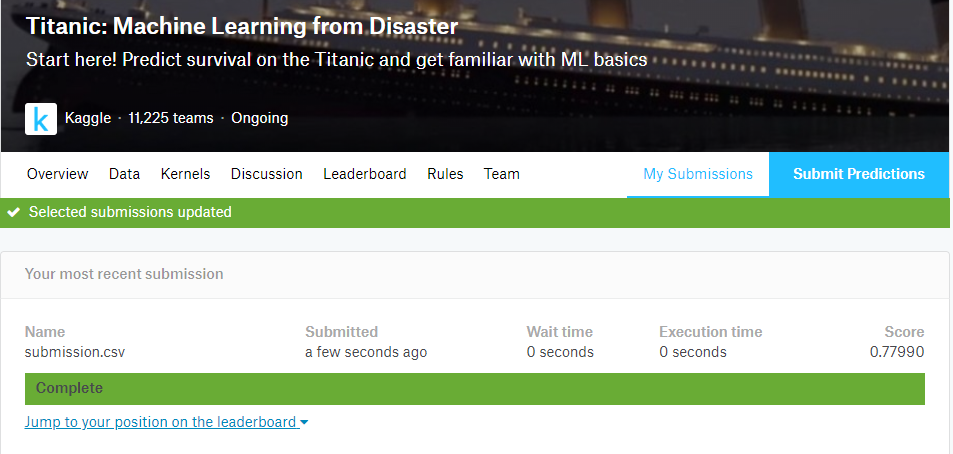
一開始我執行100000次來找權重（update=10000）



“變” 的時候，代表找到錯誤更少的向量權重，這在執行第25313次的時候發現最小只錯158次的權重向量



最終在執行10000次的時候也沒有找到更小的，最後update設為250000



最後丟給kaggle上評分的結果為0.77分