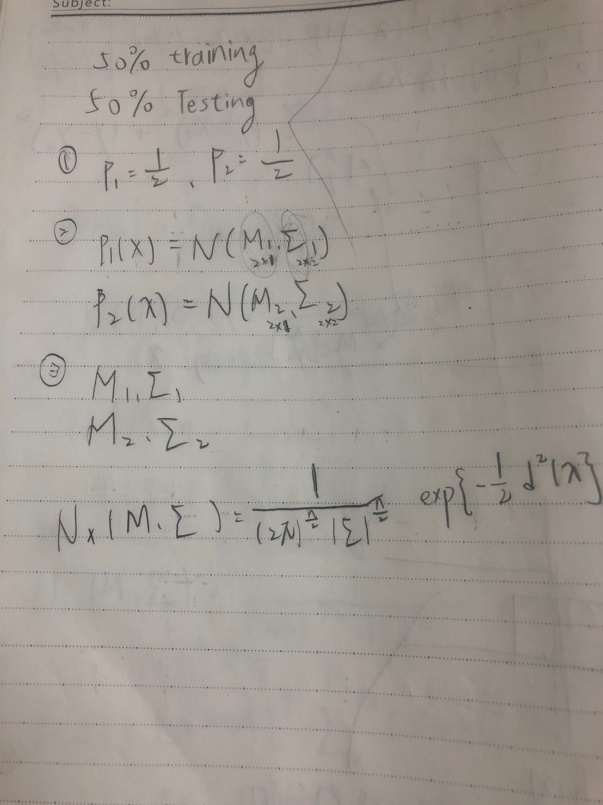
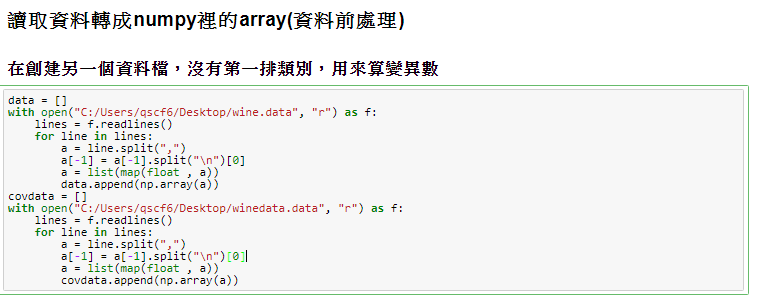
7108056022 陳冠霖

作法:



第一步:

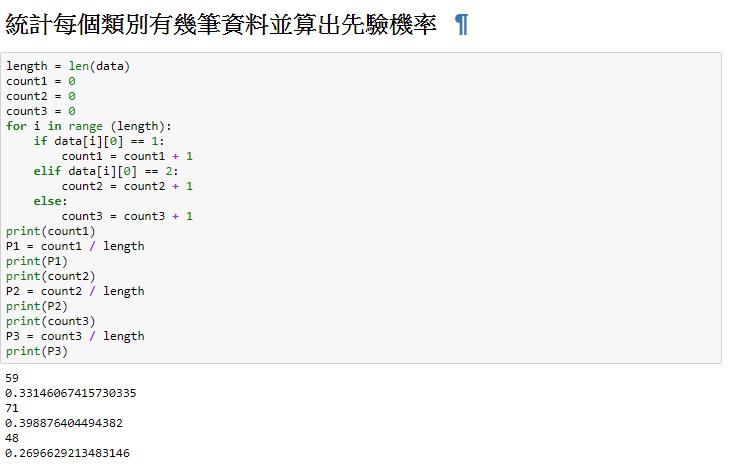
讀取資料檔，因為資料檔裡面的資料都是連再一起的，所以要使用檔案裡面的資料時，需要進行資料前處理，分割資料檔案裡面的資料，並將檔案中文字格式轉為數字格式，並將資料用numpy array的格式儲存下來，因為numpy中的function皆是以numpy array的格式輸入資料，另外創建一個資料集是沒有第一排類別，用來方便我們算變異數。



第二步:

算出每個類別在全部的資料中佔有多少個，得到每個類別的

先驗機率。



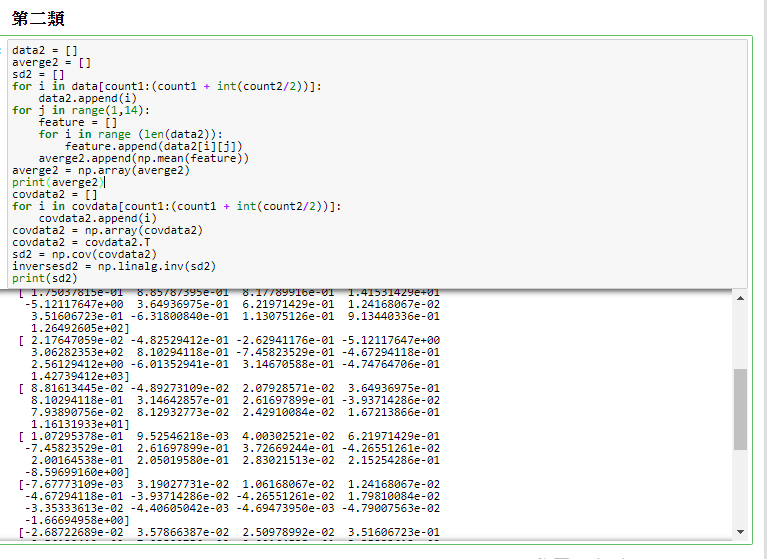
第三步:

將資料取一部分出來作為分類器的訓練資料，如果直接取資料檔的一半資料作為訓練資料的話，會出現問題，因為資料檔是以每筆資料的類別作為排序，所以不能直接取一半，而是每個類別的資料取一半作為訓練資料，每個類別剩下的另一半作為測試資料，來檢驗分類器是否準確。

將資料檔取一半作為訓練資料，資料裡總共有13個特徵，依照每個類別，算出資料的平均數為13乘1的矩陣。

將另一個創建的資料取一半作為訓練資料，資料裡總共有13個特徵，經過轉置矩陣，依照每個類別，變異數為13乘13的矩陣。

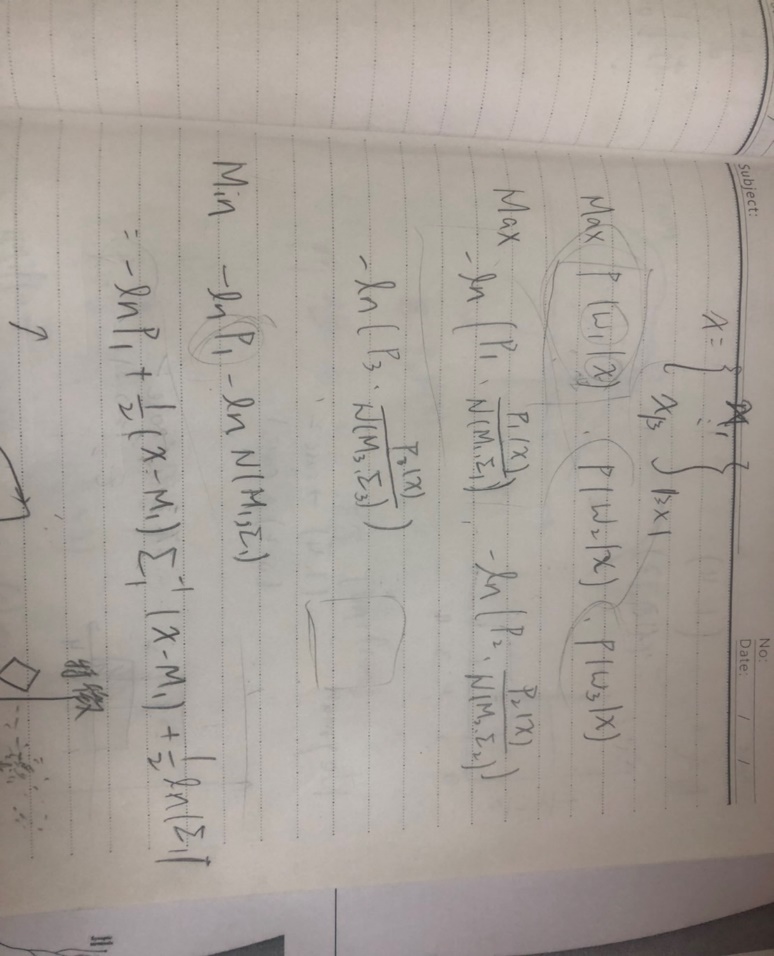




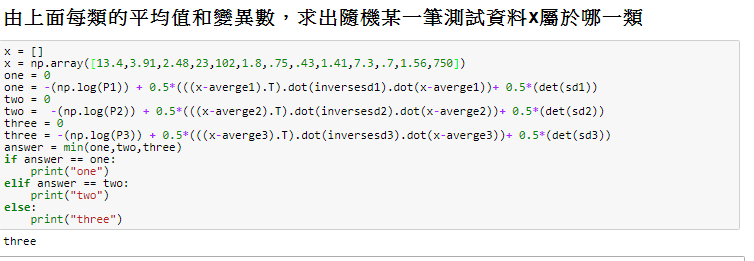


第四步:

利用上面每個類別的特徵的平均數和變異數帶入區分函式discrimination function，用來分辨讀取進來的測試資料屬於哪一個類別(取最小值為類別)。



取第三類的其中一筆測試資料13.4,3.91,2.48,23,102,1.8,.75,.43,1.41,7.3,.7,1.56,750



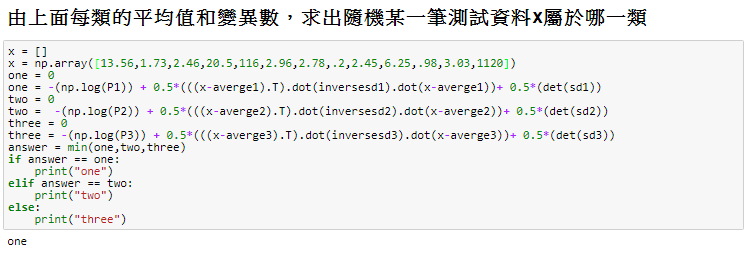
取第二類的其中一筆測試資料

12.04,4.3,2.38,22,80,2.1,1.75,.42,1.35,2.6,.79,2.57,580



取第一類的其中一筆測試資料

13.56,1.73,2.46,20.5,116,2.96,2.78,.2,2.45,6.25,.98,3.03,1120



結果:

將所有的測試資料，進行區分找出各自屬於哪個類別，算出準確率，使用分到正確的類別的個數除以全部有幾筆測試資料。

