ML HW2

0756526 林俊翰

Question1:

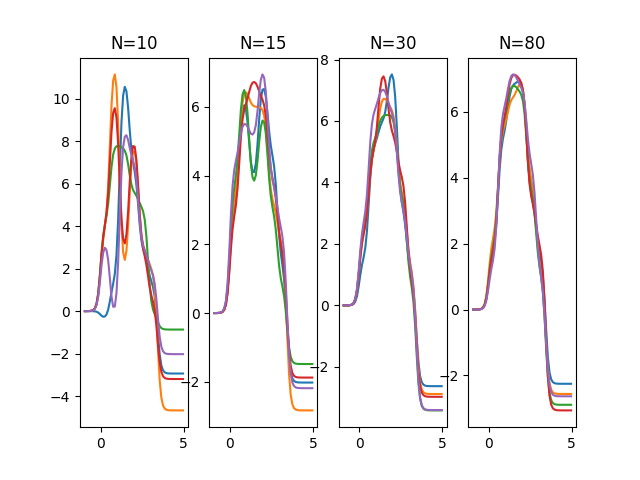
Part1: 先算mean vector & covariance matrix (SN)

先算SN 用課本3.51式可得

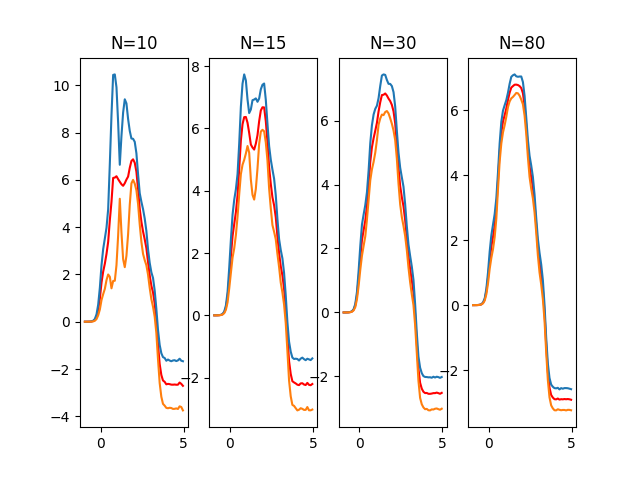
後算Mean vector 用課本3.50式

Part2: samples form parameter posterior distribution

用python numpy套件，只要輸入mN跟SN，就可隨機sample normal distribution



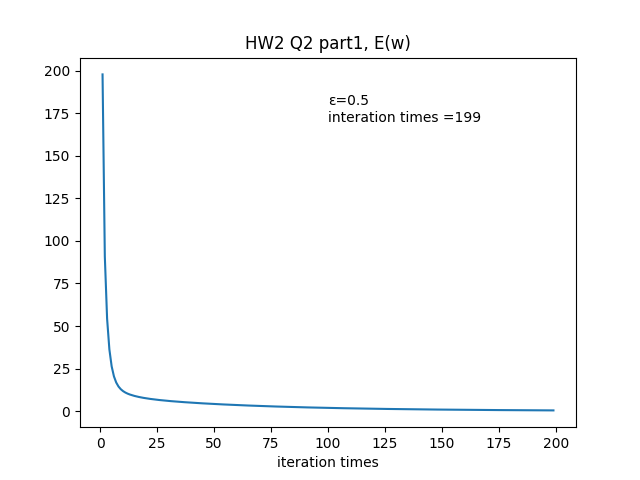
part3 : mean curve and the region of variance with one standard deciation 藍線是+1標準差、橘線是-1標準差、紅線是mean curve   
mean curve 我是對固定的X軸的點進行sample取平均值＆標準差，後將mean point 連成一線即可得．然後正負一個標準差的兩條curve也可得．



Question 2

Part 1 : Learning curve of E(W)

更新W的方式，我使用課本4.99式，原因是都是matrix的運算的話，numpy 可以算得比較快．  
我設定的 /epsilon 是0.5 ，需要計算199次之後，E(W)會小於 /epsilon



Part2 : classification result of test data

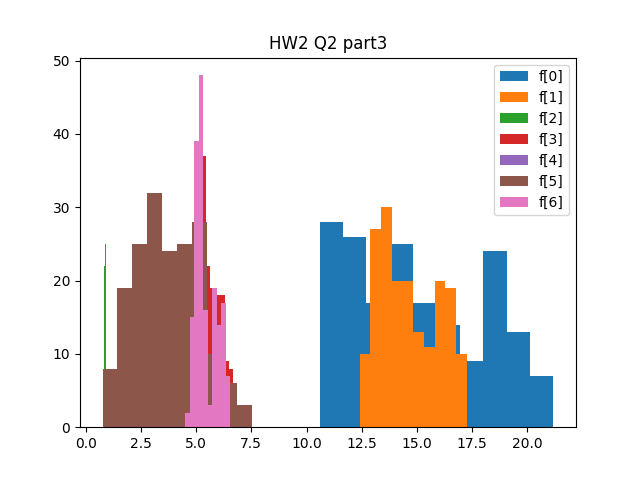
用最後的Ｗ 得到的分類結果



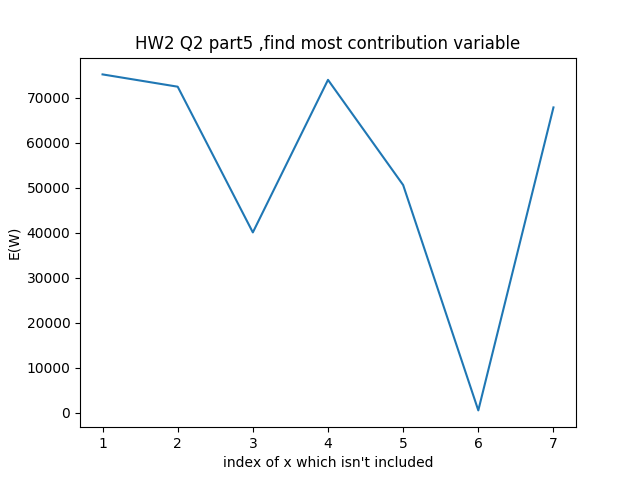
Part 3 distribution of the variable

X軸是各feature 的值，用不同顏色代表不同feature

Y軸是不同值出現次數的累積



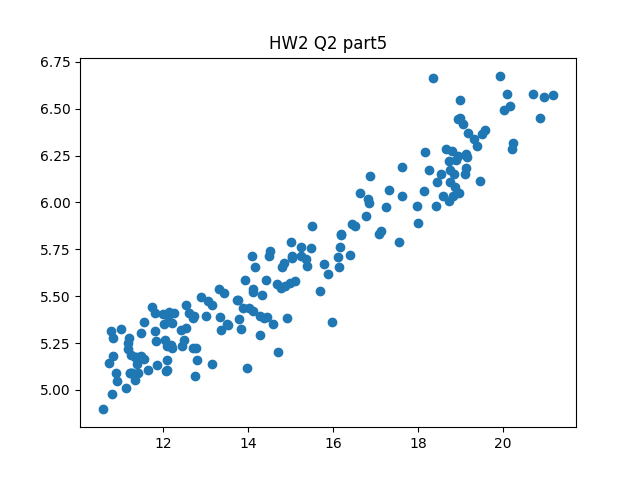
part5 : a pair of the most variables



由圖可知 少了第1 跟4 個variable後，

Error 上升最多

所以可知這兩個variable是most contribution variable

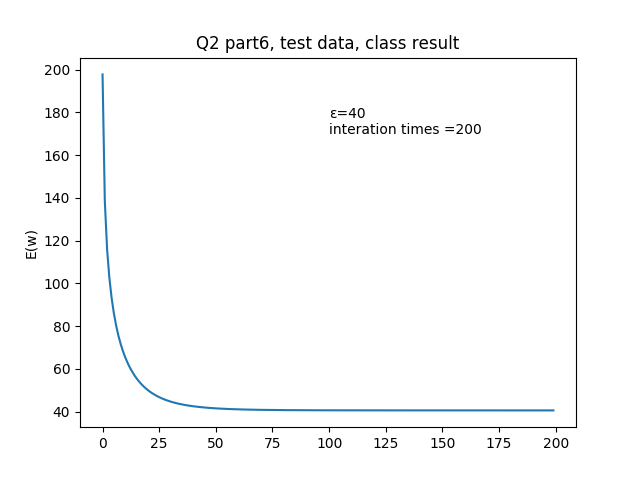


X軸是variable 1

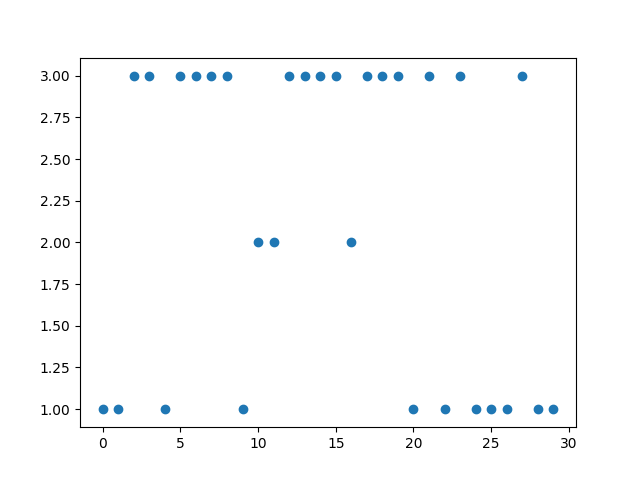
Y軸是variable 4

part6

E(w) curve 由兩個most contribution variable重新得到



由兩個variable train出來的W所分類的結果

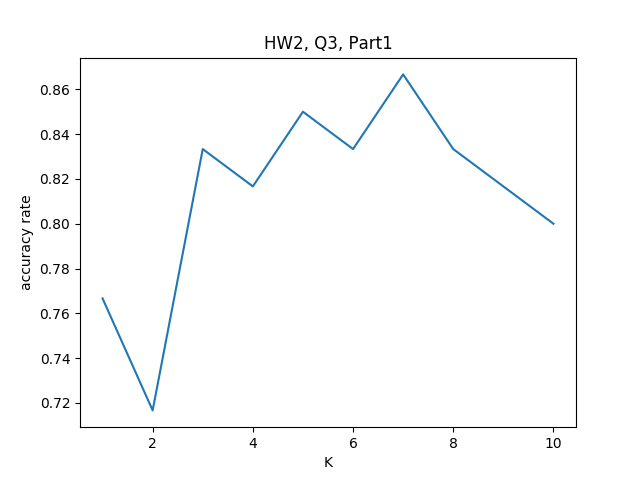


Question 3

Part1:

X軸是K值，Ｋ表示要取幾個Euclidean distance最接近的點來投票

Ｋ從1~10



Part2 :

X軸是V, V表示多維空間中的距離，即Euclidean distance．

V從2~10

