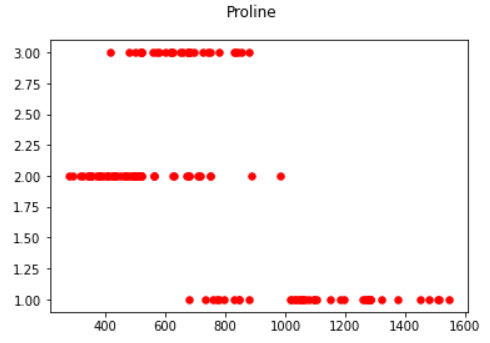
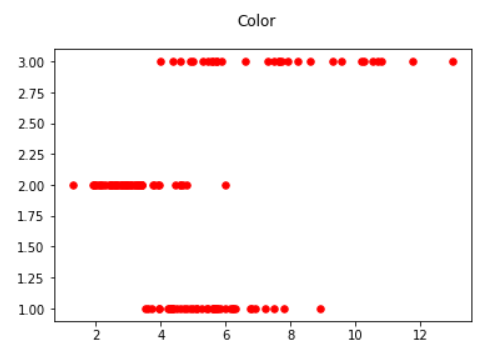
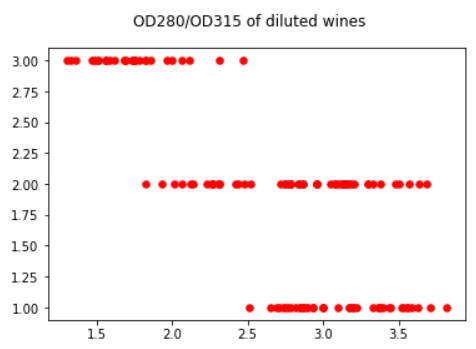
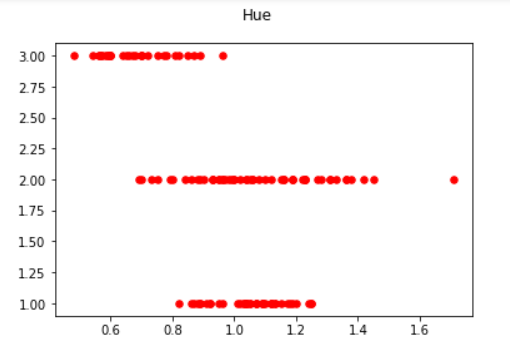
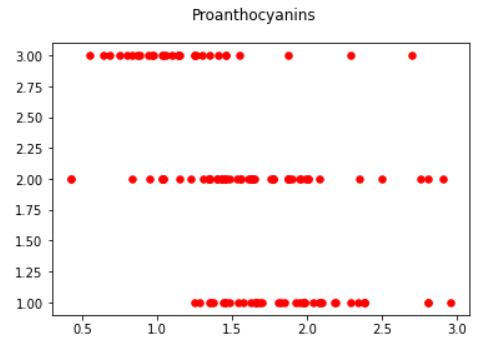
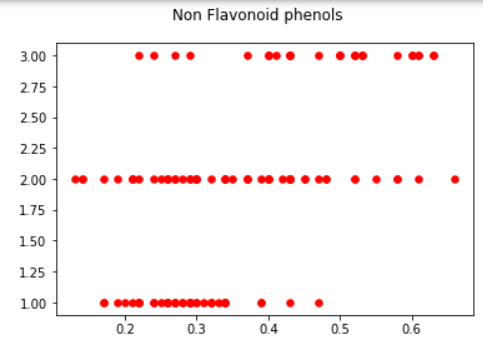
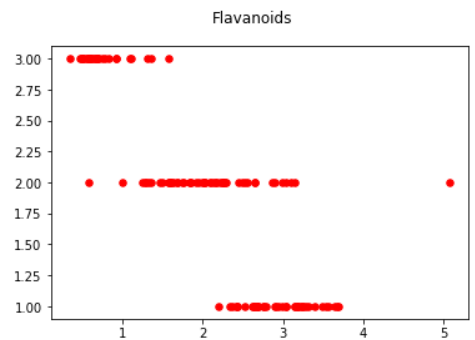
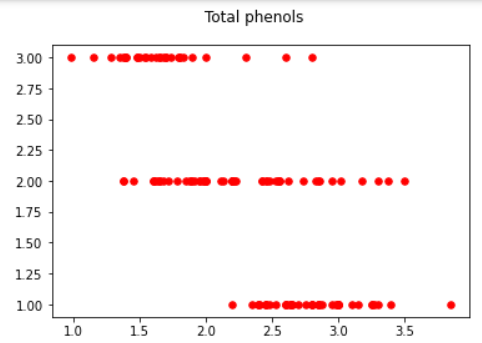
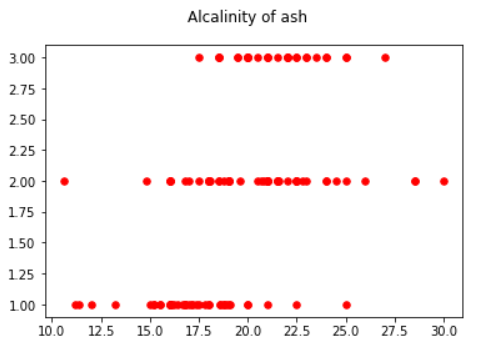
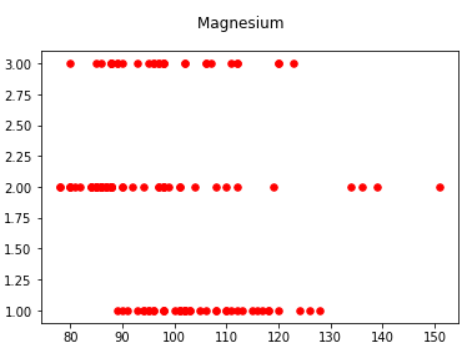
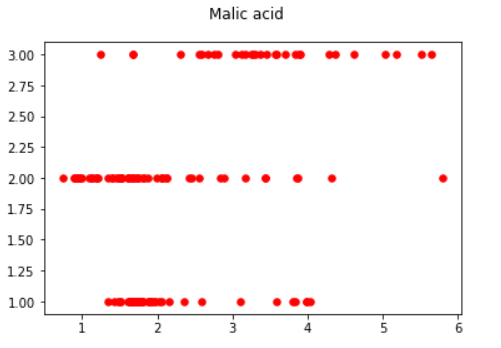
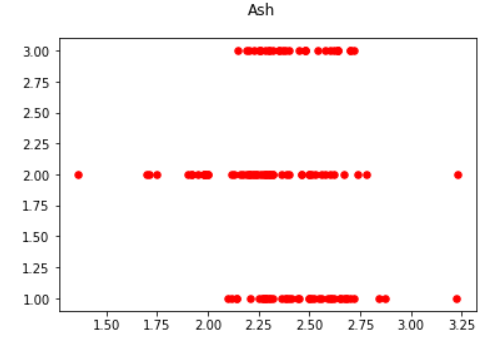
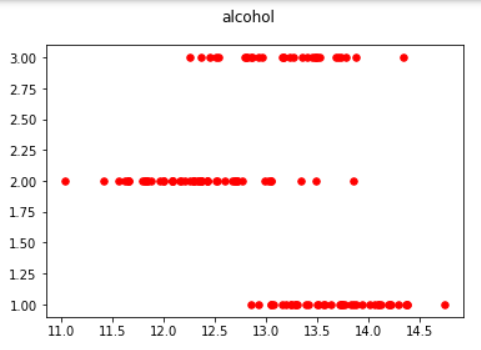
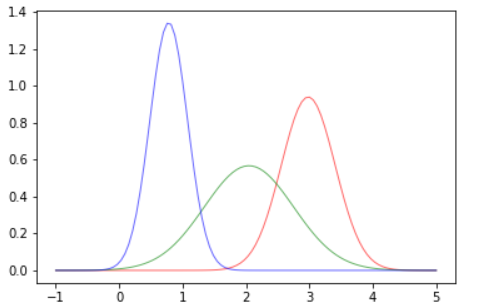
Characteristics

下圖是13個特徵當x軸對應該分類當y軸的二維平面圖，在某些特徵，像是 Flavanoids或是Total phenols下可以畫出兩條垂直線將這三類分開，但在Ash或是Magnesium下就無法明確地畫出兩條直線將三類分開，大部分的特徵在一三類有比較類似的數值，而第二類則跟一三類有較大的差異。



實作方法

根據MAP的公式可以拆分成likelihood乘prior並且除訓練資料出現的機率。所以我設定三個MAP分別計算此13個特徵對應到第一類的機率、第二類的機率和第三類的機率，由於訓練資料出現機率在三個MAP裡面都是一樣的數值所以忽略不計算。prior的計算方法為整個資料去除測試資料後三個標籤出現的機率。likelihood的計算方法則是根據題目敘述，所有資料的都是高斯分布並且獨立事件，因此可以分別計算每一個特徵的出現機率在將其相乘起來算總likelihood，而我們要取likelihood的最大值，根據likelihood取log後對平均值和標準差微分，我們可以得出將該高斯分布的平均值和標準差應設定為training data的平均值和標準差，即可得到最大likelihood，而該平均值和標準差可利用numpy函示庫計算，取得的數值可以利用scipy.stats.norm.pdf計算高斯機率密度函數來取得該Characteristic的出現機率，下圖為三個標籤對應紅綠藍線在Flavanoids的高斯機率密度分布。最後只要將所有特徵的likelihood相乘起來在乘prior然後取數值最大的MAP當作預測結果就完成了。

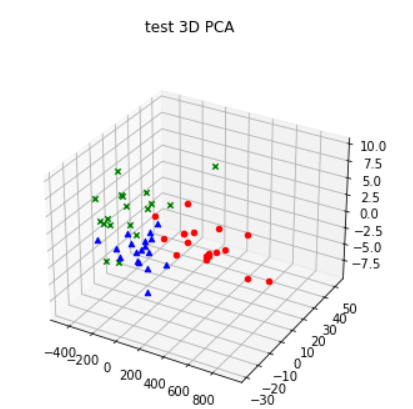
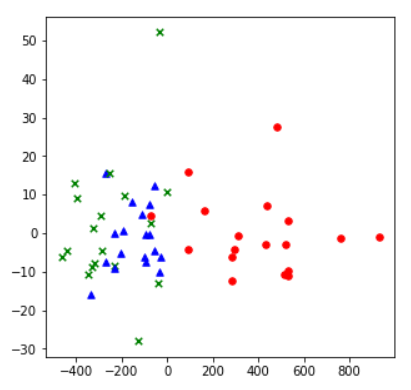


MAP和MLE

在作業裡面如果將資料分為54筆的測試和124筆測試資料，可以做到測試資料100%的準確度，訓練資料98~99%的準確度。為了測試MAP是否真的優於MLE，所以做了以下實驗，如果測試資料和訓練資料都是隨機切分出來的，那MAP的正確率會優於只做MLE，但在本次作業裡因為測試資料比例和訓練資料是不一樣的加上正確率極高並無法有效觀察出來，但假使將訓練資料和測試資料比例調整到34:144，會因為prior的準確度跟實際差異過大導致MAP準確度只有66%但是MLE仍然可以到達98%。

PCA

下圖分別是將訓練資料用PCA後產生的一維和二維向量，對應分類標籤所產生的圖。第一類是紅色，第二類是綠色，第三類則是藍色，可以再二維圖片上區分第一類和另外兩類，而到三維上也可以稍微區分第二和第三類，因此如果在更高維的空間裡應該可以更明確的分類出來。

而下面兩張圖則是訓練資料利用PCA繪製出來的二維和三維圖像，可以發現分布和測試資料類似。

