1.

我讀取'wine.csv'數據集，並用 for 迴圈去判斷'target'是 0、1、2，然後每個種類隨機取 20 個數據併為'test.csv'，然後剩下的數據併為'train.csv'。

2.

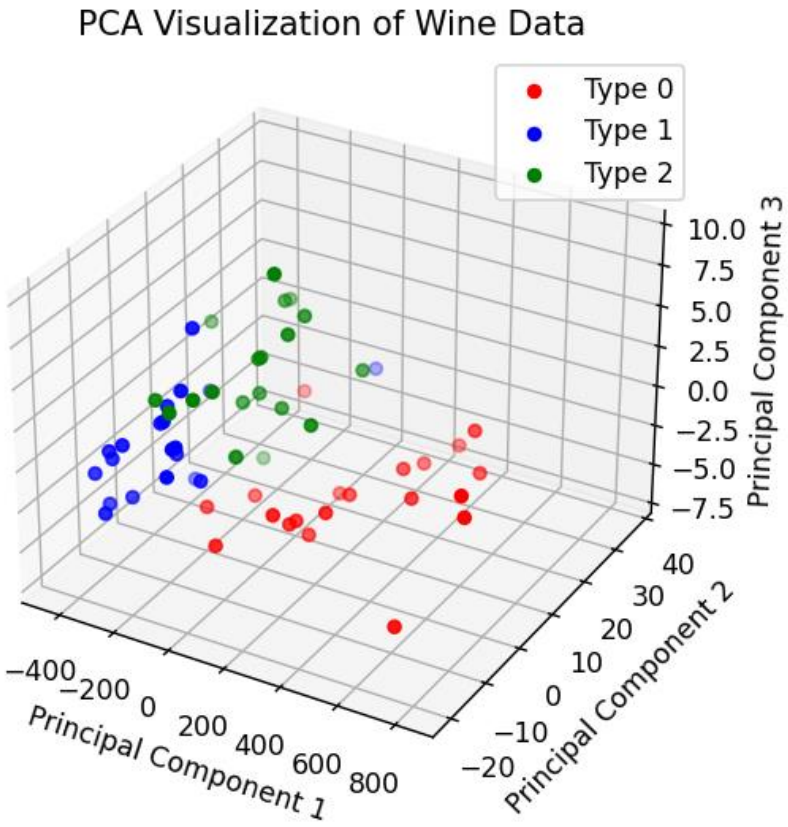
為了計算測試數據集中每個實例的後驗機率並預測其標籤，我使用變數 num_correct 初始為 0，來計算正確的預測數量。
使用 iterrow()逐個讀取測試數據及中的每個實例，並將其真實標籤儲存仔變量 true_label 中。接下來對每個可能的標籤值進行計算機率。對於每個標籤值 j，從先驗機率中取的 j 的機率值，並對每個特徵 k，對應的高斯分布中獲取該實例的可能性，然後取其對數相加。所有的標籤值的機率值都儲存在 probs 數組裡。最後根據計算得到的機率值，選擇最有可能的標籤並預測其值，如果預測的標籤和真實標籤相同，則 num_correct 變量加 1，代表預測正確。最後根據預測正確的數量以及測試數據集的總數量相除來計算精度值。

```
Accuracy rate: 0.9833333333333333
```

3.

describe the role of PCA:
主成分分析(PCA)是一種統計技術，用於減少數據集中的變量數量，同時保留盡可能多的信息，它有助於通過辨識變量之間的模式和相關性，將複雜的數據簡化成新的一組變量，這些變量捕獲數劇中最重要的信息。它也是常用的降維技術，找到數據中最大的變化方向並將數據投影到這些方向上，由此產生的主成分彼此不相關，意味著他們包含關於數據的獨立信息。通選擇捕獲數據變化的主成分，PCA 可以將高維數據降為低維數據，以方便可視化。

Code:
首先讀取' test_csv' 檔案，再來進行 PCA 分析，我設置的主成分是 3，並將 PCA 應用在我的測試數據上，並將其儲存為變數 transformed 中，接者我用散布圖 scatter plot 將分析結果進行呈現 3 維圖。
我定義了 3 種不同的顏色(colors)和標籤(labels)，我用 for 迴圈將每個目標值(target)分別用不同顏色進行繪製，然後我用 np.where 找尋每個目標值的索引，然後使用這些索引進行繪圖，最後添加圖例(legend)和標題(title)。



4.

在 Bayesian inference 中，先驗分布代表我們在看到數據之前對參數的機率分布看法，當我們觀察數據之後更新機率分布，就是後驗機率。先驗分布在後驗機率上可以產生重大影響，尤其當我們只有有限的數據或是數據存在噪聲或不明確性。

對於葡萄酒數據集，我們可以使用先驗分布來表示對於每種葡萄酒類型的每個特徵分布的信息。例如，我們可能假設酒精含量的平均值和變異數對於每個葡萄酒類型都不同，因此我們使用先驗分布表示這些差異。然後通過考慮特徵分布的先驗知識，我們可以獲得更準確的後驗機率，做出更精確的預測決定。

另外先驗分布的選擇對後驗機率影響也很大，如果先驗分布過強或過弱，會導致我們的估計，得到不準確的預測，此外先驗分布過窄也可能導致過度擬合和對新數據的預測能力不佳。