Homework I

- **Task**

  The provided dataset contains three types of wine, each represented by 13 distinct features. The objective is to implement a Maximum A Posteriori (MAP) classifier to classify the given wine samples based on these features.

- **Solution**

  The whole progress is in the following steps:

  - **Split the dataset:** The dataset is divided into a training set and a testing set.
  - **Compute the prior distribution:** The prior probability for each class x is estimated from the training set as:

    $$P(x) = \frac{\text{number of samples with type } x}{\text{total number of samples in the training set}}, \quad x \in \{0, 1, 2\}.$$

  - **Estimate the likelihood using a Gaussian distribution:** Assuming that all features are independent and follow a Gaussian distribution, we estimate the mean and variance for each feature and class:

    $$\mu_x^{(j)} = \frac{1}{N_x} \sum_{i=1}^{N_x} X_i^{(j)},$$

    $$\sigma_x^{(j)2} = \frac{1}{N_x} \sum_{i=1}^{N_x} \left(X_i^{(j)} - \mu_x^{(j)} \right)^2,$$

    where:

    - $X_i^{(j)}$ is the value of the j-th feature for the i-th sample in class x.
    - N_x is the number of samples in class x.
    - $\mu_x^{(j)}$ and $\sigma_x^{(j)2}$ are the estimated mean and variance of the j-th feature for class x.

  - **Compute the posterior probability:** Using Bayes' theorem, the posterior probability for a sample belonging to class x is given by:

    $$P(x \mid X) \propto P(X \mid x) P(x),$$

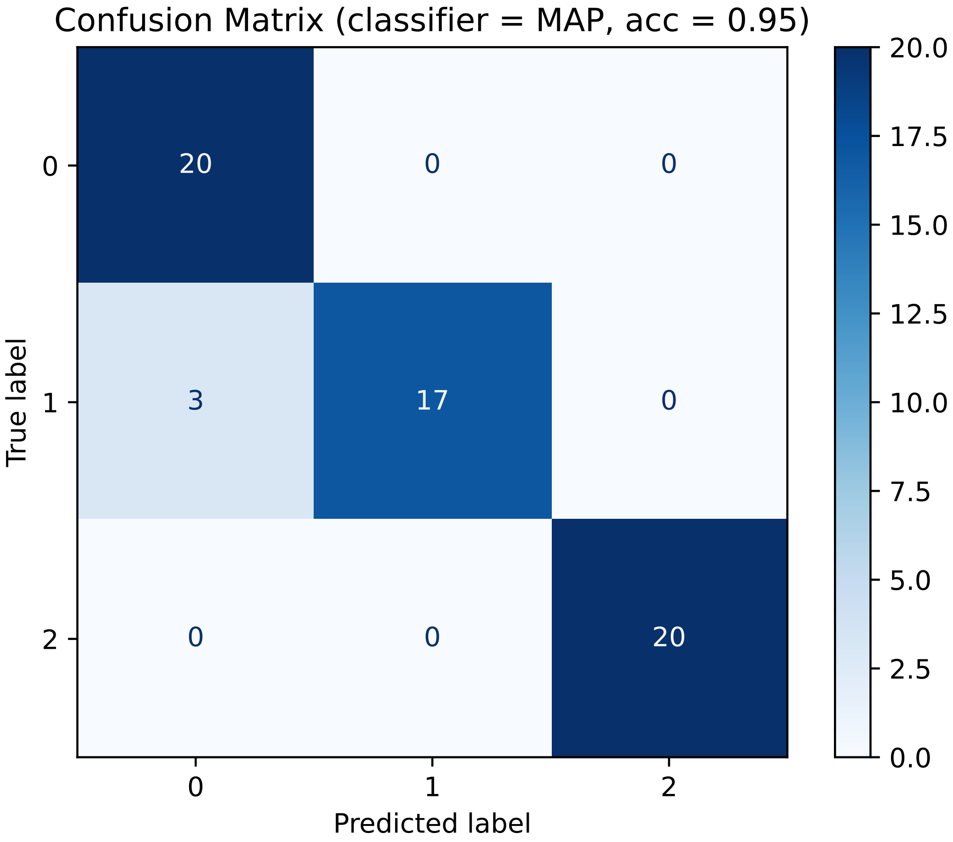
    where the likelihood term is calculated as:

    $$P(X \mid x) = \prod_{j=1}^{13} \frac{1}{\sqrt{2\pi\sigma_x^{(j)2}}} \exp\left(-\frac{(X^{(j)} - \mu_x^{(j)})^2}{2\sigma_x^{(j)2}} \right).$$

- **Classification decision:** A given test sample is assigned to the class with the highest posterior probability:

$$\hat{x} = \arg\max_{x \in \{0,1,2\}} P(x \mid X).$$

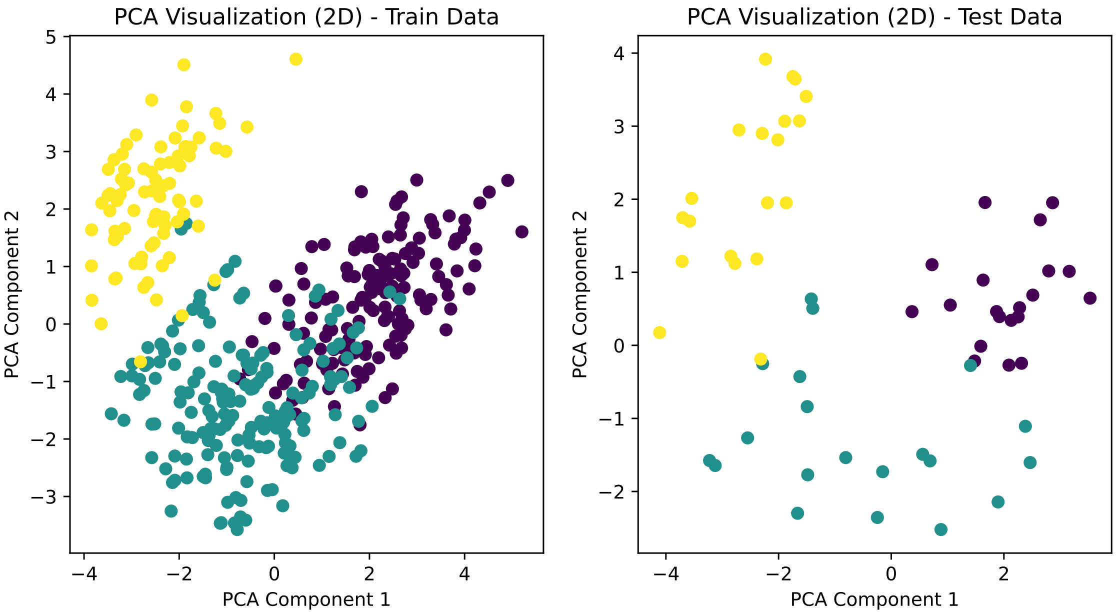
- **Performance**
  The performance of the MAP classifier is shown below. The accuracy is 95% on the testing set.



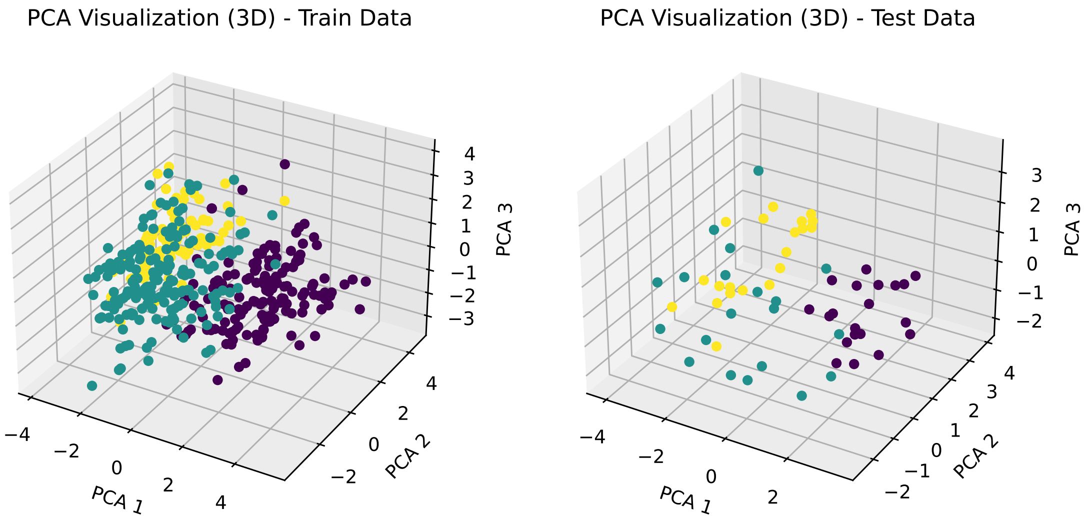
Confusion matrix of the MAP classifier.

- **Discussion**
  Below are some key aspects we discussed regarding the dataset and the performance:

  - **Data Visualization:** To better understand the distribution of different wine types, we visualize the dataset in both 2D and 3D spaces. The following figures illustrate the feature distribution and potential class separability.

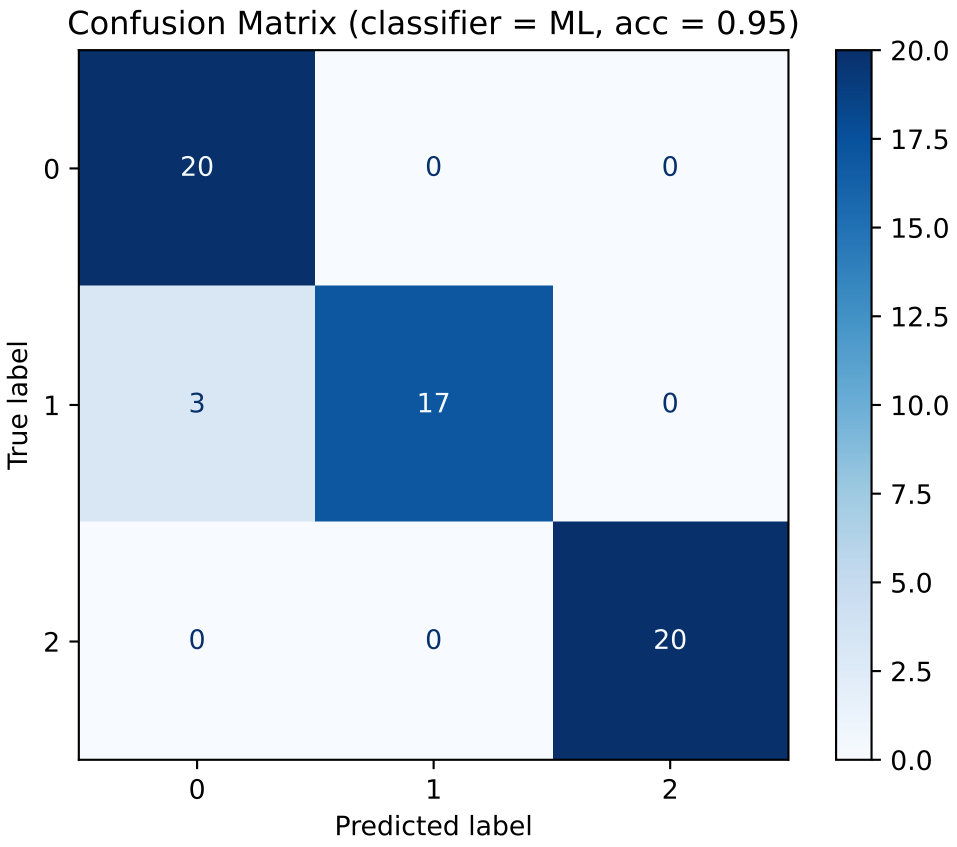


2D visualization of the dataset.



3D visualization of the dataset.

- **Effect of Prior Distribution:** To evaluate the influence of prior probabilities on classification performance, we implement a Maximum Likelihood (ML) classifier, which completely disregards prior information. The performance of the ML classifier is shown below:



Confusion matrix of the ML classifier.

Interestingly, the results are identical to those obtained with the MAP classifier. This is because, in this problem, the likelihood values are much smaller than the prior probabilities. Since the posterior probability is the product of the likelihood and prior, the likelihood becomes the dominant factor, making the effect of the prior negligible.

– **Contribution of Each Feature:** To analyze the importance of each feature, we train the MAP classifier using all features except one and measure the resulting accuracy. A larger drop in accuracy indicates that the removed feature is more important for classification. The results are summarized in Table 1. The accuracy of the original MAP classifier is 95%. It can be seen that when removing the first six features in Table 1, the accuracy decreases, indicating that these features contribute positively to the classification performance. On the other hand, removing the last few features leads to an increase in accuracy, suggesting that these features introduce noise or redundant information rather than aiding the classification.

Feature Removed	**Accuracy after remove the feature**
alcohol	93.3%
malic_acid	93.3%
ash	93.3%
flavanoids	93.3%
hue	93.3%
od280/od315_of_diluted_wines	93.3%
alcalinity_of_ash	95.0%
magnesium	95.0%
total_phenols	95.0%
nonflavanoid_phenols	95.0%
proanthocyanins	95.0%
proline	95.0%
color_intensity	96.7%

Table 1. Classification accuracy when removing each feature.