

Please use python3 to execute the code.

In this problem , since it is similar to the previous one with 2 features and two classes 0 and 1. Here I use previous code with little variation in extracting the data form the csv files. All the conditions of bayesian classification are used as they were used in the previous problem .

For the performance analysis , I use the confusion matrix which is defined as $M = \begin{bmatrix} \text{true 0} & \text{false 1} \\ \text{false 0} & \text{true 1} \end{bmatrix}$ and classification percentage.

Here this code was analysed for various cases as follows:

Case a: $\text{cov_0} = \text{cov_1} = \text{cov} = \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}$ where $\text{cov} = p_0 * \text{cov_0} + p_1 * \text{cov_1}$ and $a = 0.5 * (\text{cov}[0][1] + \text{cov}[1][1])$:

the confusion matrix: $\begin{bmatrix} 47 & 0 \\ 3 & 40 \end{bmatrix}$

the classification percentage of 0 and 1 resp. : 94.0 100.0

Case b : $\text{cov_0} = \text{cov_1} = \text{cov} = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$ where $\text{cov} = p_0 * \text{cov_0} + p_1 * \text{cov_1}$ and $a = \text{cov}[0][1]$ and $b = \text{cov}[1][1]$:

the confusion matrix: $\begin{bmatrix} 50 & 1 \\ 0 & 39 \end{bmatrix}$

the classification percentage of 0 and 1 resp. : 100.0 97.5

Case c-0: $\text{cov_0} = \text{cov_1} = \text{cov}$:

the confusion matrix: $\begin{bmatrix} 50 & 1 \\ 0 & 39 \end{bmatrix}$

the classification percentage of 0 and 1 resp. : 100.0 97.5

Case c-1: $\text{cov_1} = \text{cov_0}$:

the confusion matrix: $\begin{bmatrix} 50 & 1 \\ 0 & 39 \end{bmatrix}$

the classification percentage of 0 and 1 resp. : 100.0 97.5

Case c-2: $\text{cov_0} = \text{cov_1}$:

the confusion matrix: $\begin{bmatrix} 45 & 0 \\ 5 & 40 \end{bmatrix}$

the classification percentage of 0 and 1 resp. : 90.0 100.0

Case d: cov_0 and cov_1 are the true values calculating from the data :

the confusion matrix: $\begin{bmatrix} 50 & 0 \\ 0 & 40 \end{bmatrix}$

the classification percentage of 0 and 1 resp. : 100.0 100.0

From the above cases it is clear that using the true values of covariance matrices give the highest accuracy and it is 100%.

