Machine Learning for Image Processing

Programming Assignment 2

Name: Mohammad Al Fahim K **Roll No.:**EE17B021

1. **Brief Description:**

   To present brief details of the datasets provided

   Table:

Name of Dataset	Feature vector dimension	No. of classes	Prior Prob. of each class	Mean vector dimension	Covariance matrix dimension
Image segmentation dataset1	19 X 1	7	1/7 for all classes	19 X 1	19 X 19
Iris Dataset 3	4 X 1	3	1/3 for all classes	4 X 1	4 X 4
Letter Recognition dataset 4	26 X 1	26	variable	26 X 1	26 X 26
Solar Flare dataset 6	10 X 1	3	variable	30 X 1	30 X 30
Wisconsin prognostic breast cancer dataset	30 X 1	2	variable	30 X 1	30 X 30
Wisconsin diagnostic breast cancer dataset	32 X 1	22	variable	32 X 1	32 X 32

2. **Aim:**

   To plot 1-D and 2-D histograms on one of the datasets provided and apply Bayesian classification

   Short Theory:

   If $P(X/W_1)$ and $P(X/W_2)$ are the likelihoods of the classes W_1 and W_2 respectively, the decision boundary obtained from,

$$P(X/W_1) = P(X/W_2)$$

$$P\left(\frac{X}{W}\right) = \left[\frac{1}{(2\pi)^{0.5}}\right] * e^{-0.5*\frac{(x-\mu)^2}{(\sigma)^{\wedge}2}}$$

   The decision boundary is where the probability curves of two classes meet.

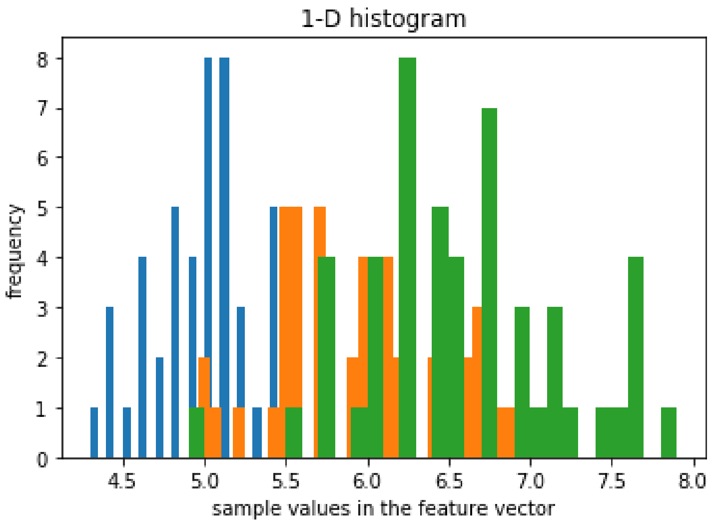
   I have worked on the iris dataset

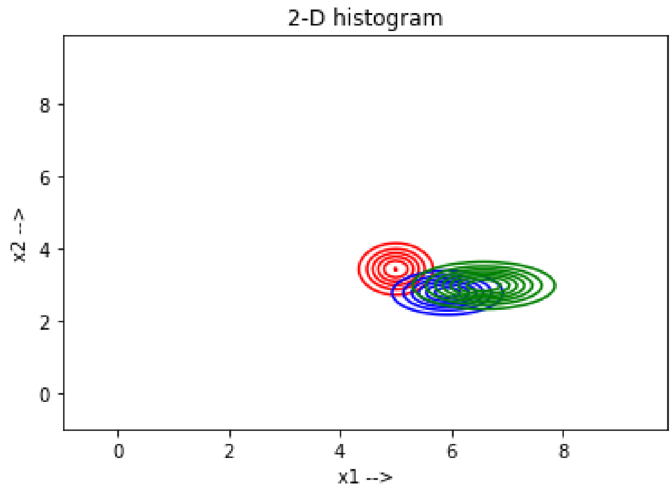
   Procedure:

   → After importing the dataset, split the samples into each class

   → Define one_d and two_d functions, with parameters stating which feature to use for the plots, to plot the 1-D and 2-D histograms

   → Input the feature index you wish to use to plot the histograms. By default, for the one_d, 0^{th} feature is used and for the two_d, 0^{th} and 1^{st} features are used.





Interferences:

The edges where the plots belonging to two different classes, denoted by different colours, meet are the decision boundaries

3. **Aim:**

To perform Bayesian classification on the dataset, que3.xlxs, for the given conditions

i) Same covariance matrices for all the classes

ii) Different covariance matrices

iii) Diagonal covariance matrices

Short Description:

If we have three classes, W_1, W_2, and W_3 respectively, a test case X will belong to that class which has the highest value among $P(X/W_1)$, $P(X/W_2)$ and $P(X/W_3)$. The Bayesian classification done here is only using the likelihood functions and I have not used the prior probabilities.

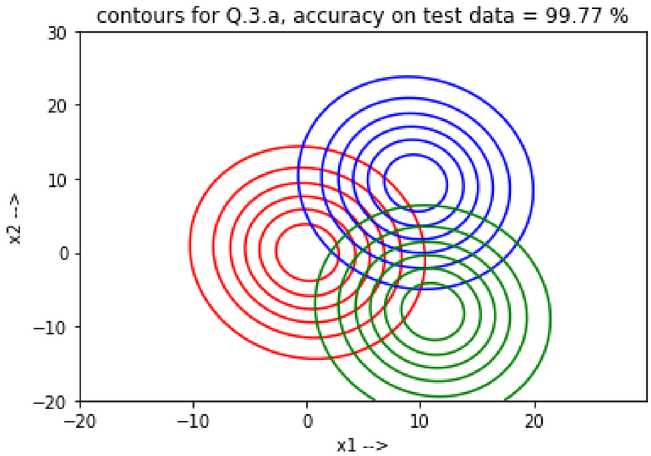
Procedure:

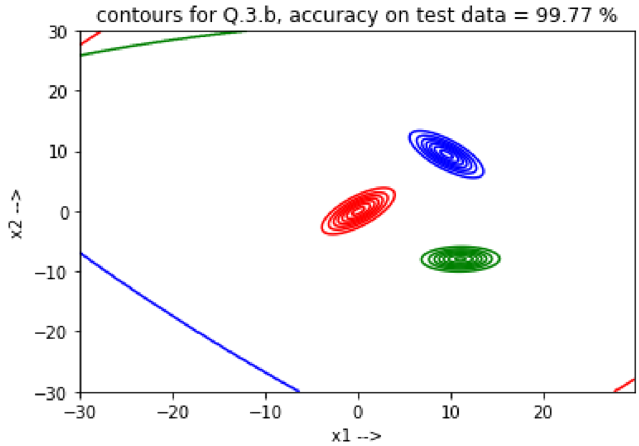
→ Import the dataset and split them into three classes and in each class, split it into 70% and 30% for training and test data respectively.

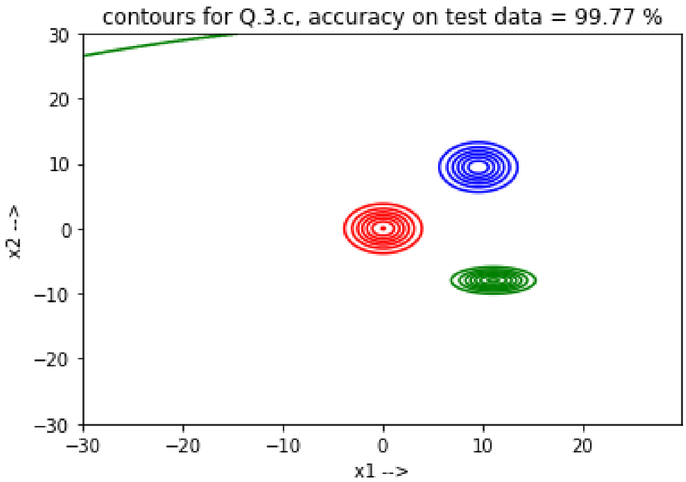
→ Define a custom function, gauss, to calculate probability and assign respective class

→ Calculate mean and sigma matrices for each case, calculate accuracy on test data using the cusom gauss function and plot their respective contours individually for each case

Plots:







Inferences:

In the 1st case, when all the classes have the same sigma matrices, the decision boundary is a perpendicular bisector between the mean points of every class.