Indian Institute of Technology Tirupati

Image Processing Lab

Lab sheet. No: 06

Name: Pedasingu Rajesh **Roll. No:** EE21M019

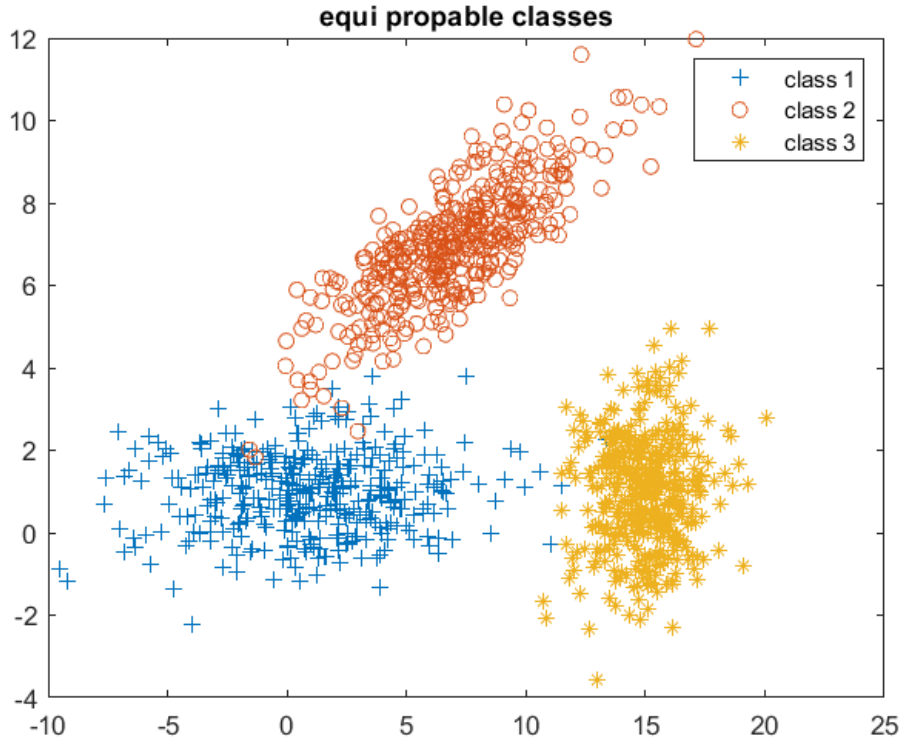
1. (a) Generate and plot a dataset of N = 1200 two-dimensional vectors that stem from three equiprobable classes modelled by normal distributions with mean vectors $m_1 = [1, 1]^T$, $m_2 = [7, 7]^T$, $m_3 = [15, 1]^T$ and covariance matrices $S_1 = \begin{bmatrix} 12 & 0 \\ 0 & 1 \end{bmatrix}$, $S_2 = \begin{bmatrix} 8 & 3 \\ 3 & 2 \end{bmatrix}$ and $S_3 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$.

   (b) Repeat (a) when the *a priori* probabilities of the classes are given by the vector $P = [0.6, 0.3, 0.1]^T$. Show the samples of each class in different color for a better visualization.

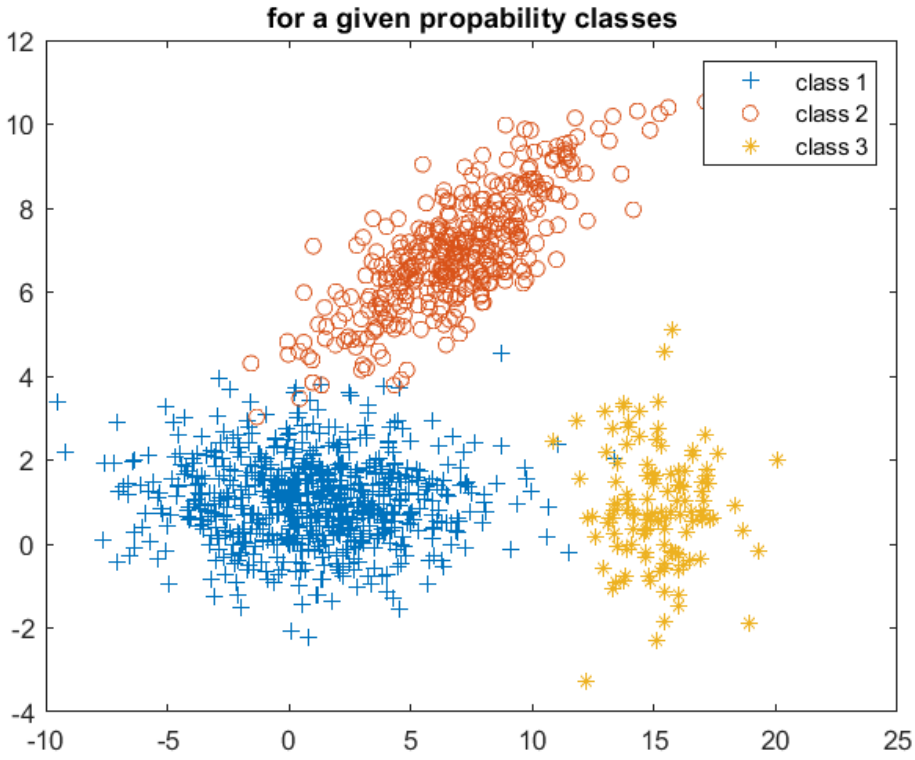
Aim: Generate and plot a dataset of N = 1200 two-dimensional vectors that stem from three equiprobable classes modelled by normal distributions with given mean and covariance matrices.
b) Repeat (a) when the a priori probabilities of the classes are given by the vector P = [0.6, 0.3, 0.1] `.

Output:

For equal probability case:



For given probability case:



Inferences:

1) Given equip ropable and there are 3 classes, so prior probabilities are 1/3 for all classes.
2) So, in 1200 two-dimensional vectors, each class having 400, 2 dimensional vectors.
3) Bu using the hold command we can plot the all the scatter points of each point in one plot its self.
4) In case of different probabilities each class get the number of two-dimensional vectors according to the ration of given probabilities.
5) So in this case, class 1 get 720 two dimensional vectors and class 2 contains 360 and class 3 contains 120 two dimensional feature vectors respectively.

2. Generate and plot a dataset of N = 1000 two-dimensional vectors that stem from three equiprobable classes modelled by normal distributions with mean vectors $m_1 = [1, 1]^T$, $m_2 = [14, 7]^T$, $m_3 = [16, 1]^T$ and covariance matrices $S_1 = S_2 = S_3 = \begin{bmatrix} 5 & 3 \\ 3 & 4 \end{bmatrix}$.

   Classify the test samples $X_1 = [5, 2]^T$, $X_2 = [17, 5]^T$ and $X_3 = [9, 2]^T$ based on

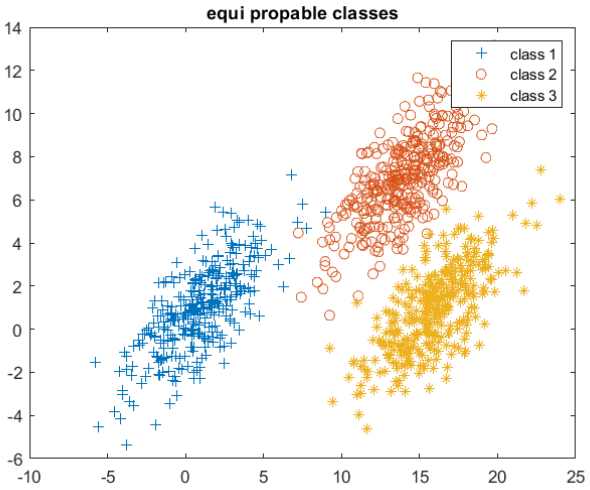
   (i) Bayesian classification

   (ii) Mahalanobis Distance

   (iii) Euclidean Distance and write the inferences.

Aim: Generate and plot a dataset of N = 1000 two-dimensional vectors that stem from three equiprobable classes modelled by normal distributions with given mean and covariance matrices. And classify then using Bayesian, mohalanbis distance and Euclidian distance for a given mean and co variance matrices.

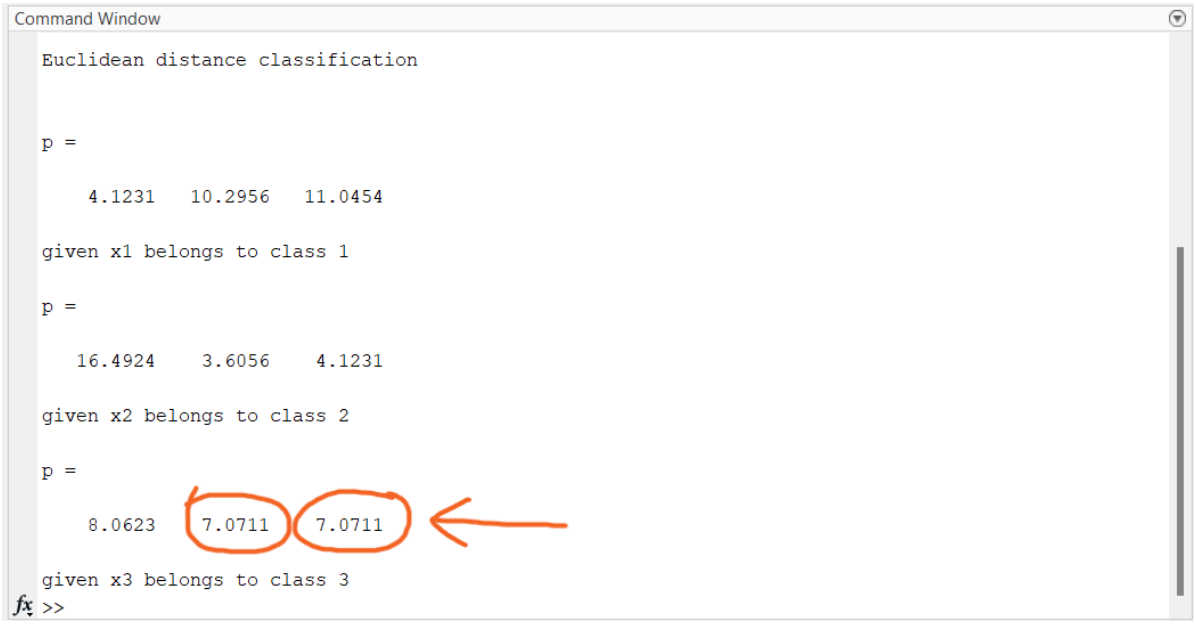
Output:




```
Command Window
  Basian classification

  given x1 belongs to class 1
  given x2 belongs to class 3
  given x3 belongs to class 2

  Mahalanobis distance classification

  given x1 belongs to class 1
  given x2 belongs to class 3
  given x3 belongs to class 2

  Euclidean distance classification

  given x1 belongs to class 1
  given x2 belongs to class 2
  given x3 belongs to class 3
fx >>
```

Inferences:

1) Given equip ropable and there are 3 classes, so prior probabilities are 1/3 for all classes.
2) So, in 1000 two-dimensional vectors, each class having 333, 2 dimensional vectors. The remaining one vectors we assign to any one of the class.
3) We classify the given x1,x2 and x3 By using given mean vector and covariance vector for 3 different Bayesian classification, mahalanobis distance classification and Euclidian distance classification. And the results are listed above.
4) For Euclidian distance classification theX3 input vector classified into both class 2 and class 3, so there is an ambiguity in Euclidian distance classification.



Q3. Given Iris Dataset. (3 Files – 'iristrain.xlsx', 'iristest.xlsx', 'readme.txt')

Aim: classify the given test iris data using Bayesian classification by using given training iris dataset.

Output:

The 3 classes are

Iris Setosa--1/3

Iris Versicolour –1/3

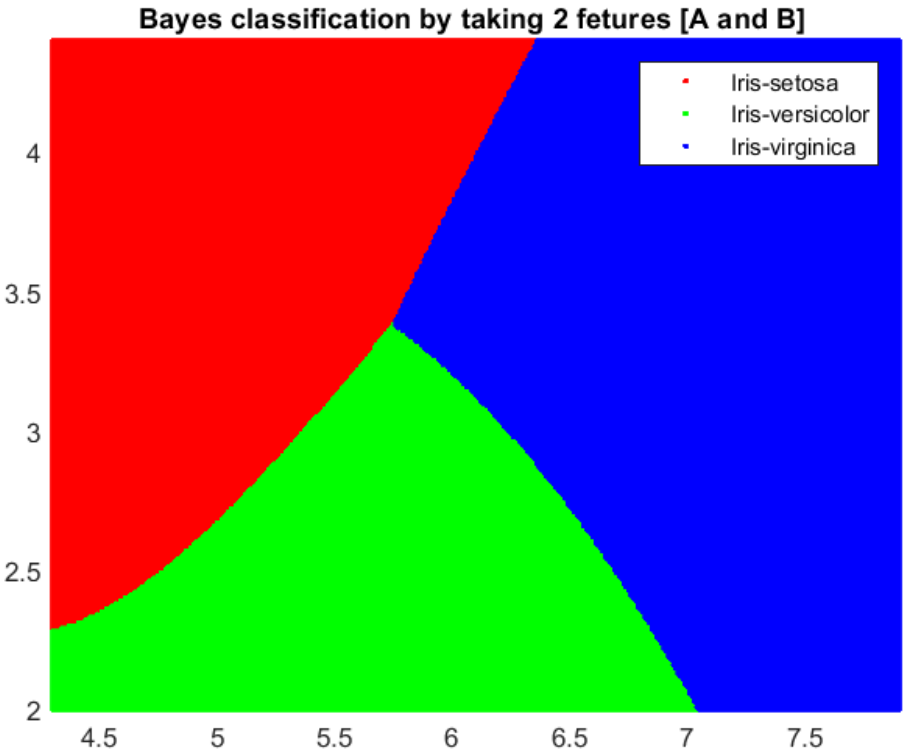
Iris Virginica—1/3

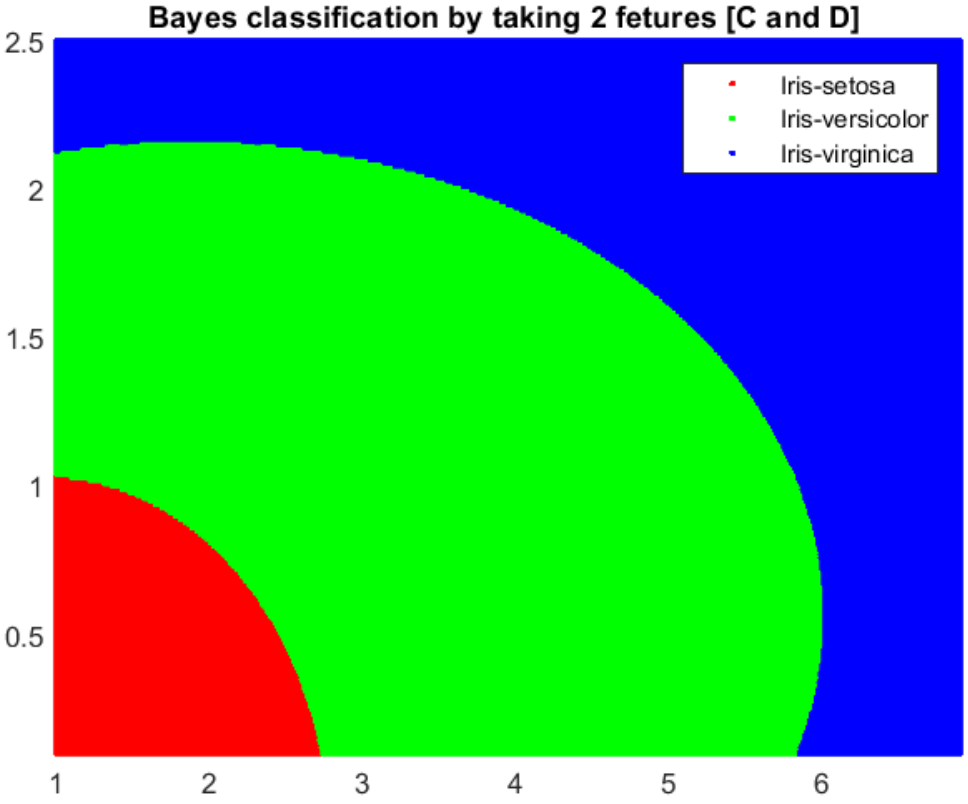
Feature Vector Dimension	No. of classes	Prior Probability of each class	Mean Vector & its Dimension **(by considering all 4 features at once)**	Covariance Matrix & its Dimension **(by considering all 4 features at once)**
4*1	3	Class1=1/3 Class2=1/3 Class3=1/3	M1=[5.0111 3.4311 1.4622 0.2489]^T M2=[5.9644 2.7644 4.2933 1.3356]^T M3=[6.6178 2.9733 5.5933 2.0222]^T **Size of m1= 4*1** **Size of m2= 4*1** **Size of m3= 4*1**	Cv1= 0.1315 0.1037 0.0168 0.0117; 0.1037 0.1495 0.0105 0.0112; 0.0168 0.0105 0.0329 0.0069; 0.0117 0.0112 0.0069 0.0121 Cv2= 0.2746 0.0912 0.1770 0.0567; 0.0912 0.1060 0.0850 0.0454; 0.1770 0.0850 0.2079 0.0743; 0.0567 0.0454 0.0743 0.0419 Cv3= 0.4329 0.1055 0.3247 0.0521; 0.1055 0.1065 0.0757 0.0490; 0.3247 0.0757 0.3197 0.0536; 0.0521 0.0490 0.0536 0.0790 **Size of CV1= 4*4** **Size of CV2= 4*4** **Size of CV3= 4*4**

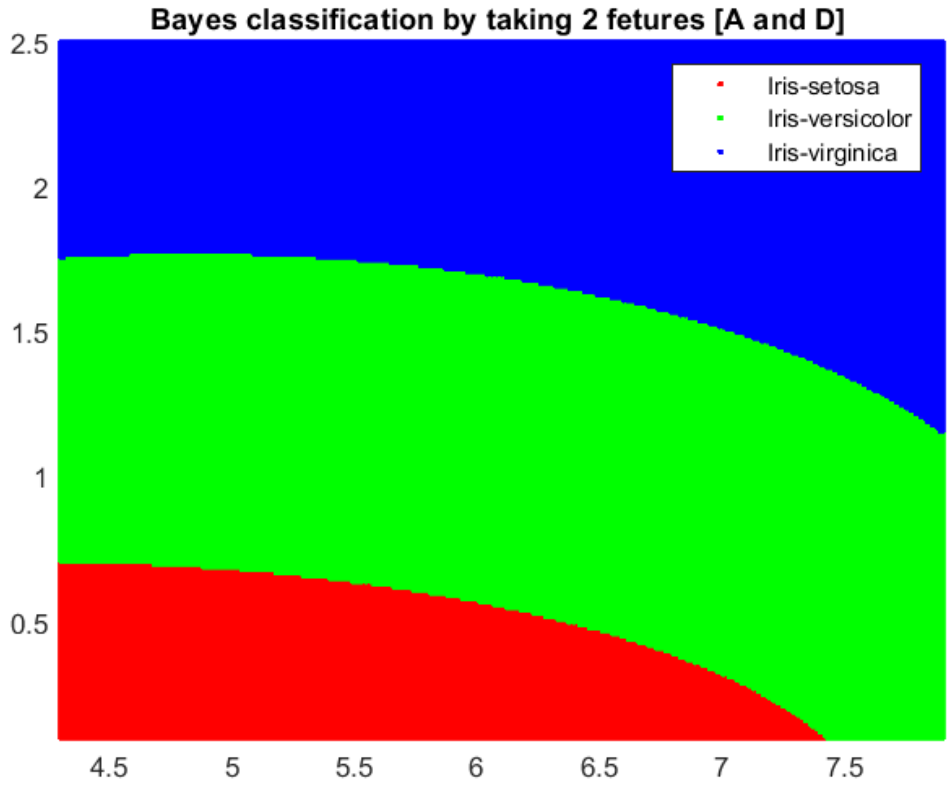
b)

```
Command Window
given x1 belongs to class Iris-setosa
given x2 belongs to class Iris-versicolor
given x3 belongs to class Iris-virginica

errorm =

   26.6667
         0
         0
         0
         0
         0

fx >>
```

c)


Bayes classification by taking 2 fetures [A and B]
Iris-setosa
Iris-versicolor
Iris-virginica
4
3.5
3
2.5
2
4.5
5
5.5
6
6.5
7
7.5


Bayes classification by taking 2 fetures [C and D]
Iris-setosa
Iris-versicolor
Iris-virginica
2.5
2
1.5
1
0.5
1
2
3
4
5
6



d)

Command Window

```
error_mohalanobis =

    6.6667
         0
         0
         0
         0
         0


error_Euclidian =

   13.3333
   13.3333
   13.3333
    6.6667
         0
         0

>>
```

e)

By using Bayesian classification

Feature pair	Classification Error
1&2	26.66 % **or** (4/15)
1 & 3	0 %
1 & 4	0 %
2 & 3	0 %
2 & 4	0 %
3 & 4	0 %

By using Mahalanobis classification

Feature pair	Classification Error
1&2	6.66 % **or** (4/15)
1 & 3	0 %
1 & 4	0 %
2 & 3	0 %
2 & 4	0 %
3 & 4	0 %

By using Euclidean distance-based classification

Feature pair	Classification Error
1&2	13.33 %
1 & 3	13.33 %
1 & 4	13.33 %
2 & 3	6.66 %
2 & 4	0 %
3 & 4	0 %

Inferences:

1) Given training data set contains, 135 samples of 4 feature vector are given. And each classes contains 45 numbers of samples of 4 feature vectors.
2) the testing data set contains, 1 samples of 4 feature vectors.
3) by considering the 2 features at a time and finding the mean and covariance matrices for each class using the training data set.
4) Perform the Bayesian classification, Euclidian distance and mahanalobis distance classification for the given testing data of 15 number of samples by selecting two feature vector combination and the error percentage is listed above.
5) Out of these 3 classifications Euclidian distance classification has high error probability compared to others and. Mahalanobis classification has the least error percentage compared to Bayesian classification.