**CM0669 Machine Learning and Computer Vision**

**Lab 5**

**Part 1:** Introduction to probabilities: Bayes classification

**1. Exercise –Probabilities and MAP decision rule**

Consider the following data set:

|  |  |  |  |
| --- | --- | --- | --- |
| Feature 1 | Feature 2 | Feature 3 | Class |
| 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 1 |
| 1 | 0 | 0 | 0 |
| 1 | 1 | 1 | 1 |
| 0 | 1 | 1 | 1 |
| 0 | 1 | 1 | 0 |
| **0** | **0** | **1** | **0** |

If we then have a test pattern with feature 1 as 0, feature 2 as 0 and feature 3 as 1, classify this pattern using the Naive Bayes classifier.

**2. Classification with Naive Bayes classifier**

Create six Matlab codes ‘Bayes\_SL\_SW\_out\_SE\_VERS.m’, ‘Bayes\_PL\_PW\_out\_SE\_VERS.m’, ‘Bayes\_SL\_SW\_PL\_PW\_out\_SE\_VERS.m’, ‘Bayes\_SL\_SW\_out\_VERS\_VIRG.m’, ‘Bayes\_PL\_PW\_out\_VERS\_VIRG.m’, ‘Bayes\_SL\_SW\_PL\_PW\_out\_VERS\_VIRG.m’ in a folder ‘Week5’.

Each Matlab code should implement a Bayes classifier. However, as you have seen in the previous labs, they should differ in terms of the input and output used. For instance, ‘Bayes\_SL\_SW\_out\_SE\_VERS.m’ should be a classifier using as input **S**epal **L**ength and **S**epal **W**idth and gives as output the ‘**SE**tosa’ and ‘**VERS**icolor’ class labels.

Similar to what you have seen in the previous Lab sessions for the evaluation of each classifier, 80 samples (specimens) should be used for the training while the testing should be performed on 20 samples.

**Hints:**

To estimate the parameters of the Bayes classifier use:

**classifier = NaiveBayes.fit(input\_training\_set,output\_training\_set);**

To get the output of the classifier for the test samples use:

**Y\_testing = classifier.predict(input\_testing\_set);**

Also, you may browse ‘help using the desktop’ and get help on the built-in functions ‘NaiveBayes.fit’ and ‘predict’.

1. Run all the codes and complete the results in the table given below.
2. Compare and analyse the results.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Bayes classifier | | | | | | | | |
| Input | | | | Output | | | Evaluation | |
| Sepal  Length | Sepal  Width | Petal  Length | Petal  Width | Setosa | Versicolor | Virginica | Training Error (%) | Testing  Error (%) |
| Yes | Yes | No | No | Yes | Yes | No |  |  |
| No | No | Yes | Yes | Yes | Yes | No |  |  |
| Yes | Yes | Yes | Yes | Yes | Yes | No | 0 | 0 |
| Yes | Yes | No | No | No | Yes | Yes |  |  |
| No | No | Yes | Yes | No | Yes | Yes |  |  |
| Yes | Yes | Yes | Yes | No | Yes | Yes | 7.5 | 40 |

**Part 2:** Unsupervised learning: K-means clustering

**1. Exercise –Euclidean distance**

Consider the following patterns in 2-D space:

V1=(1,1), V2=(7,1), V3=(6,7), V4=(6,2)

Calculate the following distances by hand

***D*(V1,V2), D(V2,V3), D(V3,V4), D(V2,V4)**

Create a Matlab function (in folder ‘Week5’) which calculates the Euclidean distance between two input vectors of any dimension. Execute the program for the previous vectors and ensure you get the same results.

**2. Clustering with the k-means method**

A Matlab code ‘kmeans\_SL\_SW\_SE\_VERS.m’ is provided. The program uses the k-means clustering method to group each feature vector of ‘Sepal Length’ and ‘Sepal Width’ in two clusters. Because there are two features, it is possible to display the distribution of the grouped samples in a 2-D graph. Similarly, the actual clusters (‘Setosa’ and ‘Versocolor’) are displayed to analyse the performance of k-means clustering.

1. Open up Matlab and type in ‘help kmeans’. Ensure you understand how to use the syntax of the built-in function ‘kmeans’. Open ‘kmeans\_SL\_SW\_ SE\_VERS.m’ and understand its content. Run the code and calculate the clustering error which is defined as

Clustering\_error = ; 0.0375

(average of below table)

**Hint**: you need to check ‘idx’ and see how many times it differs from Actual\_clusters\_set.

1. Create new Matlab codes ‘kmeans\_PL\_PW\_SE\_VERS.m’, ‘kmeans\_SL\_SW\_PL\_PW\_SE\_VERS.m’, ‘kmeans\_SL\_SW\_VERS\_VIRG.m’, ‘kmeans\_PL\_PW\_VERS\_VIRG.m’, ‘kmeans\_SL\_SW\_PL\_PW\_VERS\_VIRG.m’, ‘kmeans\_SL\_SW\_PL\_PW\_SE\_VERS\_VIRG.m’
2. Complete the table given below and analyse the results

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Input | | | | clusters | | |
| Sepal  Length | Sepal  Width | Petal  Length | Petal  Width | Setosa | Versicolor | Virginica | Clustering Error (%) |
| Yes | Yes | No | No | Yes | Yes | No |  |
| No | No | Yes | Yes | Yes | Yes | No |  |
| Yes | Yes | Yes | Yes | Yes | Yes | No |  |
| Yes | Yes | No | No | No | Yes | Yes |  |
| No | No | Yes | Yes | No | Yes | Yes |  |
| Yes | Yes | Yes | Yes | No | Yes | Yes |  |
| Yes | Yes | Yes | Yes | Yes | Yes | Yes |  |