Computer Science and Software Engineering Départment

**COMP6731 & COMP473**
**Pattern Recognition**

# Team Assignment 2 & 3

## ALL Team members:


| Name | Family Name | StudentID# | Email ID |
|---|---|---|---|
| Prasanth | Ambalam Jawaharlal | 40042116 | santhtce@gmail.com |
| Anil | Heggodu Raghavendra | 40042818 | anilconcordia@gmail.com |
| Simarpreet | Singh | 40049885 | simar.admissions@gmail.com |
| Sofia | Gakhar | 40067862 | Sofia.gakhar1991@gmail.com |

Q1) ①

a) Maximum Likelihood Estimation of $\theta$ (Binomial Distribution)

As given, in any given set of N test samples the Probability that K- Samples are misclassified is given by the binomial distribution:

$$p(K|\theta) = \begin{bmatrix} N \\ K \end{bmatrix} \theta^K (1-\theta)^{N-K}$$

Where $\begin{bmatrix} N \\ K \end{bmatrix} = \frac{N!}{K!(N-K)!}$ is the number of ways that any K out of the N samples can be misclassified

Taking the log gives us:

$$l(\theta) = \log p(K|\theta) = \log \begin{bmatrix} N \\ K \end{bmatrix} + \log(\theta^K) + \log(1-\theta)^{N-K}$$

$$l(\theta) = \log P(K|\theta) = \log \begin{bmatrix} N \\ K \end{bmatrix} + K \log(\theta) + (N-K)\log(1-\theta)$$

taking the derivative and setting to zero:

$$\frac{\partial l(\theta)}{\partial \theta} = \frac{\partial}{\partial \theta}\left[\log\left[\begin{matrix} N \\ K \end{matrix}\right] + K\log(\theta) + (N-K)\log(1-\theta)\right]$$

$$\frac{\partial l(\theta)}{\partial \theta} = 0 + \frac{K}{\theta} - \left[\frac{N-K}{1-\theta}\right]$$

equating $\frac{\partial l(\theta}{\partial \theta} = 0$.

we get

$$0 = \frac{K}{\theta} - \left[\frac{N-K}{1-\theta}\right]$$

$$\frac{K}{\theta} = \frac{N-K}{1-\theta}$$

$$(1-\theta)K = \theta(N-K)$$

$$K - \cancel{\theta K} = \theta N - \cancel{\theta K}.$$

$$\theta N = K.$$

$$\boxed{\hat{\theta}_{ML} = \frac{K}{N}}$$

Q1. b)

From the MLE expression for Binomial Distribution

$$\hat{\theta}_{ML} = \frac{K}{N}$$

We can conclude that the ML estimate of error-rate is just the fraction of samples misclassified.

④

Q2)

b) $m = [10 \quad 0]^T$

$$S = \begin{bmatrix} 16 & -12 \\ -12 & 34 \end{bmatrix}$$

To find the eigen values we have the following equation

$$\det(S - \lambda I) = 0$$

where $\lambda$ is the eigen values

$I$ is the Identity matrix i.e.

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Substituting the S & I in the above equation to find $\lambda$

$$|S - \lambda I| = \left| \begin{bmatrix} 16 & -12 \\ -12 & 34 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = 0$$

$$\Rightarrow \det\left( \begin{bmatrix} 16-\lambda & -12 \\ -12 & 34-\lambda \end{bmatrix} \right) = 0$$

Solving the above eq$^n$ using determinant

$$(16-\lambda)(34-\lambda) - (-12)(-12) = 0$$

$$544 - 16\lambda - 34\lambda + \lambda^2 - 144 = 0$$

$$\lambda^2 + 400 - 50\lambda = 0$$

$$\Rightarrow \begin{bmatrix} -24 & -12 \\ -12 & -6 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} -24a - 12b \\ -12a - 6b \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Rightarrow -24a - 12b = 0 \ldots (1)$$

$$-12a - 6b = 0 \ldots (2)$$

From either $eq^n$ (1) or (2) we can inference

$$-24a = 12b.$$

$$a = -\frac{12}{24} b.$$

$$a = -\frac{1}{2} b.$$

$\therefore$ $a = -\frac{1}{2}$ $b = 1$ (by the ratio role).

$$\boxed{\phi_2 \text{ or } V_2 = \begin{pmatrix} a \\ b \end{pmatrix} \Rightarrow \phi_2 \text{ or } V_2 = \begin{bmatrix} -1/2 \\ 1 \end{bmatrix}}$$

For $\underline{\lambda_1 = 10}$.

$(S - \lambda I)\phi_1 = 0$ where $\phi_1 = \begin{bmatrix} a \\ b \end{bmatrix}$

Where $S = \begin{bmatrix} 16 & -12 \\ -12 & 24 \end{bmatrix}$ $\lambda = 10$ $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

substituting above values in the above $eq^n$.

$$\left( \begin{bmatrix} 16 & -12 \\ -12 & 24 \end{bmatrix} - 10 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \begin{bmatrix} a \\ b \end{bmatrix} = 0$$

Solving the above eqn for 'λ'

$\Rightarrow \lambda^2 - 40\lambda - 10\lambda + 400$

$\Rightarrow \lambda(\lambda-40) - 10(\lambda-40) = 0$

$(\lambda-10)(\lambda-40) = 0 \quad \Rightarrow \lambda_1 - 10 = 0 / \lambda_2 - 40 = 0$

$\boxed{\lambda_2 = 40 \quad \lambda_1 = 10}$

$\boxed{\text{eigen values are } 10, 40.}$

Q2.c) The eigenvectors of Covariance matrix S can be calculated by substituting $\lambda_1$ & $\lambda_2$ values in the below equation

$$(S - \lambda I)\phi_2 = 0 \quad \text{where } \phi_2 = \begin{bmatrix} a \\ b \end{bmatrix}$$

Substituting the $\lambda_2$ & S, I values in the above equation

$$\underline{\lambda_2 = 40} \quad S = \begin{bmatrix} 16 & -12 \\ -12 & 34 \end{bmatrix} \quad I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\Rightarrow \left( \begin{bmatrix} 16 & -12 \\ -12 & 34 \end{bmatrix} - 40 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \begin{bmatrix} a \\ b \end{bmatrix} = 0$$

$$\left( \begin{bmatrix} 16 & -12 \\ -12 & 34 \end{bmatrix} - \begin{bmatrix} 40 & 0 \\ 0 & 40 \end{bmatrix} \right) \begin{bmatrix} a \\ b \end{bmatrix} = 0.$$

$$\begin{bmatrix} 16-40 & -12 \\ -12 & 34-40 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = 0.$$

$$\Rightarrow \left( \begin{bmatrix} 6 & -12 \\ -12 & 14 \end{bmatrix} \right) \begin{bmatrix} a \\ b \end{bmatrix} = 0$$

$$\Rightarrow \begin{bmatrix} 6a - 12b \\ -12a + 14b \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Rightarrow \quad 6a - 12b = 0 \quad \cdots (1)$$

$$-12a + 14b = 0 \quad \cdots (2)$$

From eq$^n$ (1) & (2) we can inferred that

$6a = 12b$

$\therefore a = 2b$

$\therefore \quad a = 2 \quad b = 1$

$$\boxed{\phi_1 \text{ or } V_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}}$$

$\therefore$ eigen vector $= [\phi_1 \quad \phi_2]$

$$\boxed{\text{eigen vector} = \begin{bmatrix} 2 & -1/2 \\ 1 & 1 \end{bmatrix}}$$

Q2.c) From the above, we can tell that

$\phi_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \qquad \phi_2 = \begin{bmatrix} -1/2 \\ 1 \end{bmatrix}$

$\therefore$ eigen vector $= [\phi_1 \quad \phi_2]$

$$\boxed{\text{eigen vector}(\phi) = \begin{bmatrix} 2 & -1/2 \\ 1 & 1 \end{bmatrix}}$$

Q2. f) matrix $\phi = \begin{bmatrix} 2 & -1/2 \\ 1 & 1 \end{bmatrix}$

The new covariance matrix is given by

$\Sigma = \phi^T S \phi$.

we know that $\phi = \begin{bmatrix} 2 & -1/2 \\ 1 & 1 \end{bmatrix}$

$\phi^T = \begin{bmatrix} 2 & 1 \\ -1/2 & 1 \end{bmatrix}$ $S = \begin{bmatrix} 16 & -12 \\ -12 & 34 \end{bmatrix}$

$\Sigma = \phi^T S \phi$

Substituting above values in the above eq$^n$

$$\Rightarrow \begin{bmatrix} 2 & 1 \\ -1/2 & 1 \end{bmatrix} \begin{bmatrix} 16 & -12 \\ -12 & 34 \end{bmatrix} \begin{bmatrix} 2 & -1/2 \\ 1 & 1 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 32-12 & -24+34 \\ -\frac{1}{2}(16)-12 & -\frac{1}{2}(-12)+34 \end{bmatrix} \begin{bmatrix} 2 & -1/2 \\ 1 & 1 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 20 & +10 \\ -20 & 40 \end{bmatrix} \begin{bmatrix} 2 & -1/2 \\ 1 & 1 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 40+10 & -\frac{1}{2}(20)+10 \\ -40+40 & -\frac{1}{2}(-20)+40 \end{bmatrix}$$

$$\boxed{\Sigma \Rightarrow \begin{bmatrix} 50 & 0 \\ 0 & 50 \end{bmatrix}}$$

When S is transformed to $\Sigma$, the covariance values become '0'. So in '$\Sigma$', the features are un-related. In other words, Features are well seperated

Q2.g) The new value to plot contour are.

$$m' = \phi^T m$$

$$m = \begin{bmatrix} 10 \\ 0 \end{bmatrix} \qquad \phi = \begin{bmatrix} 2 & -1/2 \\ 1 & 1 \end{bmatrix}$$

$$\phi^T = \begin{bmatrix} 2 & 1 \\ -1/2 & 1 \end{bmatrix}$$

$$m' = \begin{bmatrix} 2 & 1 \\ -1/2 & 1 \end{bmatrix} \begin{bmatrix} 10 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 20 - 0 \\ -\frac{1}{2}(10) - 0 \end{bmatrix}$$

$$m' = \begin{bmatrix} 20 \\ -5 \end{bmatrix}.$$

$$\boxed{m' = [20 \ \ -5]^T}$$

From the Q2.5 we know that

$$\Sigma = \begin{bmatrix} 50 & 0 \\ 0 & 50 \end{bmatrix}.$$

From the eq$^n$ $S' = \Sigma$

$$\boxed{\therefore \quad S' = \begin{bmatrix} 50 & 0 \\ 0 & 50 \end{bmatrix}}$$

When compared to first contour plot, Axis of the second contour plot is aligned in the direction of maximum variance.

Scanned by CamScanner

**Note:**

- The Training set is same as the training set given in Question 5 in Assignment 5 (**Training_Dataset.csv**)
- The Test data is generated randomly based on mean and covariance of Training data set of each class (10 samples per class is generated and stored in **Test_Dataset.csv and is used in all the problems in Question 3**)

# Question 3a)General Classification with SVM, KNN,PNN, MLP

**i) SVM : (*Accuracy: 70%*)**

A Linear SVM classifier with 3 classes is trained using the Training_Dataset.csv and tested using the Test_Dataset.csv. We got the following confusion Matrix.

| | | *Actual* | | |
|---|---|---|---|---|
| | | **1** | **2** | **3** |
| *Pred* | **1** | 7 | 5 | 0 |
| | **2** | 1 | 5 | 1 |
| | **3** | 2 | 0 | 9 |

**ii) KNN Classifier**:

K Nearest Neighbour Classifier with **Euclidean distance** is used to find the nearest neighbor. The classification experiment is done for **K = 3 and K=5.** We got the following confusion matrices.

a) **KNN-3: (*Accuracy: 70%*)**

*Confusion Matrix:*

| | | Prediction | | |
|---|---|---|---|---|
| Actual | | **1** | **2** | **3** |
| | **1** | 5 | 2 | 3 |
| | **2** | 2 | 7 | 1 |
| | **3** | 1 | 0 | 9 |

b) **KNN-5:(*Accuracy: 67%*)**

| Actual | Prediction 1 | 2 | 3 |
|---|---|---|---|
| **1** | 5 | 2 | 3 |
| **2** | 3 | 5 | 1 |
| **3** | 1 | 0 | 10 |

**<u>iii) PNN:</u> (*Accuracy: 73%*)**

The confusion matrix that we got for the test data.

| Predicted | Actual 1 | 2 | 3 |
|---|---|---|---|
| **1** | 8 | 4 | 2 |
| **2** | 0 | 5 | 0 |
| **3** | 2 | 0 | 9 |

**<u>iv) MLP:</u> (Accuracy: 76% )**

The multilayer perceptron neural network has been trained using the training data. Sigmoidal activation function has been used as the transfer function for the hidden layer neurons.

The confusion matrix that we got for the test data as follows.

| **Actual** | **Predicted 1** | **2** | **3** |
|---|---|---|---|
| **1** | 7 | 1 | 2 |
| **2** | 1 | 7 | 2 |
| **3** | 1 | 0 | 9 |

## Conclusion for Q3a:

- MLP has the best classification rate of 76%

## <u>Question 3b):</u> Choosing Parameters or Hyper Parameters to improve the classification

**<u>SVM:</u>**

- Kernel has been changed to ***'Radial'***
- Scaling has been Disabled.
- Gamma has been set as 0.1

By doing so, the classification rate improved, the confusion matrix which we got is

| | | **Actual** | |
|---|---|---|---|
| **Predicted** | **1** | **2** | **3** |
| **1** | 9 | 0 | 2 |
| **2** | 1 | 9 | 2 |
| **3** | 0 | 0 | 7 |

The accuracy has improved from **70%** to **83%**

**<u>KNN:</u>**

- **K has set to 7 (seven nearest neighbours are considered)**

By doing so, the classification rate improved, the confusion matrix which we got is

| | | **Prediction** | |
|---|---|---|---|
| **Actual** | **1** | **2** | **3** |
| **1** | 6 | 2 | 2 |
| **2** | 3 | 7 | 0 |
| **3** | 1 | 1 | 8 |

The accuracy has improved from **67%** to **70%**

**PNN:**

- The Smoothing parameter has been changed.
- Got the best possible result when sigma set to 0.8

| | Actual | | |
|---|---|---|---|
| **Predicted** | **1** | **2** | **3** |
| **1** | 8 | 0 | 0 |
| **2** | 1 | 9 | 1 |
| **3** | 1 | 0 | 10 |

The accuracy has improved from **70%** to **90%**

**MLP:**

- Max iteration has been increased
- Momentum has been set to 0.9
- Learning rate has set to constant

| | Predicted | | |
|---|---|---|---|
| **Actual** | **1** | **2** | **3** |
| **1** | 6 | 2 | 2 |
| **2** | 1 | 8 | 1 |
| **3** | 0 | 0 | 10 |

The accuracy rate has improved from **76%** to **83%**

## Conclusion for Q3b:

- PNN has the best classification rate of 90%
- MLP/SVM share the second best classification rate with 83%
- KNN has an average classification rate of 70%

## Question 3c: Max Voting Classifier

- Max voting algorithm is used.
- The improved version of SVM, KNN-1 and PNN (Solution to Q3b) has been used for Max Voting classification.
- Train all the classifiers with the same Training Data set (Training_Dataset.csv)
- Predict each Test Point from Test_Dataset.csv by passing it to all the three classifiers.
- Choose the **'Class'** which has got the highest vote from the above classifiers.

The confusion Matrix which we got from the Max voting algorithm is

| | | **max_voted_output** | |
|---|---|---|---|
| **actual_data** | **1** | **2** | **3** |
| **1** | 8 | 1 | 1 |
| **2** | 0 | 9 | 1 |
| **3** | 1 | 0 | 9 |

The accuracy of the Max voting classifier is **86%**

**Conclusion:**

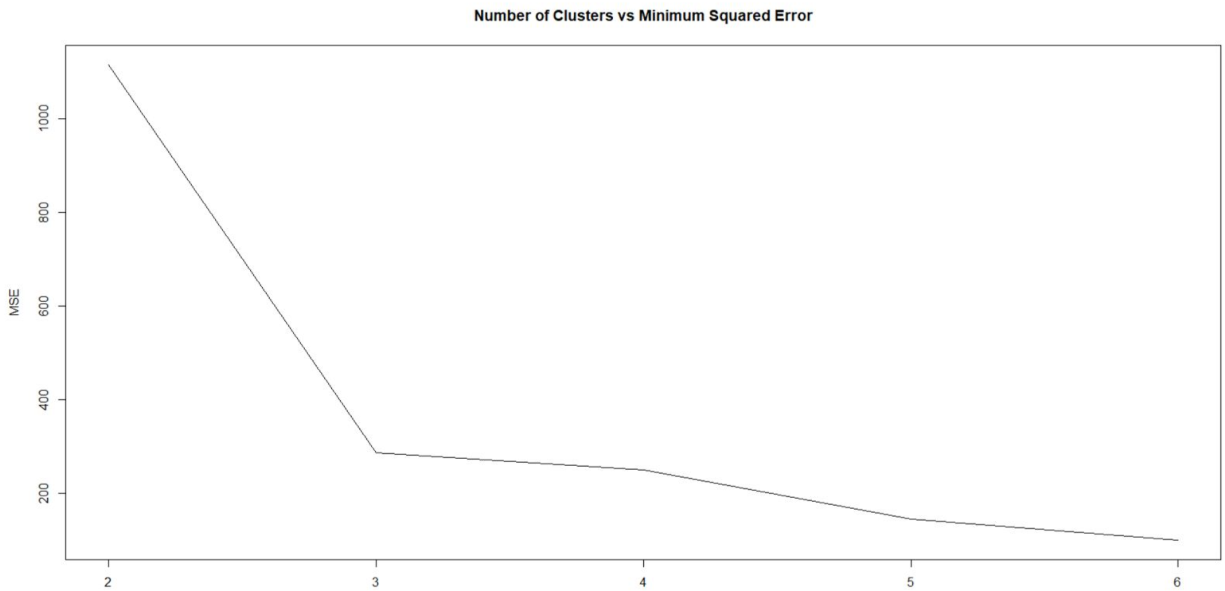
**Since all the classifiers are combined, the Max Voting Classifier gave the Best classification rate.**

## Question 3d):K Means Clustering

- The Testing data set has been clustered using K Means algorithm
- Different values of K (2,3,4,5,6) has been used to cluster the data.
- Sum of Squared Errors has been computed for all the clusters (For k = 2,3,4,5,6)
- Elbow Method:
    - The SSE has been plotted against the Value of corresponding K.
    - From the graph, as per this data, elbow is created at k=3, so we will take 3 clusters in this case.
    - With K=3, Our clustering algorithm is partially able to detect original distribution of data belonging to three classes of W1, W2, and W3.
    - **Accuracy : 63%**
    - Confusion Matrix

| | | **Actual** | |
|---|---|---|---|
| **Predicted** | **1** | **2** | **3** |
| **1** | 6 | 5 | 0 |
| **2** | 2 | 5 | 2 |
| **3** | 2 | 0 | 8 |

Graph:-



Number of Clusters vs Minimum Squared Error
MSE
1000
800
600
400
200
2
3
4
5
6

## REFERENCES: -


[1] http://www.eng.uwaterloo.ca/~a28wong/

[2] http://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html

[3] https://www.rdocumentation.org/packages/DMwR/versions/0.4.1/topics/kNN

[4] https://www.rdocumentation.org/packages/e1071/versions/1.6-8/topics/svm

[5] https://www.rdocumentation.org/packages/Rcmdr/versions/2.0-4/topics/KMeans

[6] https://www.rdocumentation.org/packages/pnn/versions/1.0.1/topics/pnn-package