

## **PROJECT PART 1**

### **Feature Extraction, Density Estimation and Bayesian Classification.**

#### **1)Steps followed in the Project:**

1) First, we import the data from the .mat files.

2) We then extract the data and store them in Xtrain,ytrain,Xtest,ytest respectively.

3) Then we create the following functions.

“skewness”- to calculate the skewness of every image in the given data.

”brighttodark”- to calculate the ratio of bright pixels to dark pixels of every image in the given data using a threshold T.

“normalize”- which uses these functions to normalize our data.

4) We then, convert these normalized data into a dataframe, which makes it easier for us to do the remaining tasks.

5) We then use Maximum Likelihood estimation, to find the estimates of mean and co-variance of both the classes.

6) We then calculate the probability density function using these estimates.

7) Then we define a function which does minimum error rate classification given the prior probabilities and X.

8) Then we calculate the error rate for every configuration.

## 2)Head of the Normalized Features when the Value of T is 150:

Head of Train Dataframe

	0	1	label
0	-0.312630	-0.051244	3
1	0.965460	-0.847999	3
2	0.516517	-0.815870	3
3	0.297382	-0.489563	3
4	1.906885	-1.441811	3

Head of test Dataframe

	0	1	label
0	-1.174485	1.142590	3
1	-0.444426	0.203891	3
2	-1.184313	1.324994	3
3	-1.922119	2.850372	3
4	-0.139868	-0.286138	3

## 3)Head of the Normalized Features when the Value of T is 200:

Head of Train Dataframe

	0	1	label
0	-0.312630	-0.029591	3
1	0.965460	-0.826533	3
2	0.516517	-0.725447	3
3	0.297382	-0.382378	3
4	1.906885	-1.157556	3

Head of test Dataframe

	0	1	label
0	-1.174485	1.214555	3
1	-0.444426	0.447426	3
2	-1.184313	0.843338	3
3	-1.922119	3.075649	3
4	-0.139868	-0.374038	3

#### 4)Mean and Co-variance when T is 150:

The maximum likelihood estimates for mean of 3 is

```
0    -0.379520
```

```
1     0.358546
```

```
dtype: float64
```

The maximum likelihood estimates for mean of 7 is

```
0     0.371585
```

```
1    -0.351050
```

```
dtype: float64
```

The maximum likelihood estimates for covariance of 3 is

```
[[ 0.8437565  -0.91737529]
```

```
[-0.91737529  1.10355626]]
```

The maximum likelihood estimates for covariance of 7 is

```
[[ 0.87417217 -0.72010492]
```

```
[-0.72010492  0.64980588]]
```

#### 5)Mean and Co-variance when T is 200:

The maximum likelihood estimates for mean of 3 is

```
0    -0.379520
```

```
1     0.306744
```

```
dtype: float64
```

The maximum likelihood estimates for mean of 7 is

```
0     0.371585
```

```
1    -0.300330
```

```
dtype: float64
```

The maximum likelihood estimates for covariance of 3 is

```
[[ 0.8437565  -0.91624532]
```

```
[-0.91624532  1.136679   ]]
```

The maximum likelihood estimates for covariance of 7 is

```
[[ 0.87417217 -0.72986478]
```

```
[-0.72986478  0.68416819]]
```

**6)Error Rates for all possible values of T and all possible prior probabilities:**

<b>Threshold (T)</b>	<b>P(3) and p(7)</b>	<b>Train Error Rate</b>	<b>Test Error Rate</b>
150	0.5 and 0.5	0.3428299272601316	0.33853083853083854
150	0.3 and 0.7	0.40656390717007274	0.4078309078309078
200	0.5 and 0.5	0.28827502597852445	0.2955647955647956
200	0.3 and 0.7	0.40318669899549703	0.41302841302841303