CS669: Pattern Recognition

Programming Assignment 4

Instructor: Dileep A.D.

By-

Group 12

Ayush Garg (B13111)

Himanshu Singal (B13312)

Hitesh Tarani (B13139)

**Introduction:**

**Datasets:**

**Dataset 1:** 2-dimensional artificial data of 3 or 4 classes:

(a) Linearly separable dataset

(b) Nonlinearly separable data set **(Only for SVM)**

**Dataset 2:** Real world data set:

(a) Two dimensional speech dataset (used in Assignment 1)

(b) Image dataset

**Classifiers to be built:**

1. Bayes classifier using unimodal Gaussian distribution on Dataset-1(a) and Dataset-2.

2. Bayes classifier using GMM on Dataset-2. GMM is built using the K-means clustering to initialize the parameters.

3. Build Bayes classifier using unimodal Gaussian distribution on the 1-dimensional representation of Dataset-1(a) and Dataset-2(a) obtained using PCA.

4. Build Bayes classifier using unimodal Gaussian distribution and GMM on the reduced dimensional representations of Dataset-2 obtaind using PCA.

5. Fisher linear discriminant analysis (FDA) based classifier on Dataset-1 and Dataset-2. Use both Bayes classifier using unomodal Gaussian and GMM.

6. Perceptron-based classifier on Dataset-1.

7. SVM-based classifier using (a) linear kernel, (b) polynomial kernel and (c) Gaussian/RBF kernel on Dataset-1 and Dataset-2

8. SVM-based classifier using (a) linear kernel, (b) polynomial kernel and (c) Gaussian/RBF kernel on the reduced dimensional representations of Dataset-2 obtained using PCA.

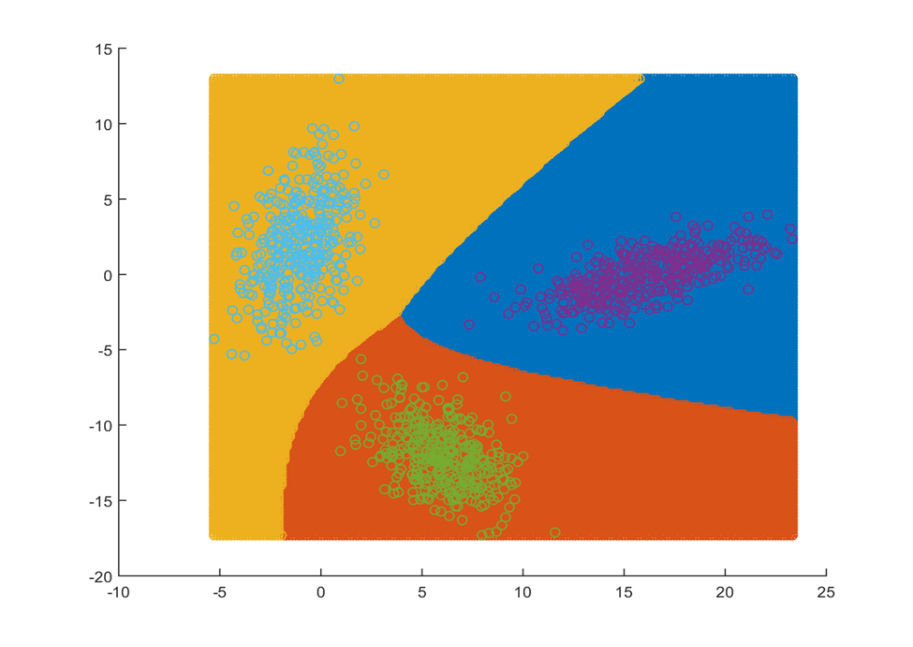
**Perform the experiments on different values of *K* in GMM for different reduced dimensions. Also perform experiments on different values of SVM and kernel parameters for different reduced dimensions.**

**Note: Concatenate the local feature vectors in each image to form a super vector. Now each image will become a vector with 36\*23-dimension.**

**1. Bayes classifier using Unimodal Gaussian distribution on Dataset-1(a) and Dataset-2.**

**a. Dataset-1(a): Linearly Separable**

1. **Classification Accuracy** = 100%
2. **Confusion Matrix based on performance for test data.**



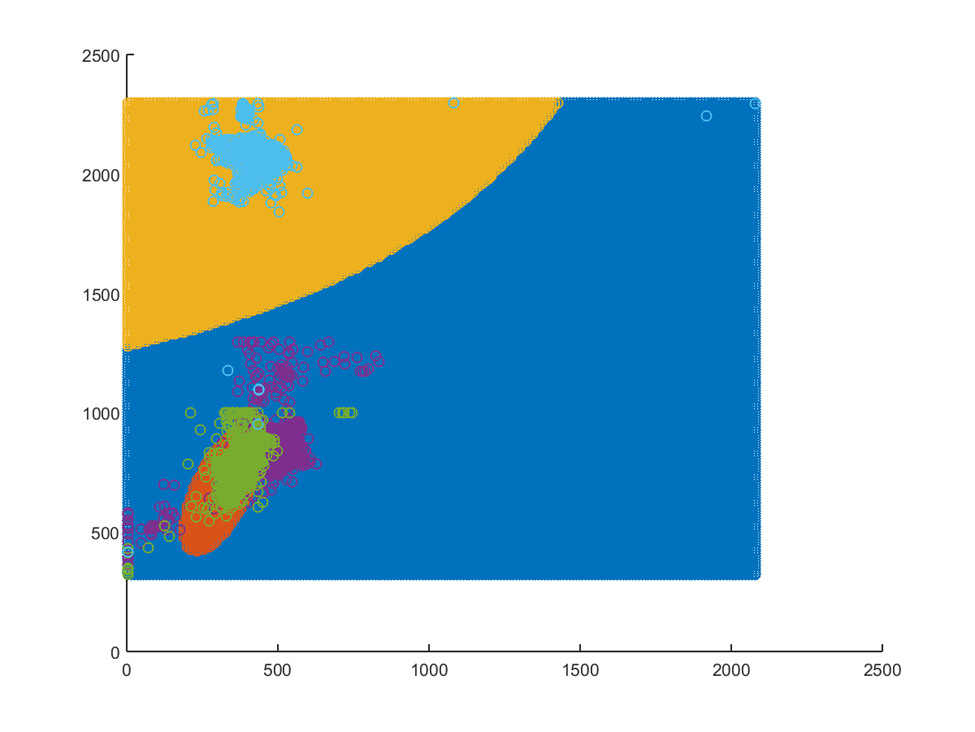
**b. Dataset-2**

Two dimensional speech dataset (used in Assignment 1)

**Classification Accuracy: - 82.698%**

**Confusion Matrix**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Class1 | Class2 | Class3 |
| Class1 | 411 | 203 | 0 |
| Class2 | 91 | 531 | 0 |
| Class3 | 11 | 8 | 554 |



**2. Bayes classifier using GMM on Dataset-2. GMM is built using the K-means clustering to initialize the parameters.**

|  |  |  |
| --- | --- | --- |
| **Dataset** | **Confusion Matrix** | **Accuracy** |
| Two dimensional speech dataset | **K = 8,8,4** | 84% |
| Image dataset | **K= 16** | 93.173% |

**3. Build Bayes classifier using unimodal Gaussian distribution on the 1- dimensional representation of Dataset-1(a) and Dataset-2(a) obtained using PCA.**

|  |  |  |
| --- | --- | --- |
| **Dataset** | **Confusion Matrix** | **Accuracy** |
| Linearly Separable |  | 98.133 % |
| Two dimensional speech dataset |  | 78.165% |

**4. Build Bayes classifier using unimodal Gaussian distribution and GMM on the reduced dimensional representations of Dataset-2 obtained using PCA.**

1. Two dimensional speech dataset

GMM

|  |  |  |  |
| --- | --- | --- | --- |
| **L** | **No. of Clusters** | **Confusion Matrix** | **Accuracy** |
| 1 | **3** |  | 86.78% |

1. Image dataset

Univariate

|  |  |  |
| --- | --- | --- |
| **L** | **Confusion Matrix** | **Accuracy** |
| 1 |  | 60.24% |
| 2 |  | 64.659% |
| 4 |  | 79% |
| 8 |  | 83.53% |
| 16 |  | 84.73% |

Multivariate

|  |  |  |  |
| --- | --- | --- | --- |
| **L** | **No. of Clusters** | **Confusion Matrix** | **Accuracy** |
| 1 | 3 |  | 64.25% |
| 4 | 3 |  | 81.124% |
| 8 | 3 |  | 83.93% |

**5. Fisher linear discriminant analysis (FDA) based classifier on Dataset-1 and Dataset-2. Use both Bayes classifier using unimodal Gaussian and GMM.**

**Dataset-1: Linearly Separable**

|  |  |  |
| --- | --- | --- |
| **Model** | **Confusion Matrix** | **Accuracy** |
| Unimodal |  | 100% |
| GMM |  | 100% |

**Dataset-2: Speech Dataset**

|  |  |  |
| --- | --- | --- |
| **Model** | **Confusion Matrix** | **Accuracy** |
| Unimodal |  | 86.62% |
| GMM |  | 86.67% |

**Dataset-2: Image Dataset**

Matrix was coming out to be singular.

**7. SVM-based classifier using (a) linear kernel, (b) Gaussian/RBF kernel on Dataset-1 and Dataset-2.**

**Dataset-1(a): Linearly Separable**

|  |  |  |
| --- | --- | --- |
| **Kernel** | **Confusion Matrix** | **Accuracy** |
| Linear |  | 100% |
| RBF/Gaussian |  | 100% |

**Dataset-1(b): Non-Linearly Separable**

**Ring**

|  |  |  |
| --- | --- | --- |
| **Kernel** | **Confusion Matrix** | **Accuracy** |
| Linear |  | 80% |
| RBF/Gaussian |  | 100% |

**Spiral**

|  |  |  |
| --- | --- | --- |
| **Kernel** | **Confusion Matrix** | **Accuracy** |
| Linear |  | 53.68% |
| RBF/Gaussian |  | 90.798% |

**Interlock**

|  |  |  |
| --- | --- | --- |
| **Kernel** | **Confusion Matrix** | **Accuracy** |
| Linear |  | 96.8% |
| RBF/Gaussian |  | 100 % |

**Dataset-2(a): Two dimensional Speech**

|  |  |  |
| --- | --- | --- |
| **Kernel** | **Confusion Matrix** | **Accuracy** |
| Linear |  | 87.065% |
| RBF/Gaussian |  | 85.13% |

**Dataset-2(b): Image Dataset**

|  |  |  |
| --- | --- | --- |
| **Kernel** | **Confusion Matrix** | **Accuracy** |
| Linear |  | 95.582% |
| RBF/Gaussian |  | 36.145% |

**8. SVM-based classifier using (a) linear kernel, (b) polynomial kernel and (c) Gaussian/RBF kernel on the reduced dimensional representations of Dataset-2 obtained using PCA.**

**Dataset-2(a): Speech Data**

|  |  |  |  |
| --- | --- | --- | --- |
| **Kernel** | **L** | **Confusion Matrix** | **Accuracy** |
| Linear | 1 |  | 85.24% |
| RBF/Gaussian | 1 |  | 88.115% |

**Dataset-2(b): Image Data**

|  |  |  |  |
| --- | --- | --- | --- |
| **Kernel** | **L** | **Confusion Matrix** | **Accuracy** |
| Linear | 1 |  | 65.863% |
| RBF/Gaussian | 1 |  | 63.855% |
| Linear | 16 |  | 83.534% |
| RBF/Gaussian | 16 |  | 61.044% |