**A Report on**

**CS669: Assignment-2&3**

**PATTERN RECOGNITION**

**Team Members (Group-01):**

**D22140 (Vandita Dutt)**

**D22180 (Nandnai Sharma)**

**S23083 (Kajal)**

**Course Instructor**

**Dr. Dileep A. D.**

## 

## 

**IIT Mandi Academic Year 2023-2024**

# TABLE OF CONTENT

## 

[**TABLE OF CONTENT 2**](#_heading=h.cj7hslwk5jlf)

[**1. Problem Statement 6**](#_heading=h.v58x94jvleym)

[**2. Approach to finding a solution 8**](#_heading=h.33fkaqd5765x)

[**3. DATASET 11**](#_heading=h.2qo6j81c6ag7)

[**3.1 Non Linearly Separable data (Dataset-1) 11**](#_heading=h.p2lytng0o6xi)

[3.1.1 Decision Boundary 12](#_heading=h.c9ibqncxi2t1)

[**Cluster =1 12**](#_heading=)

[**Cluster =2 13**](#_heading=)

[**Cluster =4 14**](#_heading=)

[**Cluster =8 15**](#_heading=)

[**Cluster =16 16**](#_heading=)

[**Cluster =32 17**](#_heading=)

[**3.1.2 log-likelihood vs Iterations plots 18**](#_heading=)

[**Cluster=1 18**](#_heading=)

[**Cluster=2 18**](#_heading=)

[**Cluster=4 19**](#_heading=)

[**Cluster=8 19**](#_heading=)

[**Cluster=16 20**](#_heading=)

[**Cluster=32 20**](#_heading=)

[**3.1.3 Confusion Matrix, Precision, Recall and F-measure 21**](#_heading=)

[**Cluster=1 21**](#_heading=)

[(Class 1,2) 21](#_heading=h.xc1ztc613oeb)

[(Class 1,3) 21](#_heading=h.izydqmsqdgxl)

[(Class 2,3) 21](#_heading=h.k04sjup3z4ra)

[(Class 1,2,3) 22](#_heading=h.jvq8dyxhe25p)

[**Cluster=2 22**](#_heading=)

[(Class 1,2) 22](#_heading=h.aplgmmvdryv9)

[(Class 1,3) 22](#_heading=h.eq6xegc4hsl1)

[(Class 2,3) 23](#_heading=h.qkggrjl62toy)

[(Class 1,2,3) 23](#_heading=h.83kg9h1f516u)

[**Cluster=4 23**](#_heading=)

[(Class 1,2) 23](#_heading=h.hh00eoe2djxx)

[(Class 1,3) 24](#_heading=h.fq06d081bfua)

[(Class 2,3) 24](#_heading=h.2sbh8htjmred)

[(Class 1,2,3) 24](#_heading=h.fk80ktcte6f3)

[**Cluster=8 25**](#_heading=)

[(Class 1,2) 25](#_heading=h.a9mypnwsrf5y)

[(Class 1,3) 25](#_heading=h.2qbnavzafh4o)

[(Class 2,3) 25](#_heading=h.bmfxsz3f1dlh)

[(Class 1,2,3) 26](#_heading=h.5kyipqa29i7j)

[**Cluster=16 26**](#_heading=)

[(Class 1,2) 26](#_heading=h.z1cz416et2c8)

[(Class 1,3) 26](#_heading=h.50p0co7aifmd)

[(Class 2,3) 27](#_heading=h.7ol4jai8u6ap)

[(Class 1,2,3) 27](#_heading=h.7yr55cvqqew5)

[**Cluster=32 27**](#_heading=)

[(Class 1,2) 27](#_heading=h.17mtu07a5y6u)

[(Class 1,3) 28](#_heading=h.2ru836wcjbrp)

[(Class 2,3) 28](#_heading=h.rsfckvi4e90b)

[(Class 1,2,3) 28](#_heading=h.pbxlepngsw3i)

[**Cluster=64 29**](#_heading=)

[(Class 1,2) 29](#_heading=h.iiudhj51mgbp)

[(Class 1,3) 29](#_heading=h.3lvg866dj52w)

[(Class 2,3) 29](#_heading=h.jq6orok8bgp0)

[(Class 1,2,3) 30](#_heading=h.ak44r2viyb7d)

[**3.1.4 Contour Plots 31**](#_heading=)

[**3.1.5 Comparison GMM with Bayes Classifier 32**](#_heading=)

[**3.1.6 Inferences 33**](#_heading=)

[**3.2 Real World data (DATASET 2(a)) 34**](#_heading=h.h14rezq8s8ke)

[3.2.1 Decision Boundary 35](#_heading=h.cqv84ziccnid)

[Cluster =1 35](#_heading=h.8bu0gv7jeblu)

[Cluster=2 36](#_heading=h.e2js5nju226x)

[Cluster=4 37](#_heading=h.95ud04s13bg5)

[Cluster=8 38](#_heading=h.6uc1x75eh5b6)

[Cluster=16 39](#_heading=h.sa851hrryhw7)

[Cluster=32 40](#_heading=h.b5hswwtv7isl)

[Cluster=64 41](#_heading=h.z3p0f0se1otg)

[3.2.2 Log-likelihood vs Iterations plots 41](#_heading=h.s2eq6kt3l7mi)

[Cluster=1 41](#_heading=h.x3s6441qjduf)

[Cluster=2 42](#_heading=h.ofassg6s9xii)

[Cluster=4 42](#_heading=h.497g6vpohllb)

[Cluster =8 43](#_heading=h.pygpangnrj6s)

[Cluster=16 43](#_heading=h.vfbjoy18t2mh)

[Cluster=32 43](#_heading=h.kwt57iyeu9kk)

[Cluster =64 44](#_heading=h.9qq98yuk9vo4)

[3.2.3 Confusion Matrix, Precision, Recall and F-measure 44](#_heading=h.t808fm6lvewc)

[Cluster=1 44](#_heading=h.5pk9vdxva67i)

[(Class 1,2) 44](#_heading=h.hmx0v4nnvj9f)

[(Class 1,3) 44](#_heading=h.vc8lfpgr35sl)

[(Class 2,3) 45](#_heading=h.yp3a7hacyjiq)

[(Class 1,2,3) 45](#_heading=h.yagqte9nlzqu)

[Cluster=2 45](#_heading=h.dydjd8h54qll)

[(Class 1,2) 45](#_heading=h.xjuikcfms7dk)

[(Class 1,3) 46](#_heading=h.r92pihhy13c5)

[(Class 2,3) 46](#_heading=h.ewub5toraxxx)

[Table 44: Performance evaluation of class 2,3 for cluster 2 (Dataset-2a) 46](#_heading=h.9335fsl8uwuk)

[(Class 1,2,3) 46](#_heading=h.4w7dv6tm25lw)

[Cluster=4 47](#_heading=h.6f4401f9t6a)

[(Class 1,2) 47](#_heading=h.4zbdczsur1dn)

[(Class 1,3) 47](#_heading=h.wka3lbavij1v)

[(Class 2,3) 47](#_heading=h.ld1cdhr3fxs6)

[(Class 1,2,3) 48](#_heading=h.ibcyedj9ge40)

[Cluster=8 48](#_heading=h.oke7ytr5wao)

[(Class 1,2) 48](#_heading=h.bslrzkf1z4ta)

[(Class 1,3) 48](#_heading=h.8obbgpay9tit)

[(Class 2,3) 49](#_heading=h.dj3c24m7d2f8)

[(Class 1,2,3) 49](#_heading=h.wukedwp607k9)

[Cluster=16 49](#_heading=h.kipv7041yonw)

[(Class 1,2) 49](#_heading=h.v6t5htwiaan5)

[(Class 1,3) 50](#_heading=h.z1v7cfyrmxud)

[(Class 2,3) 51](#_heading=h.ukqlufg2umxv)

[(Class 1,2,3) 51](#_heading=h.ircp9usjdht8)

[Cluster=32 51](#_heading=h.ca5l7be32l8k)

[(Class 1,2) 51](#_heading=h.7ak0xoa1tx8d)

[(Class 1,3) 52](#_heading=h.ljv19ywm0xgl)

[(Class 2,3) 52](#_heading=h.ojv3w7t0vn0b)

[(Class 1,2,3) 52](#_heading=h.bi6oar8hjj7r)

[Cluster=64 52](#_heading=h.em3ntwylt3p4)

[(Class 1,2) 52](#_heading=h.6mn2simqognf)

[(Class 1,3) 53](#_heading=h.3od12zjn7oen)

[(Class 2,3) 53](#_heading=h.75tznf1xzgyu)

[(Class 1,2,3) 53](#_heading=h.swmldcf3iqyl)

[**3.2.4 Contour Plots 54**](#_heading=)

[3.2.5 Comparison GMM with Bayes ClassifierPerformance 55](#_heading=h.5132ez2vhsly)

[**3.2.6 Inferences 56**](#_heading=)

[**3.3 3-Class Scene Image (DATASET 2(b)) 57**](#_heading=h.oosiel7993dk)

[**3.3.1 32 -dimensional Bag of Visual Words Representation 58**](#_heading=)

[3.3.2 Color Coded Histogram 59](#_heading=h.thhex3xjkv4v)

[3.3.2.1 Log-likelihood vs Iterations plots 59](#_heading=h.r6viy31m0ci8)

[Cluster=1 59](#_heading=h.yxdtwxi6wjv9)

[Cluster=2 60](#_heading=h.mu1486z6qq8e)

[Cluster=4 60](#_heading=h.bnlxle9d8bzs)

[Cluster=8 60](#_heading=h.gbsqzrotojg)

[Cluster=16 60](#_heading=h.bv25sefftfib)

[3.3.2.2 Confusion Matrix, Precision, Recall and F-measure 61](#_heading=h.6x694mr5d9no)

[Cluster=1 61](#_heading=h.2tr2i9f7x26v)

[Cluster=2 61](#_heading=h.syiak6cod44i)

[Cluster=4 61](#_heading=h.h1oetoua1ff8)

[Cluster=8 62](#_heading=h.dznnobbj1ljj)

[Cluster=16 62](#_heading=h.e00d8s6vru22)

[**3.3.2.3:Inferences 62**](#_heading=)

[**3.3.3 Bag of Visual Words 63**](#_heading=)

[3.3.3.1 Log-likelihood vs Iterations plots 63](#_heading=h.ry0ms1wzzgrl)

[Cluster=1 63](#_heading=h.jo90505n7du1)

[Cluster=2 63](#_heading=h.mblye0chccn3)

[Cluster=4 63](#_heading=h.97svzaz5fsf1)

[Cluster=8 64](#_heading=h.259jalgsuprf)

[Cluster=16 64](#_heading=h.kkpha77jg56)

[Cluster=32 64](#_heading=h.egltpc3jb9co)

[Cluster=64 64](#_heading=h.uy2uwjnqwoqe)

[**3.3.3.2 Confusion Matrix, Precision, Recall and F-measure 65**](#_heading=)

[Cluster=1 65](#_heading=h.vqyh556lvj8)

[Cluster=2 65](#_heading=h.7xsqboxiajqq)

[Cluster=4 65](#_heading=h.mhp6ic2shzsa)

[Clusters=8 66](#_heading=h.naj8syujb4e)

[Clusters=16-64 66](#_heading=h.xs31huhfm7g4)

[**3.3.3.3: Inferences 67**](#_heading=)

[**3.4: Cervical cytology (cell): DATASET 2(c) 68**](#_heading=h.lcx0m2pufik9)

[**3.4.1: KMeans Clustering 68**](#_heading=)

[**3.4.2:Modified KMeans Clustering 70**](#_heading=)

[**3.4.3: Inferences 72**](#_heading=)

[**4: Conclusion and Observations 72**](#_heading=)

[**Bibliography 74**](#_heading=)

## 

# 1. Problem Statement

**Datasets:**

**Dataset 1: Nonlinearly separable classes**: 2-dimensional data of 2 or 3 classes that are non-linearly separable. (Same data used in Assignment 1.)

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

(a) Two-dimensional speech dataset (vowel data) used in

Assignment 1

(b) 3 class scene image datasets

(c) Cervical cytology (cell) image dataset

Data of each class is given separately. For Dataset 1 and Dataset 2(a), divide the data from each class into training, and test data. From each class, train, and test split should be 70% and 30% respectively. For Dataset 2(b) and Dataset 2(c), training and test sets are given.

**Note:** Each batch of students must use the datasets identified for that batch.

**1. Classification task:**

Build the Bayes classifier using GMM on Dataset-1, Dataset-2(a) and Dataset-2(b) Parameters of GMM are to be initialized using K-means clustering.

**Note:**

**i. Perform the experiments on different numbers of mixtures of GMM (For e.g. 1, 2, 4, 8, 16, 32, 64).**

**ii. Perform the experiments on Dataset 2(b) using a set of 24-dimensional color histogram feature vectors and 32-dimensional bag-of-visual-words (BoVW) feature vectors separately. Report the results for both the representations on different numbers of mixtures (For e.g. 1, 2, 4, 8, 16).**

**2. Segment the cell images by clustering the local feature vectors from cell image datasets into 3 groups using (a) K-means clustering and (b) Modified K-means clustering (using Mahalanobis distance).**

**Note:** Both the K-means clustering methods are initialized by the same initial centers.

**Objective:**

**1.** Classification accuracy, precision for every class, mean precision, recall for every class, mean recall, F-measure for every class and mean F-measure on test data (for each of the different parameters).

**2.** Confusion matrix based on the performance for test data (for the best GMM model).

**3.** Constant density contour plot for all the classes with the training data superposed **(only for Dataset-1** and **Dataset 2(a) on best model)**.

**4.** Decision regions plot with the training data superposed **(only for Dataset-1** and **Dataset 2(a) on the best model)**

**5.** Comparison with the results from the Assignment 1**.**

**6.** Result should also consist of a plot of 3 clusters on training data of **Dataset 2(c)** and the result of the cluster projected on test images (i.e., segmentation).

**7.** Report should also include the graph of ***iterations vs log likelihood*** for all the datasets with different numbers of components.

## 

# 2. Approach to finding a solution

**K means clustering:** K-means clustering is a type of unsupervised learning, which is used when you have unlabeled data (i.e., data without defined classes). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided. Data points are clustered based on feature similarity. The results of the K- means clustering algorithm are:

1. The centroids of the K clusters, which can be used to label new data

2. Labels for the training data (each data point is assigned to a single cluster (Hard Clustering))

The algorithm works as follows:

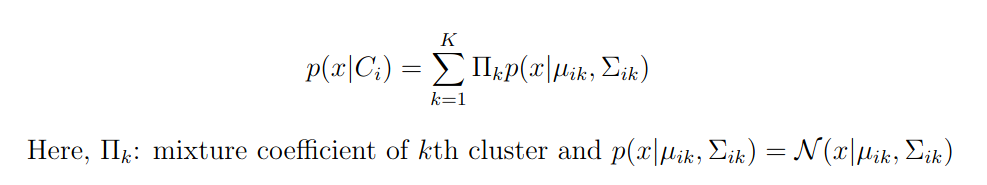
1. First, we initialize k points, called means, randomly.

2. We categorize each item to its closest mean and update the mean’s coordinates, which are the averages of the items categorized in that mean so far.

3. We repeat the process until convergence, and at the end, we have our clusters.

**Gaussian Mixture Model:** Gaussian mixture models (GMM) are often used for data clustering. Usually, fitted GMMs cluster by assigning query data points to the multivariate normal components that maximize the component posterior probability given the data. That is, given a fitted GMM, a cluster assigns query data to the component yielding the highest posterior probability. GMM clustering is more flexible because you can view it as a soft clustering method. Soft clustering methods assign a score to a data point for each cluster. The value of the score indicates the association strength of the data point to the cluster. When clustering with GMMs, the score is the posterior probability.

Let class Ci have K Gaussian clusters; then the probability of x given class Ci is,



***Features to be extracted from images of Dataset 2(b) and Dataset 2(c):***

**1. Features from images of Dataset 2(b):**

**1. Color histogram feature:**

• Consider 32 x 32 non overlapping patches on every image (from training and test sets). For example, if the image size is 256 x 256, there will be 32 numbers of 32 x 32 non overlapping patches.

• Extract 8-bin color histogram from every color channel (R, G and B) from a patch. It results in 3, 8-dimensional feature vectors. Concatenate them to form a 24-dimensional feature vector.

• Similarly extract 24-dimensional feature vectors from every patch.

• Stack the 24-dimensional feature vectors corresponding to every patch in an image and save them as a file in the corresponding class folder.

• Thus, an image is represented as set (collection) of 24-dimensional color histogram vectors representation

• Repeat the above steps to all the images in training and test sets of all the classes.

**Color histogram is computed as follows from a color channel:**

• When the given image is read, it will be read as a 3-dimensional matrix of pixel values. Each dimension is corresponding to a color channel. The pixel values in each color channel are in the range 0 to 255.

• For a color channel,

◦ Divide this range into 8 equal bins.

◦ Count the number of pixels falling into each bin. This results in a vector of 8 values.

◦ This is the 8-dimensional color histogram (from a color channel) feature vector. ◦ Normalise this vector by dividing it by the number of pixels in that image. • Do the same for other color channels. Concatenate those three 8-dimensional color histogram vectors to form a 24-dimensional vector.

**2. Bag-of-visual-words (BoVW) feature using K-means clustering:**

• Take the 24-dimensional color histogram feature vectors of all the training examples of all the classes.

• Group them into 32 clusters using K-means clustering algorithms.

• Now take an image, assign each 24-dimensional color histogram feature vector to a cluster.

• Count the number of feature vectors assigned to each of the 32 clusters. • This results in a 32-dimensional BoVW representation for that image.

• Normalize this vector by dividing it by the number of 24-dimensional histogram feature vectors in that image.

• Repeat this for every image in the training and test set.

**2. Features from images of Dataset 2(c):**

• Consider 7 x 7 overlapping patches with a shift of 1 pixel on every training cell image. • Compute mean and variance of intensities of pixels in the 7 x 7 patch.

• Thus a 7 x 7 patch is represented as a 2-dimensional feature vector.

• In the similar way compute 2-dimensional feature vectors from every patch from every training image.

To solve the given cases of classification and data analysis, we used K-Means Clustering and Gaussian Mixture Model methods. In the given datasets, we don’t know the incoming density distribution; hence, we cannot predict the classification of data points. This is a case of incomplete data problem where we don’t know the actual distribution of the incoming data and hence its parameters. Its parameters depend on each other, none of which we know beforehand.

To overcome this problem, we model our data as a combination of Gaussian Mixtures, which is quite a correct solution as most of the data in nature is a mixture of Gaussians. We now use the Expectation–Maximization type of algorithm. We cluster our data into K clusters using the K-Means clustering, initialized using a set of K randomly selected points (which may or may not belong to our sample data). We take the final mean computed by this algorithm to initialize our GMM model considering K Gaussian mixtures. Then we apply the EM algorithm for the GMM to obtain the parameters for the model. We terminate when the log-likelihood function plateaus. We obtain the mixture coefficients πk, responsibility terms γ(Znk), mean vectors, and the covariance matrices for nth data point for the kth cluster, for all the mixture components for all the data points. We obtain the GMM parameters for all the classes and use those to classify our points into the classes. For Dataset 1(b) and 2(a), we classify each point (represented as a 2-dimensional feature vector). For Dataset 2(b), we have represented each image as a collection of several 24-dimensional feature vectors. To classify an image, we first try to classify each of its feature vectors into a class. The class to which the maximum number of feature vectors were classified is taken to be the class the image should belong to. For this dataset, we also obtained the Bag of Visual Words representation (BOVW). For each image, the code-vectors (feature-vectors) obtained upon feature extraction from images are classified into one of the clusters (vocabulary) built from training data, and the frequency of code-vectors appearing in each cluster gives us the number of clusters - dimensional BOVW. For Dataset 2(c), we have represented each image as a collection of several 2-dimensional feature vectors (mean and dispersion of the concerned patch). We cluster these feature vectors into 3 clusters using K-means. Then we further sharpen our model by using a Gaussian Mixture Model of 3 mixtures initialized by the K-means algorithm. This helps us to decide which patch of the image belongs to which cluster. This segments the given images into three regions, viz., nucleus,the cell body, and the background.

# 3. DATASET

**Dataset 1: Nonlinearly separable classes**: 2-dimensional data of 2 or 3 classes that are non-linearly separable. (Same data used in Assignment 1.)

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

(a) Two-dimensional speech dataset (vowel data) used in

Assignment 1

(b) 3 class scene image datasets

(c) Cervical cytology (cell) image dataset

# 3.1 Non Linearly Separable data (Dataset-1)

Dataset 1 contains 2-dimensional non linearly separable data. There are 3 classes of data. The data is divided from each class into training, and test data. From each class, train, and test split is 70% and 30% respectively.

Table 1 gives the division of the dataset into Train and Test samples. Figure 1 shows the (a) All data points plotted for three different classes in Dataset 1.

| Class | Train Sample | Test Sample |
| --- | --- | --- |
| Class1 | 350 | 150 |
| Class2 | 350 | 150 |
| Class3 | 700 | 300 |

Table 1: Number of Training and Test Samples in Dataset-1

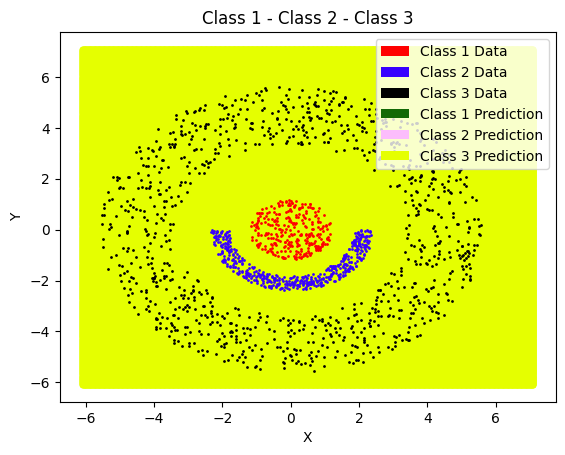


Figure 1:Visualization of Non-Linear Separable Dataset-1

## 3.1.1 Decision Boundary

### Cluster =1

|  |  |
| --- | --- |
|  |  |

Figure 2: Decision boundary plots of class 1,2, class1,3, class2,3 and class 1,2,3 respectively for cluster-1(Dataset-1)

### Cluster =2

|  |  |
| --- | --- |
|  |  |

Figure 3:Decision boundary plots of class 1,2, class1,3, class2,3 and class 1,2,3 respectively for cluster-2(Dataset-1)

### Cluster =4

|  |  |
| --- | --- |
|  |  |

Figure 4: Decision boundary plots of class 1,2, class1,3, class2,3 and class 1,2,3 respectively for cluster-4(Dataset-1)

### Cluster =8

|  |  |
| --- | --- |
|  |  |

Figure 5:Decision boundary plots of class 1,2, class1,3, class2,3 and class 1,2,3 respectively for cluster-8(Dataset-1)

### Cluster =16

|  |  |
| --- | --- |
|  |  |

Figure 6:Decision boundary plots of class 1,2, class1,3, class2,3 and class 1,2,3 respectively for cluster-16(Dataset-1)

### Cluster =32

|  |  |
| --- | --- |
|  |  |

Figure 7: Decision boundary plots of class 1,2, class1,3, class2,3 and class 1,2,3 respectively for cluster-32 (Dataset-1)

## 3.1.2 log-likelihood vs Iterations plots

### Cluster=1

|  |  |  |
| --- | --- | --- |

Figure 8:iteration vs. log-likelihood of class 1,2,3 for cluster 1 in color coded histogram (Dataset-2(a))

### Cluster=2

|  |  |  |
| --- | --- | --- |

Figure 9:iteration vs. log-likelihood of class 1,2,3 for cluster 2 in color coded histogram (Dataset-2(a))

### Cluster=4

|  |  |  |
| --- | --- | --- |

Figure 10:iteration vs. log-likelihood of class 1,2,3 for cluster 4 in color coded histogram(Dataset-2(a))

### Cluster=8

|  |  |  |
| --- | --- | --- |

Figure 11:iteration vs. log-likelihood of class 1,2,3 for cluster 8 in color coded histogram (Dataset-2(a))

### Cluster=16

|  |  |  |
| --- | --- | --- |

Figure 12:iteration vs. log-likelihood of class 1,2,3 for cluster 16 in color coded histogram (Dataset-2(a))

### Cluster=32

|  |  |  |
| --- | --- | --- |

Figure 13:iteration vs. log-likelihood of class 1,2,3 for cluster 32 in color coded histogram (Dataset-2(a))

## 3.1.3 Confusion Matrix, Precision, Recall and F-measure

### Cluster=1

#### (Class 1,2)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | | --- | --- | --- | | **Predicted Class1** | **64** | **26** | | **Predicted Class2** | **0** | **150** | | **Accuracy:** [0.892]  **Precision:** [1.0 0.852]  **Recall:** [0.711 1.0 ]  **F-Measure:** [0.831 0.920 ]  **Mean F-Measure:** 0.876  **Mean Precision:** 0.926  **Mean Recall:** 0.856 |

Table 2:Performance evaluation of class1,2 for cluster 1(Dataset-1)

#### (Class 1,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class3** | | --- | --- | --- | | **Predicted Class1** | **90** | **0** | | **Predicted Class3** | **0** | **300** | | **Accuracy:** [1.0 ]  **Precision:** [1.0 1.0]  **Recall:** [1.0 1.0]  **F-Measure:** [1.0 1.0]  **Mean F-Measure:** 0.999  **Mean Precision:** 0.999  **Mean Recall:** 0.999 |

Table 3:Performance evaluation of class 1,3 for cluster 1(Dataset-1)

#### (Class 2,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | | **Predicted Class2** | **130** | **20** | | **Predicted Class3** | **1** | **299** | | **Accuracy:** [0.953 ]  **Precision:** [0.992 0.937]  **Recall:** [0.867 0.997]  **F-Measure:** [0.925 0.966]  **Mean F-Measure:** 0.946  **Mean Precision:** 0.965  **Mean Recall:** 0.932 |

Table 4:Performance evaluation of class 2,3 for cluster 1(Dataset-1)

#### (Class 1,2,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | --- | | **Predicted Class1** | **64** | **26** | **0** | | **Predicted Class2** | **0** | **130** | **20** | | **Predicted Class3** | **1** | **10** | **299** | | **Accuracy:** [0.942 ]  **Precision:** [1.0 0.828 0.937]  **Recall:** [0.711 0.867 0.997]  **F-Measure:** [0.831 0.847 0.966]  **Mean F-Measure:** 0.881  **Mean Precision:** 0.922  **Mean Recall:** 0.858 |

Table 5:Performance evaluation of class 1,2,3 for cluster 1(Dataset-1)

### Cluster=2

#### (Class 1,2)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | | --- | --- | --- | | **Predicted Class1** | **90** | **0** | | **Predicted Class2** | **0** | **150** | | **Accuracy:** [1.0 ]  **Precision:** [1.0 1.0]  **Recall:** [1.0 1.0]  **F-Measure:** [1.0 1.0]  **Mean F-Measure:** 0.999  **Mean Precision:** 0.999  **Mean Recall:** 0.999 |

Table 6:Performance evaluation of class 1,2 for cluster 2(Dataset-1)

#### (Class 1,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class3** | | --- | --- | --- | | **Predicted Class1** | **90** | **0** | | **Predicted Class3** | **0** | **300** | | **Accuracy:** [1.0]  **Precision:** [1.0 1.0]  **Recall:** [1.0 1.0]  **F-Measure:** [1.0 1.0]  **Mean F-Measure:** 0.999  **Mean Precision:** 0.999  **Mean Recall:** 0.999 |

Table 7:Performance evaluation of class 1,3 for cluster 2(Dataset-1)

#### (Class 2,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | | **Predicted Class2** | **150** | **0** | | **Predicted Class3** | **0** | **300** | | **Accuracy:** [1.0]  **Precision:** [1.0 1.0]  **Recall:** [1.0 1.0]  **F-Measure:** [1.0 1.0]  **Mean F-Measure:** 0.999  **Mean Precision:** 0.999  **Mean Recall:** 0.999 |

Table 8:Performance evaluation of class 2,3 for cluster 2(Dataset-1)

#### (Class 1,2,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | --- | | **Predicted Class1** | **90** | **0** | **0** | | **Predicted Class2** | **0** | **150** | **0** | | **Predicted Class3** | **0** | **0** | **300** | | **Accuracy:** [1.0]  **Precision:** [1.0 1.0 1.0]  **Recall:** [1.0 1.0 1.0]  **F-Measure:** [1.0 1.0 1.0]  **Mean F-Measure:** 0.999  **Mean Precision:** 0.999  **Mean Recall:** 0.999 |

Table 9:Performance evaluation of class 1,2,3 for cluster 2(Dataset-1)

### Cluster=4

#### (Class 1,2)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | | --- | --- | --- | | **Predicted Class1** | **90** | **0** | | **Predicted Class2** | **0** | **150** | | **Accuracy**: [1.0]  **Precision**: [1.0 1.0]  **Recall**: [1.0 1.0]  **F-Measure**: [1.0 1.0]  **Mean F-Measure**: 0.999  **Mean Precision**: 0.999  **Mean Recall**: 0.999 |

Table 10:Performance evaluation of class 1,2 for cluster 4(Dataset-1)

#### 

#### (Class 1,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class3** | | --- | --- | --- | | **Predicted Class1** | **90** | **0** | | **Predicted Class3** | **0** | **300** | | **Accuracy**: [1.0]  **Precision**: [1.0 1.0]  **Recall**: [1.0 1.0]  **F-Measure**: [1.0 1.0]  **Mean F-Measure**: 0.999  **Mean Precision**: 0.999  **Mean Recall**: 0.999 |

Table 11:Performance evaluation of class 1,3 for cluster 4(Dataset-1)

#### (Class 2,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | | **Predicted Class2** | **150** | **0** | | **Predicted Class3** | **0** | **300** | | **Accuracy**: [1.0]  **Precision**: [1.0 1.0]  **Recall**: [1.0 1.0]  **F-Measure**: [1.0 1.0]  **Mean F-Measure**: 0.999  **Mean Precision**: 0.999  **Mean Recall**: 0.999 |

Table 12:Performance evaluation of class 2,3 for cluster 4(Dataset-1)

#### (Class 1,2,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | --- | | **Predicted Class1** | **90** | **0** | **0** | | **Predicted Class2** | **0** | **150** | **0** | | **Predicted Class3** | **0** | **0** | **300** | | **Accuracy**: [1.0]  **Precision**: [1.0 1.0 1.0]  **Recall**: [1.0 1.0 1.0]  **F-Measure**: [1.0 1.0 1.0]  **Mean F-Measure**: 0.999  **Mean Precision**: 0.999  **Mean Recall**: 0.999 |

Table 13:Performance evaluation of class 1,2,3 for cluster 4(Dataset-1)

### Cluster=8

#### (Class 1,2)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | | --- | --- | --- | | **Predicted Class1** | **90** | **0** | | **Predicted Class2** | **0** | **150** | | **Accuracy**: [1.0]  **Precision**: [1.0 1.0]  **Recall**: [1.0 1.0]  **F-Measure**: [1.0 1.0]  **Mean F-Measure**: 0.999  **Mean Precision**: 0.999  **Mean Recall**: 0.999 |

Table 14:Performance evaluation of class1,2 for cluster 8(Dataset-1)

#### (Class 1,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class3** | | --- | --- | --- | | **Predicted Class1** | **90** | **0** | | **Predicted Class3** | **0** | **300** | | **Accuracy**: [1.0]  **Precision**: [1.0 1.0]  **Recall**: [1.0 1.0]  **F-Measure**: [1.0 1.0]  **Mean F-Measure**: 0.999  **Mean Precision**: 0.999  **Mean Recall**: 0.999 |

Table 15:Performance evaluation of class 1,3 for cluster 8(Dataset-1)

#### (Class 2,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | | **Predicted Class2** | **150** | **0** | | **Predicted Class3** | **0** | **300** | | **Accuracy**: [1.0]  **Precision**: [1.0 1.0]  **Recall**: [1.0 1.0]  **F-Measure**: [1.0 1.0]  **Mean F-Measure**: 0.999  **Mean Precision**: 0.999  **Mean Recall**: 0.999 |

Table 16:Performance evaluation of class 2,3 for cluster 8(Dataset-1)

#### 

#### 

#### (Class 1,2,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | --- | | **Predicted Class1** | **90** | **0** | **0** | | **Predicted Class2** | **0** | **150** | **0** | | **Predicted Class3** | **0** | **0** | **300** | | **Accuracy**: [1.0]  **Precision**: [1.0 1.0 1.0]  **Recall**: [1.0 1.0 1.0]  **F-Measure**: [1.0 1.0 1.0]  **Mean F-Measure**: 0.999  **Mean Precision**: 0.999  **Mean Recall**: 0.999 |

Table 17:Performance evaluation of class 1,2,3 for cluster 8(Dataset-1)

### Cluster=16

#### (Class 1,2)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | | --- | --- | --- | | **Predicted Class1** | **90** | **0** | | **Predicted Class2** | **0** | **150** | | **Accuracy**: [1.0]  **Precision**: [1.0 1.0]  **Recall**: [1.0 1.0]  **F-Measure**: [1.0 1.0]  **Mean F-Measure**: 0.999  **Mean Precision**: 0.999  **Mean Recall**: 0.999 |

Table 18:Performance evaluation of class 1,2 for cluster 16 (Dataset-1)

#### (Class 1,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class3** | | --- | --- | --- | | **Predicted Class1** | **90** | **0** | | **Predicted Class3** | **0** | **300** | | **Accuracy**: [1.0]  **Precision**: [1.0 1.0]  **Recall**: [1.0 1.0]  **F-Measure**: [1.0 1.0]  **Mean F-Measure**: 0.999  **Mean Precision**: 0.999  **Mean Recall**: 0.999 |

Table 19:Performance evaluation of class 1,3 for cluster 16 (Dataset-1)

#### (Class 2,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | | **Predicted Class2** | **150** | **0** | | **Predicted Class3** | **0** | **350** | | **Accuracy**: [1.0]  **Precision**: [1.0 1.0]  **Recall**: [1.0 1.0]  **F-Measure**: [1.0 1.0]  **Mean F-Measure**: 0.999  **Mean Precision**: 0.999  **Mean Recall**: 0.999 |

Table 20:Performance evaluation of class 2,3 for cluster 16 (Dataset-1)

#### (Class 1,2,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | --- | | **Predicted Class1** | **90** | **0** | **0** | | **Predicted Class2** | **0** | **150** | **0** | | **Predicted Class3** | **0** | **0** | **300** | | **Accuracy**: [1.0]  **Precision**: [1.0 1.0 1.0]  **Recall**: [1.0 1.0 1.0]  **F-Measure**: [1.0 1.0 1.0]  **Mean F-Measure**: 0.999  **Mean Precision**: 0.999  **Mean Recall**: 0.999 |

Table 21:Performance evaluation of class 1,2,3 for cluster 16 (Dataset-1)

### Cluster=32

#### (Class 1,2)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | | --- | --- | --- | | **Predicted Class1** | **90** | **0** | | **Predicted Class2** | **0** | **150** | | **Accuracy**: [1.0]  **Precision**: [1.0 1.0]  **Recall**: [1.0 1.0]  **F-Measure**: [1.0 1.0]  **Mean F-Measure**: 0.999  **Mean Precision**: 0.999  **Mean Recall**: 0.999 |

Table 22:Performance evaluation of class 1,2 for cluster 32(Dataset-1)

#### (Class 1,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class3** | | --- | --- | --- | | **Predicted Class1** | **90** | **0** | | **Predicted Class3** | **0** | **300** | | **Accuracy**: [1.0]  **Precision**: [1.0 1.0]  **Recall**: [1.0 1.0]  **F-Measure**: [1.0 1.0]  **Mean F-Measure**: 0.999  **Mean Precision**: 0.999  **Mean Recall**: 0.999 |

Table 23:Performance evaluation of class 1,3 for cluster 32(Dataset-1)

#### (Class 2,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | | **Predicted Class2** | **150** | **0** | | **Predicted Class3** | **0** | **300** | | **Accuracy**: [1.0]  **Precision**: [1.0 1.0]  **Recall**: [1.0 1.0]  **F-Measure**: [1.0 1.0]  **Mean F-Measure**: 0.999  **Mean Precision**: 0.999  **Mean Recall**: 0.999 |

Table 24:Performance evaluation of class 2,3 for cluster 32(Dataset-1)

#### (Class 1,2,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | --- | | **Predicted Class1** | **90** | **0** | **0** | | **Predicted Class2** | **0** | **150** | **0** | | **Predicted Class3** | **0** | **0** | **300** | | **Accuracy**: [1.0]  **Precision**: [1.0 1.0 1.0]  **Recall**: [1.0 1.0 1.0]  **F-Measure**: [1.0 1.0 1.0]  **Mean F-Measure**: 0.999  **Mean Precision**: 0.999  **Mean Recall**: 0.999 |

Table 25:Performance evaluation of class 1,2,3 for cluster 32(Dataset-1)

### Cluster=64

#### (Class 1,2)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | | --- | --- | --- | | **Predicted Class1** | **90** | **0** | | **Predicted Class2** | **0** | **150** | | **Accuracy**: [0.375]  **Precision**: [0.375 0.5]  **Recall**: [1.0 6.66666667e-13]  **F-Measure**: [5.45454545e-01 1.33333333e-12] **Mean F-Measure**: 0.272  **Mean Precision**: 0.437  **Mean Recall**: 0.499 |

Table 26:Performance evaluation of class 1,2 for cluster 64 (Dataset-1)

#### (Class 1,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class3** | | --- | --- | --- | | **Predicted Class1** | **90** | **0** | | **Predicted Class3** | **0** | **300** | | **Accuracy**: [0.231]  **Precision**: [0.231 0.5 ]  **Recall**: [1.00000000e+00 3.33333333e-13] **F-Measure**: [3.75000000e-01 6.66666667e-13] **Mean F-Measure**: 0.187  **Mean Precision**: 0.365  **Mean Recall**: 0.499 |

Table 27:Performance evaluation of class 1,3 for cluster 64 (Dataset-1)

#### (Class 2,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | | **Predicted Class2** | **150** | **0** | | **Predicted Class3** | **0** | **300** | | **Accuracy**: [0.333]  **Precision**: [0.333 0.5 ]  **Recall**: [1.00000000e+00 3.33333333e-13] **F-Measure**: [5.00000000e-01 6.66666667e-13] **Mean F-Measure**: 0.250  **Mean Precision**: 0.417  **Mean Recall**: 0.499 |

Table 28:Performance evaluation of class 2,3 for cluster 64 (Dataset-1)

#### 

#### 

#### (Class 1,2,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | --- | | **Predicted Class1** | **90** | **0** | **0** | | **Predicted Class2** | **0** | **150** | **0** | | **Predicted Class3** | **0** | **0** | **300** | | **Accuracy**: [0.444]  **Precision**: [0.167 0.333 0.333]  **Recall**: [1.00000000e+00 6.66666667e-13 3.33333333e-13]  **F-Measure**: [2.85714286e-01 1.33333333e-12 6.66666667e-13]  **Mean F-Measure**: 0.095  **Mean Precision**: 0.278  **Mean Recall**: 0.333 |

Table 29:Performance evaluation of class 1,2,3 for cluster 64 (Dataset-1)

## 3.1.4 Contour Plots

| NLSk12  Cluster=1 | NLSk22  Cluster=2 |
| --- | --- |
| NLSk42  Cluster=4 | NLSk82Cluster=8 |
| NLSk162  Cluster=16 | NLSk322  Cluster=32 |

Figure 14: contour plots of cluster 1,2,4,8,16,32 respectively of Dataset-1

## 3.1.5 Comparison GMM with Bayes Classifier

| **Confusion Matrix** | **Precision, Recall and F-measure**  **(Cluster-wise)** |
| --- | --- |
| | **Case1** | **Accuracy**: 57.7%  **Mean Precision**: 0.373  **Mean Recall**: 0.435  **Mean F-Measure**: 0.357 | | --- | --- | | **Case2** | **Accuracy**: 57.9%  **Mean Precision**: 0.376  **Mean Recall**: 0.439  **Mean F-Measure**: 0.360 | | **Case3** | **Accuracy**: 96.7%  **Mean Precision**: 0.952  **Mean Recall**:0.931  **Mean F-Measure**: 0.939 | | **Case4** | **Accuracy**: 96.88%  **Mean** **Precision**: 0.952  **Mean Recall**: 0.936  **Mean F-Measure**: 0.92 | | | **Cluster1** | **Accuracy:** [0.942 ]  **Precision:** [1.0 0.828 0.937]  **Recall:** [0.711 0.867 0.997]  **F-Measure:** [0.831 0.847 0.966]  **Mean F-Measure:** 0.881  **Mean Precision:** 0.922  **Mean Recall:** 0.858 | | --- | --- | | **Cluster2** | **Accuracy:** [1.0]  **Precision:** [1.0 1.0 1.0]  **Recall:** [1.0 1.0 1.0]  **F-Measure:** [1.0 1.0 1.0]  **Mean F-Measure:** 0.999  **Mean Precision:** 0.999  **Mean Recall:** 0.999 | | **Cluster4** | **Accuracy**: [1.0]  **Precision**: [1.0 1.0 1.0]  **Recall**: [1.0 1.0 1.0]  **F-Measure**: [1.0 1.0 1.0]  **Mean F-Measure**: 0.999  **Mean Precision**: 0.999  **Mean Recall**: 0.999 | | **Cluster8** | **Accuracy**: [1.0]  **Precision**: [1.0 1.0 1.0]  **Recall**: [1.0 1.0 1.0]  **F-Measure**: [1.0 1.0 1.0]  **Mean F-Measure**: 0.999  **Mean Precision**: 0.999  **Mean Recall**: 0.999 | | **Cluster16** | **Accuracy**: [1.0]  **Precision**: [1.0 1.0 1.0]  **Recall**: [1.0 1.0 1.0]  **F-Measure**: [1.0 1.0 1.0]  **Mean F-Measure**: 0.999  **Mean Precision**: 0.999  **Mean Recall**: 0.999 | | **Cluster32** | **Accuracy**: [1.0]  **Precision**: [1.0 1.0 1.0]  **Recall**: [1.0 1.0 1.0]  **F-Measure**: [1.0 1.0 1.0]  **Mean F-Measure**: 0.999  **Mean Precision**: 0.999  **Mean Recall**: 0.999 | | **Cluster64** | **Accuracy**: [0.444]  **Precision**: [0.167 0.333 0.333]  **Recall**: [1.00000000e+00 6.66666667e-13 3.33333333e-13]  **F-Measure**: [2.85714286e-01 1.33333333e-12 6.66666667e-13]  **Mean F-Measure**: 0.095  **Mean Precision**: 0.278  **Mean Recall**: 0.333 | |

Table 30: Comparison GMM with base classifier taking 4 case of class 1,2,3(Dataset-1)

## 3.1.6 Inferences

* Here as the number of clusters increases we can observe that accuracy increases. Here, interestingly, in case of 32 clusters case we have 32(number of clusters) \*(24(Mean-vector) + 24(diagonal covariance matrix)) free variables, which are more than total training samples we were having, but we do not observe any kind of over-fitting. This might be due to the fact that there was not much difference between training data and test data.
* Here for the case of 1 cluster in GMM, we are observing almost the same accuracy as observed for case3 and 4 of Bayes’, which should be the case as single cluster Gaussian means single Gaussian and also in classifier 3 we have taken a diagonal matrix. As, in GMM it doesn’t matter much if we take a diagonal or full covariance matrix, so our single cluster GMM case is also comparable with Bayes’ case-4 result.
* We can say that GMM gives a more complex and better estimate of decision boundary than Bayes’ classifier which improves classification accuracy, while using GMM.
* In both Bayes classifier 3 and GMM we have taken the covariance matrix to be diagonal and also the nature of the boundary is non-linear. But the horse-shoe type distribution of data is best fitted by multivariate Gaussian mixture model because multiple Gaussians can take the shape of horse-shoe type distribution.

# 3.2 Real World data (DATASET 2(a))

Dataset 2(a) contains 2-dimensional non linearly separable data. There are 3 classes of data. The data is divided from each class into training, and test data. From each class, train, and test split is 70% and 30% respectively.

Table gives the division of the dataset into Train and Test samples. Figure 15 shows the (a) All data points plotted for three different classes in Dataset 2(a).

| Class | Train Sample | Test Sample |
| --- | --- | --- |
| Class1 | 1671 | 717 |
| Class2 | 1741 | 747 |
| Class3 | 1514 | 650 |

Table 31:Number of Training and Test Samples in Dataset2(a)

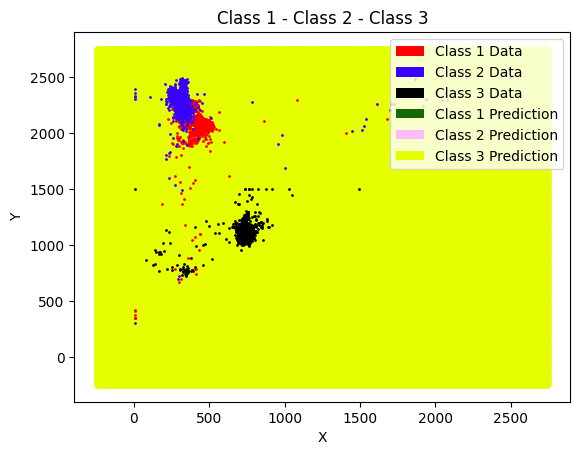


Figure 15:Visualization of Real World Dataset2(a)

## 3.2.1 Decision Boundary

### Cluster =1

| IMG_256 | IMG_256 |
| --- | --- |
| IMG_256 | IMG_256 |

Figure 16:Decision boundary plots of class 1,2, class1,3, class2,3 and class 1,2,3 respectively for cluster-1(Dataset-2(a))

### Cluster=2

| IMG_256 | IMG_256 |
| --- | --- |
| IMG_256 | IMG_256 |

Figure17:Decision boundary plots of class 1,2, class1,3, class2,3 and class 1,2,3 respectively for cluster-2(Dataset-2(a))

### 

### Cluster=4

| IMG_256 | IMG_256 |
| --- | --- |
| IMG_256 | IMG_256 |

Figure 18:Decision boundary plots of class 1,2, class1,3, class2,3 and class 1,2,3 respectively for cluster-4(Dataset-2(a))

### 

### Cluster=8

| IMG_256 | IMG_256 |
| --- | --- |
| IMG_256 | IMG_256 |

Figure19: Decision boundary plots of class 1,2, class1,3, class2,3 and class 1,2,3 respectively for cluster-8(Dataset-2(a))

### 

### 

### Cluster=16

| IMG_256 | IMG_256 |
| --- | --- |
| IMG_256 | IMG_256 |

Figure20: Decision boundary plots of class 1,2, class1,3, class2,3 and class 1,2,3 respectively for cluster-16(Dataset-2(a))

### 

### Cluster=32

| IMG_256 | IMG_256 |
| --- | --- |
| IMG_256 | IMG_256 |

Figure 21:Decision boundary plots of class 1,2, class1,3, class2,3 and class 1,2,3 respectively for cluster-32(Dataset-2(a))

### 

### Cluster=64

| IMG_256 | IMG_256 |
| --- | --- |
| IMG_256 | IMG_256 |

Figure 22: Decision boundary plots of class 1,2, class1,3, class2,3 and class 1,2,3 respectively for cluster-64(Dataset-2(a))

## 3.2.2 Log-likelihood vs Iterations plots

#### Cluster=1

| IMG_256 | IMG_256 | IMG_256 |
| --- | --- | --- |

Table 31 :Iteration vs. log-likelihood of class 1,2,3 for cluster 1 in color coded histogram (Dataset-2(a))

#### Cluster=2

| IMG_256 | IMG_256 | IMG_256 |
| --- | --- | --- |

Table 32: iteration vs. log-likelihood of class 1,2,3 for cluster 2 in color coded histogram (Dataset-2(a))

#### Cluster=4

| IMG_256 | IMG_256 | IMG_256 |
| --- | --- | --- |

Table 33: iteration vs. log-likelihood of class 1,2,3 for cluster 4 in color coded histogram (Dataset-2(a))

#### Cluster =8

| IMG_256 | IMG_256 | IMG_256 |
| --- | --- | --- |

Table 34:iteration vs. log-likelihood of class 1,2,3 for cluster 8 in color coded histogram (Dataset-2(a))

#### 

#### Cluster=16

| IMG_256 | IMG_256 | IMG_256 |
| --- | --- | --- |

Table 35: iteration vs. log-likelihood of class 1,2,3 for cluster 16 in color coded histogram (Dataset-2(a))

#### Cluster=32

| IMG_256 | IMG_256 | IMG_256 |
| --- | --- | --- |

Table 36: iteration vs. log-likelihood of class 1,2,3 for cluster 32 in color coded histogram (Dataset-2(a))

#### Cluster =64

| IMG_256 | IMG_256 | IMG_256 |
| --- | --- | --- |

Table 37: iteration vs. log-likelihood of class 1,2,3 for cluster 64 in color coded histogram (Dataset-2(a))

## 3.2.3 Confusion Matrix, Precision, Recall and F-measure

#### Cluster=1

##### (Class 1,2)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | | --- | --- | --- | | **Predicted Class1** | **340** | **348** | | **Predicted Class2** | **26** | **691** | | **Accuracy:**  [0.733]  **Precision:** [0.929 0.665]  **Recall:** [0.494 0.964 ]  **F-Measure:** [0.645 0.787]  **Mean F-Measure:** 0.716  **Mean Precision:** 0.797  **Mean Recall:** 0.728 |

Table 38 :Performance evaluation of class 1,2 for cluster 1 (Dataset-2a)

##### (Class 1,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class3** | | --- | --- | --- | | **Predicted Class1** | **670** | **18** | | **Predicted Class3** | **1** | **649** | | **Accuracy:** [0.986]  **Precision:** [0.998 0.973]  **Recall:** [0.974 0.998]  **F-Measure:** [0.986 0.986]  **Mean F-Measure:**  0.986  **Mean Precision:** 0.986  **Mean Recall:** 0.986 |

Table 39: Performance evaluation of class 1,3 for cluster 1 (Dataset-2a)

##### (Class 2,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | | **Predicted Class2** | **702** | **15** | | **Predicted Class3** | **1** | **649** | | **Accuracy:** [0.988]  **Precision:**  [0.998 0.977]  **Recall:** [0.979 0.998]  **F-Measure:** [0.988 0.988]  **Mean F-Measure:** 0.988  **Mean Precision:**  0.988  **Mean Recall:** 0.989 |

Performance evaluation of class 2,3 for cluster 1 (Dataset-2a)

##### (Class 1,2,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | --- | | **Predicted Class1** | **322** | **348** | **18** | | **Predicted Class2** | **13** | **691** | **13** | | **Predicted Class3** | **1** | **0** | **649** | | **Accuracy:** [0.872]  **Precision:** [0.958 0.665 0.954]  **Recall:** [0.468 0.964 0.998]  **F-Measure:**  [0.628 0.787 0.976]  **Mean F-Measure:** 0.797  **Mean Precision:** 0.859  **Mean Recall:** 0.810 |

Table 40: Performance evaluation of class 1,2,3 for cluster 1 (Dataset-2a)

#### Cluster=2

##### (Class 1,2)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | | --- | --- | --- | | **Predicted Class1** | **403** | **285** | | **Predicted Class2** | **21** | **696** | | **Accuracy**: [0.782]  **Precision**: [0.950 0.709]  **Recall**: [0.586 0.971]  **F-Measure**: [0.725]  **Mean F-Measure**: 0.772  **Mean Precision**: 0.829  **Mean Recall**: 0.778 |

Table 41: Performance evaluation of class 1,2 for cluster 2 (Dataset-2a)

##### 

##### (Class 1,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class3** | | --- | --- | --- | | **Predicted Class1** | **680** | **8** | | **Predicted Class3** | **23** | **627** | | **Accuracy**: [0.976]  **Precision**: [0.967 0.987]  **Recall**: [0.988 0.964]  **F-Measure**: [0.978 0.976]  **Mean F-Measure**: 0.977  **Mean Precision**: 0.977  **Mean Recall**: 0.976 |

Table 42:Performance evaluation of class 1,3 for cluster 2 (Dataset-2a)

##### (Class 2,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | | **Predicted Class2** | **715** | **2** | | **Predicted Class3** | **24** | **626** | | **Accuracy**: [0.981]  **Precision**: [0.967 0.997]  **Recall**: [0.997 0.963]  **F-Measure**: [0.982 0.979]  **Mean F-Measure**: 0.981  **Mean Precision**: 0.982  **Mean Recall**: 0.980 |

##### Table 44: Performance evaluation of class 2,3 for cluster 2 (Dataset-2a)

##### (Class 1,2,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | --- | | **Predicted Class1** | **395** | **285** | **8** | | **Predicted Class2** | **19** | **696** | **2** | | **Predicted Class3** | **21** | **3** | **626** | | **Accuracy**:[0.890]  **Precision**:[0.908 0.707 0.984]  **Recall**:[0.574 0.9707 0.963]  **F-Measure**:[0.70 0.818 0.973]  **Mean F-Measure**: 0.832  **Mean Precision**: 0.866  **Mean Recall**: 0.836 |

Table 45:Performance evaluation of class 1,2,3 for cluster 2 (Dataset-2a)

#### Cluster=4

##### (Class 1,2)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | | --- | --- | --- | | **Predicted Class1** | **520** | **168** | | **Predicted Class2** | **22** | **695** | | **Accuracy**: [0.865]  **Precision**: [0.959 0.805]  **Recall**: [0.756 0.969]  **F-Measure**: [0.846 0.879]  **Mean F-Measure**: 0.863  **Mean Precision**: 0.882  **Mean Recall**: 0.863 |

Table 46 :Performance evaluation of class 1,2 for cluster 4 (Dataset-2a)

##### (Class 1,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class3** | | --- | --- | --- | | **Predicted Class1** | **673** | **15** | | **Predicted Class3** | **2** | **648** | | **Accuracy**: [0.987]  **Precision**: [0.997 0.977]  **Recall**: [0.978 0.997]  **F-Measure**: [0.987 0.987]  **Mean F-Measure**: 0.987  **Mean Precision**: 0.987  **Mean Recall**: 0.988 |

Table 47 :Performance evaluation of class 1,3 for cluster 4 (Dataset-2a)

##### 

##### (Class 2,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | | **Predicted Class2** | **706** | **11** | | **Predicted Class3** | **1** | **649** | | **Accuracy**: [0.991]  **Precision**: [0.998 0.983]  **Recall**: [0.985 0.998]  **F-Measure**: [0.992 0.991]  **Mean F-Measure**: 0.991  **Mean Precision**: 0.991  **Mean Recall**: 0.992 |

Table 48: Performance evaluation of class 2,3 for cluster 4 (Dataset-2a)

##### (Class 1,2,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | --- | | **Predicted Class1** | **505** | **168** | **15** | | **Predicted Class2** | **19** | **695** | **3** | | **Predicted Class3** | **2** | **0** | **648** | | **Accuracy**: [0.933]  **Precision**: [0.960 0.805 0.973]  **Recall**: [0.734 0.969 0.997]  **F-Measure**: [0.832 0.879 0.985]  **Mean F-Measure**: 0.899  **Mean Precision**: 0.913  **Mean Recall**: 0.900 |

Table 49: Performance evaluation of class 1,2,3 for cluster 4 (Dataset-2a)

#### Cluster=8

##### (Class 1,2)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | | --- | --- | --- | | **Predicted Class1** | **481** | **207** | | **Predicted Class2** | **22** | **695** | | **Accuracy**: [0.837]  **Precision**: [0.956 0.770]  **Recall**: [0.699 0.969]  **F-Measure**: [0.808 0.859]  **Mean F-Measure**: 0.833  **Mean Precision**: 0.863  **Mean Recall**: 0.834 |

Table 51:Performance evaluation of class 1,2 for cluster 8 (Dataset-2a)

##### 

##### (Class 1,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class3** | | --- | --- | --- | | **Predicted Class1** | **678** | **10** | | **Predicted Class3** | **6** | **644** | | **Accuracy**: [0.988]  **Precision**: [0.991 0.985]  **Recall**: [0.985 0.991]  **F-Measure**: [0.988 0.988]  **Mean F-Measure**: 0.9880341268846131  **Mean Precision**: 0.9879687751487354  **Mean Recall**: 0.9881171735240042 |

Table 52:Performance evaluation of class 1,3 for cluster 8 (Dataset-2a)

##### (Class 2,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | | **Predicted Class2** | **709** | **8** | | **Predicted Class3** | **1** | **649** | | **Accuracy**: [0.993]  **Precision**: [0.998 0.988]  **Recall**: [0.989 0.998]  **F-Measure**: [0.994 0.993]  **Mean F-Measure**: 0.993  **Mean Precision**: 0.993  **Mean Recall**: 0.994 |

Table 53: Performance evaluation of class 2,3 for cluster 8 (Dataset-2a)

##### (Class 1,2,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | --- | | **Predicted Class1** | **472** | **207** | **9** | | **Predicted Class2** | **19** | **695** | **3** | | **Predicted Class3** | **6** | **1** | **643** | | **Accuracy**: [0.921]  **Precision**: [0.949 0.769 0.982]  **Recall**: [0.686 0.969 0.989]  **F-Measure**: [0.797 0.858 0.985]  **Mean F-Measure**: 0.880  **Mean Precision**: 0.900  **Mean Recall**: 0.881 |

Table 54:Performance evaluation of class 1,2,3 for cluster 8 (Dataset-2a)

#### 

#### Cluster=16

##### (Class 1,2)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | | --- | --- | --- | | **Predicted Class1** | **479** | **209** | | **Predicted Class2** | **23** | **694** | | **Accuracy**: [0.835]  **Precision**: [0.954 0.768]  **Recall**: [0.696 0.968 ]  **F-Measure**: [0.805 0.857]  **Mean F-Measure**: 0.831  **Mean Precision**: 0.861  **Mean Recall**: 0.832 |

Table 55:Performance evaluation of class 1,2 for cluster 16 (Dataset-2a)

##### (Class 1,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class3** | | --- | --- | --- | | **Predicted Class1** | **678** | **10** | | **Predicted Class3** | **5** | **645** | | **Accuracy**: [0.989]  **Precision**: [0.993 0.986]  **Recall**: [0.985 0.992]  **F-Measure**: [0.989 0.988]  **Mean F-Measure**: 0.989  **Mean Precision**: 0.989  **Mean Recall**: 0.989 |

Table 56: Performance evaluation of class 1,3 for cluster 16 (Dataset-2a)

##### (Class 2,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | | **Predicted Class2** | **700** | **17** | | **Predicted Class3** | **1** | **649** | | **Accuracy**: [0.987]  **Precision**: [0.999 0.974]  **Recall**: [0.976 0.998]  **F-Measure**: [0.987 0.986]  **Mean F-Measure**: 0.987  **Mean Precision**: 0.986  **Mean Recall**: 0.987 |

Table 57:Table Performance evaluation of class 2,3 for cluster 16 (Dataset-2a)

##### 

##### (Class 1,2,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | --- | | **Predicted Class1** | **470** | **209** | **9** | | **Predicted Class2** | **16** | **694** | **7** | | **Predicted Class3** | **5** | **1** | **644** | | **Accuracy:** [0.92]  **Precision:** [0.957 0.768 0.976]  **Recall:** [0.683 0.968 0.991]  **F-Measure:** [0.797 0.856 0.983]  **Mean F-Measure:** 0.879  **Mean Precision:** 0.900  **Mean Recall:** 0.881 |

Table 58: Performance evaluation of class 1,2,3 for cluster 16 (Dataset-2a)

#### Cluster=32

##### (Class 1,2)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | | --- | --- | --- | | **Predicted Class1** | **472** | **216** | | **Predicted Class2** | **25** | **692** | | **Accuracy:** [0.828]  **Precision:** [0.949 0.762]  **Recall:** [0.686 0.965 ]  **F-Measure:** [0.797 0.852]  **Mean F-Measure:** 0.824  **Mean Precision:** 0.856  **Mean Recall:** 0.826 |

Table 59:Performance evaluation of class 1,2 for cluster 32 (Dataset-2a)

##### (Class 1,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class3** | | --- | --- | --- | | **Predicted Class1** | **677** | **11** | | **Predicted Class3** | **3** | **647** | | **Accuracy:**  [0.989]  **Precision:** [0.996 0.983]  **Recall:** [0.984 0.995]  **F-Measure:** [0.989 0.989]  **Mean F-Measure:** 0.989  **Mean Precision:** 0.989  **Mean Recall:** 0.989 |

Table 60:Performance evaluation of class 1,3 for cluster 32 (Dataset-2a)

##### (Class 2,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | | **Predicted Class2** | **700** | **17** | | **Predicted Class3** | **0** | **650** | | **Accuracy:** [0.988]  **Precision:** [1.0 0.974]  **Recall:** [0.976 1. 0 ]  **F-Measure:**  [0.988 0.987]  **Mean F-Measure:** 0.987  **Mean Precision:** 0.987 |

Table 61: Performance evaluation of class 2,3 for cluster 32 (Dataset-2a)

##### (Class 1,2,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | --- | | **Predicted Class1** | **463** | **216** | **9** | | **Predicted Class2** | **23** | **692** | **2** | | **Predicted Class3** | **3** | **0** | **647** | | **Accuracy**: [0.918]  **Precision**: [0.947 0.762 0.983]  **Recall**: [0.673 0.965 0.995]  **F-Measure:** [0.787 0.852 0.989]  **Mean F-Measure**: 0.876  **Mean Precision**: 0.897  **Mean Recall**: 0.879 |

Table 62: Performance evaluation of class 1,2,3 for cluster 32 (Dataset-2)

#### Cluster=64

##### (Class 1,2)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | | --- | --- | --- | | **Predicted Class1** | **480** | **208** | | **Predicted Class2** | **34** | **683** | | **Accuracy:**  [0.828]  **Precision:** [0.933 0.767]  **Recall:** [0.698 0.953 ]  **F-Measure:**  [0.799 0.849]  **Mean F-Measure:**  0.824  **Mean Precision:** 0.850  **Mean Recall:** 0.825 |

Table 63: Performance evaluation of class 1,2 for cluster 64 (Dataset-2a)

##### (Class 1,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class3** | | --- | --- | --- | | **Predicted Class1** | **679** | **9** | | **Predicted Class3** | **4** | **646** | | **Accuracy:** [0.990]  **Precision:** [0.994 0.986]  **Recall:** [0.987 0.994]  **F-Measure:** [0.990 0.990]  **Mean F-Measure:**  0.990  **Mean Precision:** 0.990  **Mean Recall:** 0.990 |

Table 64: Performance evaluation of class 1,3 for cluster 64 (Dataset-2a)

##### 

##### (Class 2,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | | **Predicted Class2** | **700** | **17** | | **Predicted Class3** | **0** | **650** | | **Accuracy:** [0.987]  **Precision:**  [1.0 0.974]  **Recall:** [0.976 1. 0] ]  **F-Measure:** [0.988 0.987]  **Mean F-Measure:**  0.987  **Mean Precision:** 0.987  **Mean Recall:** 0.988 |

Table 65: Performance evaluation of class 2,3 for cluster 64 (Dataset-2a)

##### (Class 1,2,3)

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | --- | | **Predicted Class1** | **472** | **208** | **8** | | **Predicted Class2** | **31** | **683** | **3** | | **Predicted Class3** | **4** | **0** | **646** | | **Accuracy:**  [0.918]  **Precision:** [0.931 0.767 0.983]  **Recall:** [0.686 0.953 0.994]  **F-Measure:** [0.789 0.849 0.988]  **Mean F-Measure:** 0.876  **Mean Precision:** 0.894  **Mean Recall:** 0.877 |

Table 66: Performance evaluation of class 1,2,3 for cluster 64 (Dataset-2a)

## 3.2.4 Contour Plots

| RDK12  Cluster=1 | RDK22  Cluster=2 |
| --- | --- |
| RDK42  Cluster=4 | RDK82  Cluster=8 |
| RDK162  Cluster=16 | RDK322  Cluster=32 |
| RDK642  Cluster=64 | |

Table 67: Contour plot of cluster 1,2,4,8,16,32,64 for class 1,2,3 of dataset-2(a)

## 3.2.5 Comparison GMM with Bayes ClassifierPerformance

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Case1** | Accuracy=87.16%  Mean Precision =0.852 Mean Recall= 0.808 Mean F-Measure=0.796 | | --- | --- | | **Case2** | Accuracy=87.20%  Mean Precision=0.858  Mean Recall= 0.809 Mean F-Measure=0.796 | | **Case3** | Accuracy=87.6%  Mean Precision=0.858 Mean Recall= 0.816 Mean F-Measure=0.806 | | **Case4** | Accuracy=87.40%  Mean Precision=0.86  Mean Recall= 0.813 Mean F-Measure=0.802 | | | **Cluster1** | **Accuracy:**  [0.872]  **Mean F-Measure:** 0.797  **Mean Precision:** 0.859  **Mean Recall:** 0.810 | | --- | --- | | **Cluster2** | **Accuracy:**  [0.89]  **Mean F-Measure:** 0.832  **Mean Precision:** 0.866  **Mean Recall:** 0.835 | | **Cluster4** | **Accuracy:**  [0.933]  **Mean F-Measure:** 0.899  **Mean Precision:** 0.913  **Mean Recall:** 0.900 | | **Cluster8** | **Accuracy:**  [0.921]  **Mean F-Measure:** 0.880  **Mean Precision:** 0.900 **Mean Recall:** 0.881 | | **Cluster16** | **Accuracy:**  [0.92]  **Mean F-Measure:** 0.879  **Mean Precision:** 0.900 **Mean Recall:** 0.881 | | **Cluster32** | **Accuracy:**  [0.918]  **Mean F-Measure:** 0.876 **Mean Precision:** 0.897 **Mean Recall:** 0.878 | | **Cluster64** | **Accuracy:**  [0.918]  **Mean F-Measure:** 0.877  **Mean Precision:** 0.894 **Mean Recall:** 0.877 | |

Table 68: comparison of GMM with Bayes Classifier

## 3.2.6 Inferences

* Here as we are increasing the number of clusters the accuracy at first is decreasing and is highest for the case when the number of clusters is 2.
* This might be due to the fact that given random real world data and taking 32 clusters on it, means that we have 32(number of clusters) \* (24(Mean-vector) + 24(diagonal covariance matrix)) number of degrees of freedom. Given this number of freedoms, the training data was much less, as the case should be for each degree of freedom we should have 10 training examples but here this was not the case. So there might be over-fitting when the number of clusters were taken higher.
* In this we can observe that we have overlapping data for classes 2 and 3, so Bayes’ classifier here has done hard clustering right away, but GMM keeps some probability associated for each point to every cluster till the end so this also helps us to better estimate the overlapping points.
* If done soft assignment too(overlapping regions) then it also gives a better estimate of relation between different classes.

# 3.3 3-Class Scene Image (DATASET 2(b))

In this dataset 2(b), we have three classes as Bayou, desert\_vegetation and music\_store which contains 50,50 varying dimensional image data in Train and Test Folder in each class respectively.

| Figure 23: (Class1) Vayou dataset image samples of dataset 2(b) |
| --- |
| Figure 24: (Class2) Desert vegetation images samples of dataset 2(b) |
| Figure 25:(Class3) Music store images samples of dataset 2(b) |

## 3.3.1 32 -dimensional Bag of Visual Words Representation

| sun_aarhbeqcuzhoshba_histogram | sun_aarhbeqcuzhoshba |
| --- | --- |
| sun_abulrpiruxyrrzgh_histogram | sun_abulrpiruxyrrzgh |
| sun_aapfaivxqoynnwwp_histogram | sun_aapfaivxqoynnwwp |
| sun_aatbkteovwwsrhmt_histogram | sun_aatbkteovwwsrhmt |
| sun_daifsnxvnraggsqi_histogram | sun_daifsnxvnraggsqi |
| sun_daikaxybbddyqxsu_histogram | sun_daikaxybbddyqxsu |

Table 69: Bag of visual words representation(BoVW) for each of scene image classes(Dataset-2b)

## 3.3.2 Color Coded Histogram

### 3.3.2.1 Log-likelihood vs Iterations plots

#### Cluster=1

| hist-k=1=class1 | hist-k=1=class2 | hist-k=1=class3 |
| --- | --- | --- |

Table 70: iteration vs. log-likelihood of class 1,2,3 for cluster 1 in color coded histogram (Dataset-2(b))

#### 

#### Cluster=2

| hist-k=2=class1 | hist-k=2=class2 | hist-k=2=class3 |
| --- | --- | --- |

Table 71: iteration vs. log-likelihood of class 1,2,3 for cluster 2 in color coded histogram(Dataset-2(b))

#### Cluster=4

| hist-k=4=class1 | hist-k=4=class2 | hist-k=4=class3 |
| --- | --- | --- |

Table 72: iteration vs. log-likelihood of class 1,2,3 for cluster 4 in color coded histogram (Dataset-2(b))

#### Cluster=8

| hist-k=8=class1 | hist-k=8=class2 | hist-k=8=class3 |
| --- | --- | --- |

Table 73: iteration vs. log-likelihood of class 1,2,3 for cluster 8 in color coded histogram(Dataset-2(b))

#### Cluster=16

| hist-k=16=class1 | hist-k=16=class2 | hist-k=16=class3 |
| --- | --- | --- |

Table 74: iteration vs. log-likelihood of class 1,2,3 for cluster 16 in color coded histogram (Dataset-2(b))

### 3.3.2.2 Confusion Matrix, Precision, Recall and F-measure

#### Cluster=1

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | --- | | **Predicted Class1** | **15** | **6** | **29** | | **Predicted Class2** | **4** | **30** | **16** | | **Predicted Class3** | **1** | **1** | **48** | | **Accuracy:**  [0.62]  **Precision:** [0.75 0.81 0.52]  **Recall:** [ 0.3 0.6 0.96]  **F-Measure:** [0.43 0.69 0.67]  **Mean F-Measure:** 0.60  **Mean Precision:** 0.70  **Mean Recall:** 0.62 |

Table 75: Performance evaluation of class 1,2,3 for cluster 1 (Dataset-2b)

#### Cluster=2

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | --- | | **Predicted Class1** | **35** | **4** | **11** | | **Predicted Class2** | **13** | **26** | **11** | | **Predicted Class3** | **4** | **1** | **45** | | **Accuracy:**  [0.706]  **Precision:** [0.67 0.84 0.67]  **Recall:** [ 0.7 0.52 0.9]  **F-Measure:** [0.69 0.64 0.77]  **Mean F-Measure:** 0.70  **Mean Precision:** 0.73  **Mean Recall:** 0.71 |

Table 76: Performance evaluation of class 1,2,3 for cluster 2 (Dataset-2b)

#### Cluster=4

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | --- | | **Predicted Class1** | **28** | **4** | **18** | | **Predicted Class2** | **8** | **35** | **7** | | **Predicted Class3** | **3** | **1** | **46** | | **Accuracy:**  [0.727]  **Precision:** [0.72 0.88 0.65]  **Recall:** [ 0.56 0.70 0.92]  **F-Measure:** [0.63 0.78 0.76]  **Mean F-Measure:** 0.72  **Mean Precision:** 0.75  **Mean Recall:** 0.73 |

Table 77: Performance evaluation of class 1,2,3 for cluster 4 (Dataset-2b)

#### Cluster=8

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | --- | | **Predicted Class1** | **14** | **15** | **21** | | **Predicted Class2** | **5** | **32** | **13** | | **Predicted Class3** | **0** | **5** | **45** | | **Accuracy:**  [0.61]  **Precision:** [0.74,0.62, 0.57]  **Recall:** [ 0.28 0.64 0.90]  **F-Measure:** [0.41 0.63 0.70]  **Mean F-Measure:** 0.58  **Mean Precision:** 0.64  **Mean Recall:** 0.61 |

Table 78:Performance evaluation of class 1,2,3 for cluster 8 (Dataset-2b)

#### Cluster=16

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | --- | | **Predicted Class1** | **14** | **17** | **19** | | **Predicted Class2** | **4** | **38** | **8** | | **Predicted Class3** | **0** | **5** | **45** | | **Accuracy:**  [0.64]  **Precision:** [0.77 0.63 0.62]  **Recall:** [ 0.28 0.76 0.90]  **F-Measure:** [0.41 0.7 0.74]  **Mean F-Measure:** 0.62  **Mean Precision:** 0.68  **Mean Recall:** 0.65 |

Table 79: Performance evaluation of class 1,2,3 for cluster 16 (Dataset-2b)

### 3.3.2.3:Inferences

* The maximum accuracy of 65% is obtained at K = 16 clusters per class and decreases thereafter.
* The classification accuracy of the image dataset is highly dependent on the feature extraction process.
* In this case, every image is represented by 32 fixed size blocks and each block is a feature vector set of 24-dimensional color histogram feature vectors.

## 3.3.3 Bag of Visual Words

### 3.3.3.1 Log-likelihood vs Iterations plots

#### Cluster=1

| bovw-k=2,class1 | bovw-k=2,class2 | bovw-k=2,class3 |
| --- | --- | --- |

Table 80: iteration vs. log-likelihood of class 1,2,3 for cluster 1 in BOVW(Dataset-2(b))

#### Cluster=2

| bovw-k=2,class1 | bovw-k=2,class2 | bovw-k=2,class3 |
| --- | --- | --- |

Table 81: iteration vs. log-likelihood of class 1,2,3 for cluster 2 in BOVW(Dataset-2(b))

#### Cluster=4

| bovw-k=4,class1 | bovw-k=4,class2 | bovw-k=4,class3 |
| --- | --- | --- |

Table 82: iteration vs. log-likelihood of class 1,2,3 for cluster 4 in BOVW(Dataset-2(b))

#### 

#### Cluster=8

| bovw-k=8,class1 | bovw-k=8,class2 | bovw-k=8,class3 |
| --- | --- | --- |

Table 83: iteration vs. log-likelihood of class 1,2,3 for cluster 8 in BOVW(Dataset-2(b))

#### Cluster=16

| bovw-k=16,class1 | bovw-k=16,class2 | bovw-k=16,class3 |
| --- | --- | --- |

Table 84: iteration vs. log-likelihood of class 1,2,3 for cluster 16 in BOVW(Dataset-2(b))

#### Cluster=32

| bovw-hist=32-class1 | bovw-hist=32-class2 | bovw-hist=32-class3 |
| --- | --- | --- |

Table 85: iteration vs. log-likelihood of class 1,2,3 for cluster 32 in BOVW(Dataset-2(b))

#### Cluster=64

| bovw-hist=64-class1 | bovw-hist=64-class2 | bovw-hist=64-class3 |
| --- | --- | --- |

Table 86: iteration vs. log-likelihood of class 1,2,3 for cluster 64 in BOVW(Dataset-2(b))

### 3.3.3.2 Confusion Matrix, Precision, Recall and F-measure

#### Cluster=1

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | --- | | **Predicted Class1** | **30** | **10** | **10** | | **Predicted Class2** | **5** | **37** | **8** | | **Predicted Class3** | **10** | **8** | **32** | | **Accuracy:**  [0.66]  **Precision:** [0.67, 0.67, 0.64]  **Recall:** [ 0.6 0.74 0.64]  **F-Measure:** [0.63 0.70 0.64]  **Mean F-Measure:** 0.656  **Mean Precision:** 0.66  **Mean Recall:** 0.66 |

Table 87: Performance evaluation of class 1,2,3 for cluster 1 (Dataset-2b)

#### Cluster=2

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | --- | | **Predicted Class1** | **30** | **13** | **7** | | **Predicted Class2** | **9** | **36** | **5** | | **Predicted Class3** | **14** | **7** | **29** | | **Accuracy:**  [0.63]  **Precision:** [0.57, 0.64, 0.70]  **Recall:** [ 0.6 0.72 0.58]  **F-Measure:** [0.58 0.68 0.63]  **Mean F-Measure:** 0.63  **Mean Precision:** 0.64  **Mean Recall:** 0.63 |

Table 88: Performance evaluation of class 1,2,3 for cluster 2 (Dataset-2b)

#### Cluster=4

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | --- | | **Predicted Class1** | **29** | **14** | **7** | | **Predicted Class2** | **11** | **33** | **6** | | **Predicted Class3** | **7** | **11** | **32** | | **Accuracy:**  [0.63]  **Precision:** [0.62,0.57, 0.71]  **Recall:** [ 0.58 0.66 0.64]  **F-Measure:** [0.58 0.61 0.67]  **Mean F-Measure:** 0.63  **Mean Precision:** 0.62  **Mean Recall:** 0.63 |

Table 89: Performance evaluation of class 1,2,3 for cluster 4 (Dataset-2b)

#### Clusters=8

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | --- | | **Predicted Class1** | **34** | **8** | **8** | | **Predicted Class2** | **16** | **27** | **7** | | **Predicted Class3** | **16** | **2** | **32** | | **Accuracy:**  [0.62]  **Precision:** [0.51,0.73, 0.68]  **Recall:** [ 0.68 0.54 0.64]  **F-Measure:** [0.59 0.62 0.66]  **Mean F-Measure:** 0.62  **Mean Precision:** 0.64  **Mean Recall:** 0.62 |

Table 90: Performance evaluation of class 1,2,3 for cluster 8 (Dataset-2b)

#### Clusters=16-64

| **Confusion Matrix** | **Precision, Recall and F-measure** |
| --- | --- |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | --- | | **Predicted Class1** | **26** | **12** | **12** | | **Predicted Class2** | **4** | **39** | **7** | | **Predicted Class3** | **10** | **6** | **34** | | **Accuracy:**  [0.66]  **Precision:** [0.65,0.68, 0.64]  **Recall:** [ 0.52 0.78 0.68]  **F-Measure:** [0.58 0.73 0.66]  **Mean F-Measure:** 0.66  **Mean Precision:** 0.66  **Mean Recall:** 0.66 |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | --- | | **Predicted Class1** | **31** | **10** | **9** | | **Predicted Class2** | **10** | **29** | **11** | | **Predicted Class3** | **10** | **9** | **31** | | **Accuracy:**  [0.61]  **Precision:** [0.61,0.60, 0.61]  **Recall:** [ 0.62 0.58 0.62]  **F-Measure:** [0.61 0.59 0.61]  **Mean F-Measure:** 0.60  **Mean Precision:** 0.61  **Mean Recall:** 0.61 |
| | **Actual\Predicted** | **Actual Class1** | **Actual Class2** | **Actual Class3** | | --- | --- | --- | --- | | **Predicted Class1** | **40** | **2** | **8** | | **Predicted Class2** | **19** | **22** | **9** | | **Predicted Class3** | **14** | **4** | **32** | | **Accuracy:**  [0.63]  **Precision:** [0.55,0.78, 0.65]  **Recall:** [ 0.8 0.44 0.64]  **F-Measure:** [0.65 0.56 0.64]  **Mean F-Measure:** 0.62  **Mean Precision:** 0.66  **Mean Recall:** 0.63 |

Table 91: Performance evaluation of class 1,2,3 for cluster 16,32,64 (Dataset-2b)

### 3.3.3.3: Inferences

* The maximum accuracy of 71.667% is obtained at K = 4 clusters per class.
* In this case, every image is represented by 32 dimensional Bag-of-visual-words  (BoVW) obtained by 32 clustering of the color coded histograms.
* The accuracy in case of BoVW representation is better than that obtained by color  histogram feature vector.

# 3.4: Cervical cytology (cell): DATASET 2(c)

In the dataset 2(c), there is a train and test folder that contains 60 and 3 images respectively of 512x512 pixels.



Figure 26: Sample images of Dataset-2(c)

### 3.4.1: KMeans Clustering

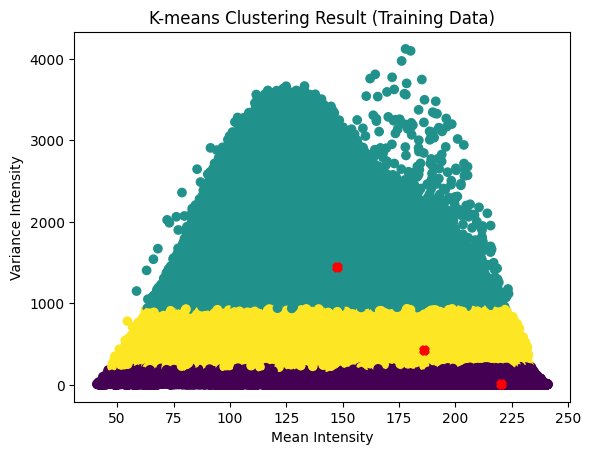


Figure 27: Mean vs Variance intensity plot of K-Means Clustering (Training Data)

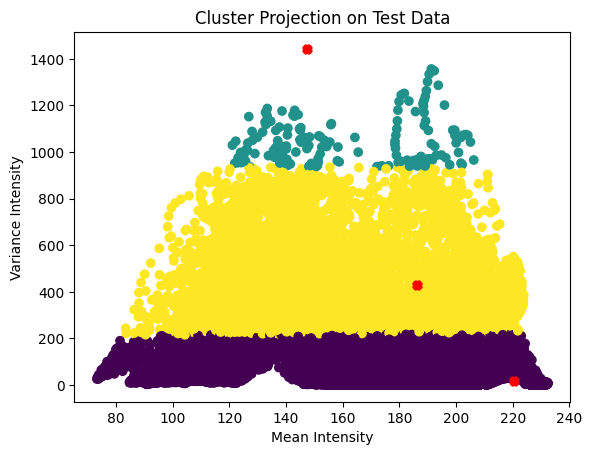


Figure 28: Mean vs Variance intensity plot of K-Means Clustering (Testing Data)

| **Cell Image** | **Cluster Projection** | **Segmentation Result** |
| --- | --- | --- |
| IMG_256 | IMG_256 | IMG_256 |
| IMG_256 | IMG_256 | IMG_256 |
| IMG_256 | IMG_256 | IMG_256 |

Table 92 : K-Means clustering results

### 3.4.2:Modified KMeans Clustering

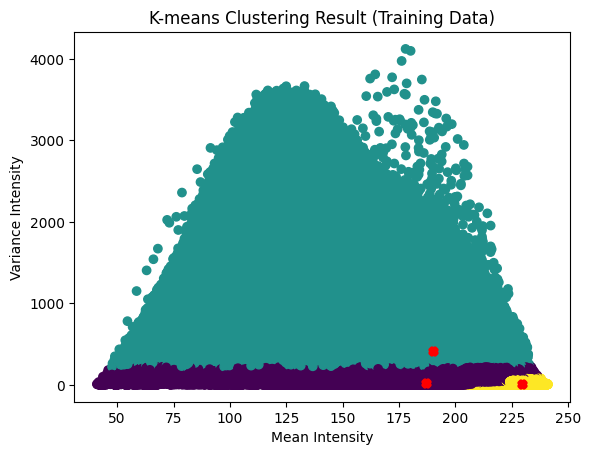


Figure 29: Mean vs Variance intensity plot of Modified K-Means Clustering (Training Data)

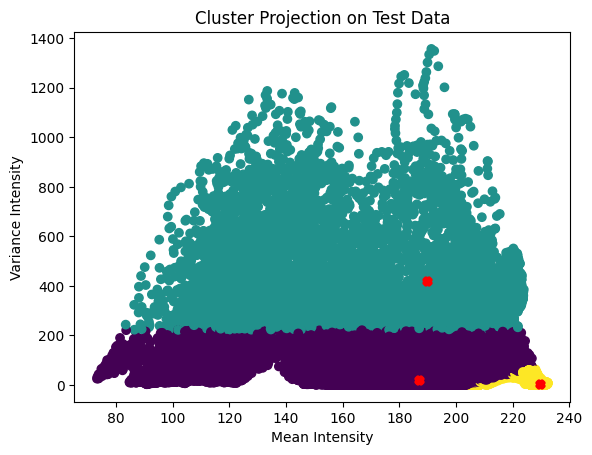


Figure 30: Mean vs Variance intensity plot of Modified K-Means Clustering (Testing Data)

| **Cell Image** | **Cluster Projection** | **Segmentation Result** |
| --- | --- | --- |
| IMG_256 | IMG_256 | IMG_256 |
| IMG_256 | IMG_256 | IMG_256 |
| IMG_256 | IMG_256 | IMG_256 |

Table 93: Modified K-Means clustering results

### 3.4.3: Inferences

* Better segmentation is observed in case of GMM as compared to that of K-Means  clustering.
* GMM allows non linear decision boundary therefore better segmenting between the  three parts.
* Clustering accuracy depends on the way image feature vectors are extracted.
* In  this case 7x7 non overlapping patches were used to form feature vectors of stack 2  dimensional feature vectors of mean and variance of the patch of the training set. ∙ The test images were represented by overlapping 7x7 such patches.

# 4: Conclusion and Observations

* The decision boundaries are more precise when the data is modeled using a mixture  of multiple Gaussians as compared to unimodal Gaussian.
* Different classification accuracies are obtained when the number of clusters K are varied.
* Although the accuracy tends to increase with the number of clusters assumed for a  class, due to overlapping data, over clustering may cause the class to cover  non-belonging points as well, this is evident in the real world data.
* The boundary between 2 clusters in K-Means is linear because the distance measure  used in K-means classification is Euclidean distance.
* As the number of iterations increases, cluster assignment of the data points changes  and approaches convergence.
* The accuracies obtained for the nonlinear datasets are far better than that obtained  in the unimodal case. Accuracies for real dataset are behind that obtained in unimodal  cases.
* The value of log likelihood increases with the number of iterations till it reaches the  local / global optima.

# Bibliography

[1] Lecture notes of CS669 August 2023-24 by Dr.Dileep A D

[2] K-Means clustering

https://en.wikipedia.org/wiki/K-means\_clustering https:

//www.geeksforgeeks.org/k-means-clustering-introduction/

https://towardsdatascience.com/understanding-k-means-clustering-\

in-machine-learning-6a6e67336aa1

[3] GMM

https://en.wikipedia.org/wiki/Mixture\_model http://www.

computerrobotvision.org/2010/tutorial\_day/GMM\_said\_crv10\_

tutorial.pdf

https://www.cse.iitk.ac.in/users/piyush/courses/pml\_winter16/

slides\_lec7.pdf

[4] Bayes Classifier

https://en.wikipedia.org/wiki/Bayes\_classifier

[5] Statistical Classification

https://en.wikipedia.org/wiki/Statistical\_classification

[6] Naive Bayes Classifier

https://en.wikipedia.org/wiki/Naive\_Bayes\_classifier

[7] Matplotlib Contours

https://matplotlib.org/api/\_as\_gen/matplotlib.pyplot.contour.

html

[7] Stack Overflow

https://stackoverflow.com