SPR Assignment 3

Jayanth S (201081003), Praveen Kumar N (201082001), Rishabh Roy (201082002)

Contents

1 MLE Estimation

- Let $mathbbD_c$ be the set containing all the sample points that belong to class c $\mathbb{D}_c = x_{i=1}^n$
- In MLE estimation, we would like to maximise the likelihood

$$l(\theta) = \mathcal{P}(\mathbb{D} : \theta)$$

- Wkt maximising log-likelihood and likelihood are same. Hence we try to maximise the log-likelihood, i.e.,

$$L(\theta) = log(l(\theta))$$

- The estimated parameter $\hat{\theta}$ is given by,

$$\hat{\theta} = arg \max_{\theta} L(\theta)$$

1.1 Gaussian Distribution

- Estimated mean for i^{th} class is a vector with 784 mean feature values(pixel values) with j^{th} pixel value equal to:

$$\mu_{ij} = \sum_{l=1}^{m} \frac{x_{jl}}{m}, \tag{1}$$

  where, m is the number of images of i^{th} particular class and
  x_{jl} is the j^{th} pixel value of l^{th} image.

- The covaiance matrix for i^{th} class it is ,

$$\Sigma_i = \sum_{l=1}^{m} \frac{z_l}{m}, \tag{2}$$

  where, $z_l = \frac{(x_l - \mu_i) * (x_l - \mu_i)^T}{n}$
  Here x_l represents l^{th} image of i^{th} class and μ_i represents the mean of i^{th} class images.

1.2 Exponential Distribution

Estimate of paramter λ for i^{th} class is a vector with 784 values with j^{th} value equal to:

$$\lambda_{ij} = \sum_{l=1}^{m} \frac{m}{x_{jl}}, \tag{3}$$

where, m is the number of images of i^{th} particular class and
x_{jl} is the j^{th} pixel value of l^{th} image.

2 Observations

2.1 Observations for Q1 and Q3

- Modeling the class conditional densities as Multivariate Gaussian distribution results in high accuracy compared to the Multivariate Exponential model.
- For MNIST dataset with multivariate gaussian model class 1 and 3 are misclassified as 8 many times. Hence combining these 3 as a single class reduced our misclassification. Hence increasing the accuracy of the model.
- For MNIST Fashion dataset with multivariate exponential model, class 6 is misclassified as 0 many times. Hence combining these 2 would reduce the misclassification thereby increasing the accuracy.
- We have added a noise of 0.35 to the diagonal elements of covariance matrix. Adding a value below this resulted in determinant of covariance matrices of some classes to be zero. When inbuit function multivariate_normal.logpdf(x,mean,covariance) is used to estimate the pdf at x, we were able to get an accuaracy of 87
- When Multivariate Exponential Model is used, lambda parameter goes to infinity for some feature components. Hence we have added some non-zero value while estimating the lamdba parameter. But always the accuracy of multivariate exponential model is less than the multivariate gaussian distribution model. We think this is because both MNIST dataset and MNIST fashion dataset are obtained from Multivariate Gaussian Distribution

2.2 Plots obtained for Q1

- MNIST handwritten digits dataset contains 60,0000 training images (28x28 pixels) of a handwritten single digit (0–9) and 10,000 testing images and our goal is to learn the parameters of each class using MLE and then classify the image as the appropriate digit using the estimated parameters.
- When Bayes classifier was implemented on MNIST Dataset by modeling the class conditonal densities as multivariate gaussian distribution we got the below plots.

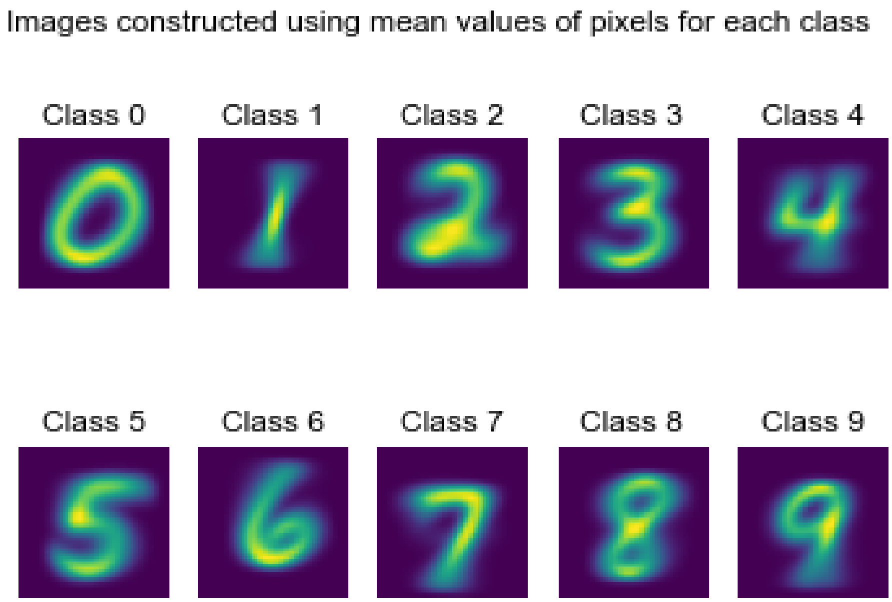


Figure 1: Images obtained from estimated parameters for each class by modeling class conditional densities as multivariate gaussian distribution

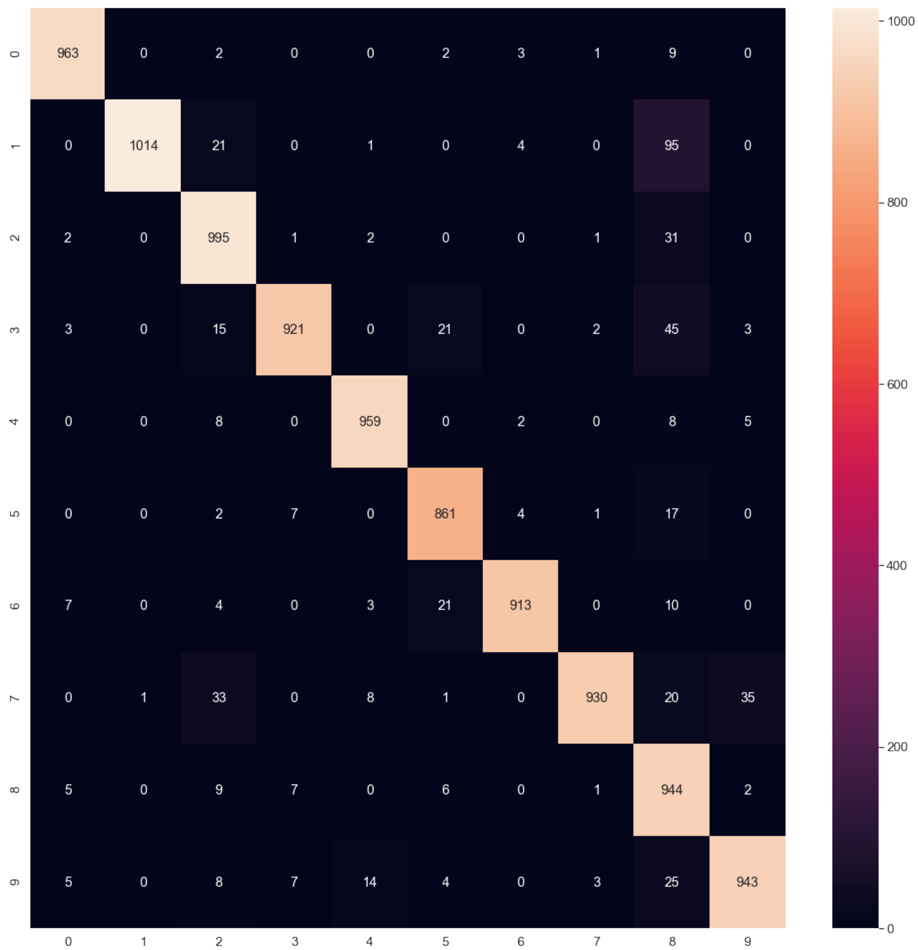
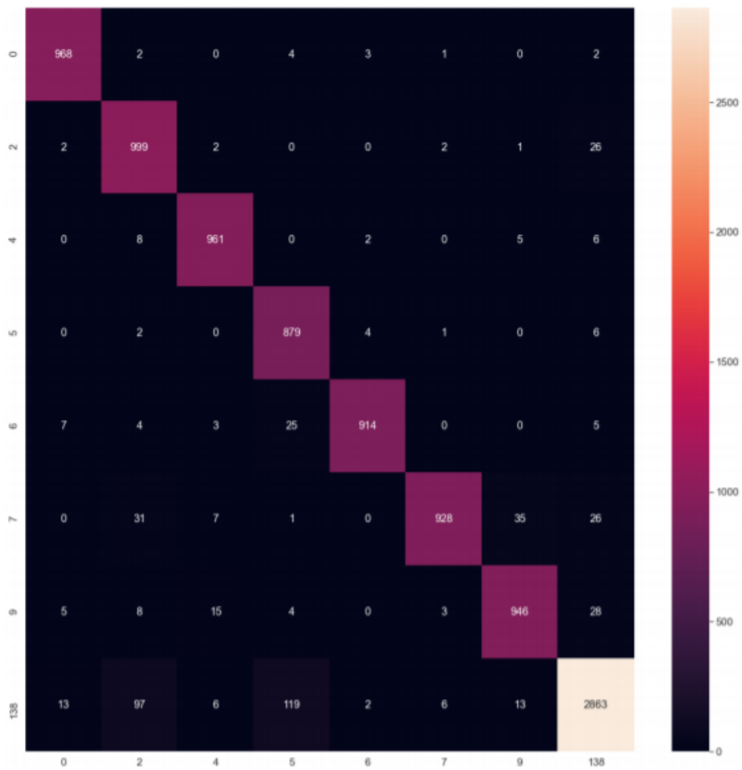


Figure 2: Confusion matrix obtained when class condition densities are modeled as multivariate gaussian distribution. x-axis represents predicted labels and y-axis represents true labels. The test accuracy is also given.




```
The % of misclassification is 5.42% for MLE estimation assuming the data is
sampled from Multivariate Gaussian Distribution with classes 1,3,8 combined
```

Figure 3: Confusion matrix obtained when class condition densities are modeled as multivariate exponential distribution and the classes 1, 3 and 8 are combined. x-axis represents predicted labels and y-axis represents true labels. The test accuracy is also given.

- When Bayes classifier was implemented on MNIST Dataset by modeling the class conditonal densities as multivariate exponential distribution we got the below plots.

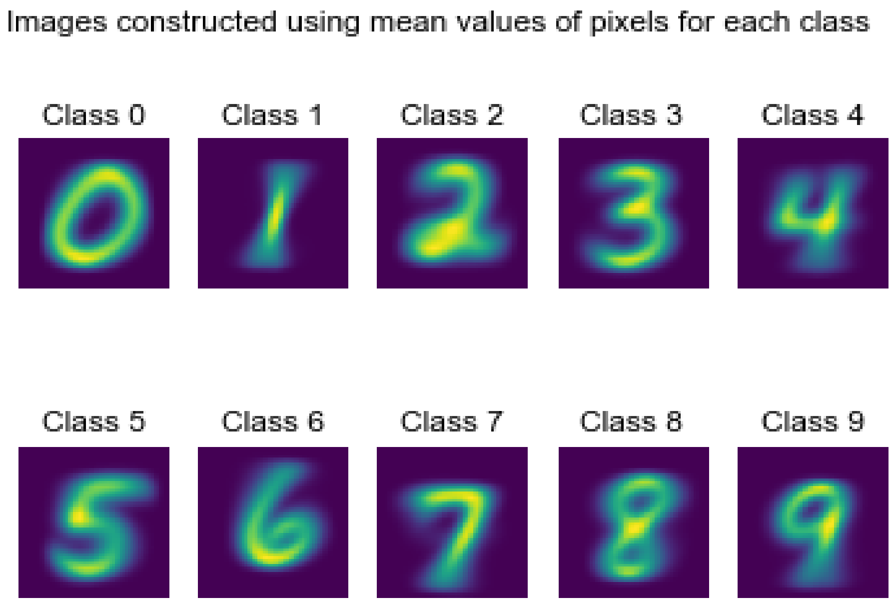


Figure 4: Images obtained from estimated parameters for each class by modeling class conditional densities as multivariate exponential distribution

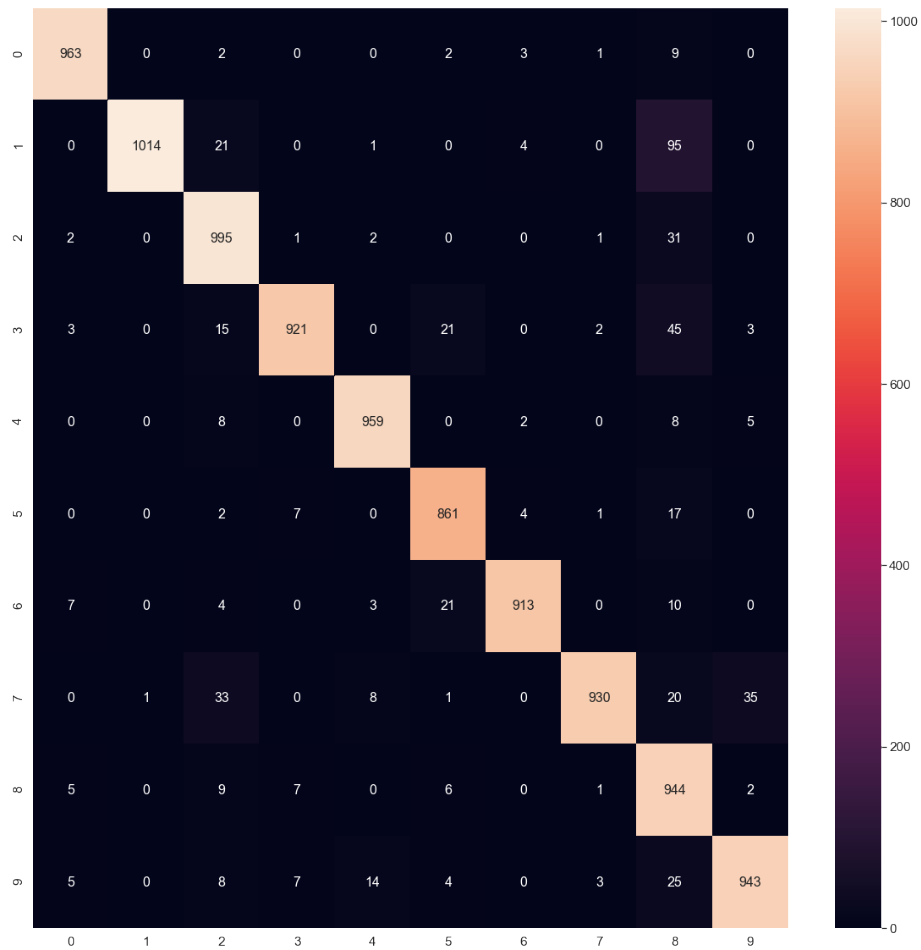


Figure 5: Confusion matrix obtained when class condition densities are modeled as multivariate exponential distribution. x-axis represents predicted labels and y-axis represents true labels. The test accuracy is also given.

2.3 Question 2

2.3.1 Dataset

Read and implement Multinomial Naive Bayes classifier on 20 Newsgroups Dataset. Compare it with sklearn implementation of Multinomial Naive Bayes Classifier.

DATASET SOURCE:http://qwone.com/~jason/20Newsgroups/

Background:

- Dataset consists of 18,886 files(11,334 for Training and 7,552 for Testing) of 20 different categories.
- The data is organized into 20 different newsgroups, each corresponding to a different topic. Some of the newsgroups are very closely related to each other, while others are highly unrelated. Here is a list of the 20 newsgroups.

Figure 6: Some information about 20 news group dataset

```
Number of news(files) in Training data :11314

Number of news(files) in Testing data  :7532
```

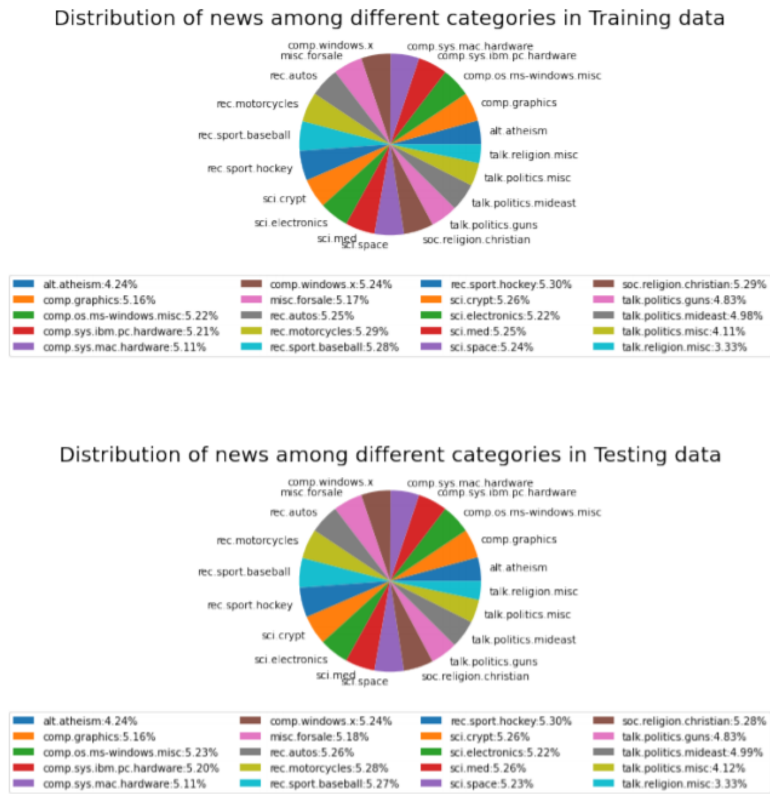


Figure 7: Distribution of news among different categories in the training and testing data

2.3.2 Building Multinomial Naive Bayes classifier:

- Here we have M(11314) number of training data(files) with class labels, distributed among K(20) classes.
- First we estimate prior probabilities(class probabilities) $P(C_k)$ for all $k = 1, \ldots, K$ from M training examples.
- Then we build a vocabulary(set of all distinct words in training data). here each word in vocabulary is a feature(there are n features).
- Next we transform each training data(file) into a feature vector $X = [x_1, \ldots x_n]$.

- Now using training data we estimate $P(X/C_k) = P(x_1/C_k) \times \cdots \times P(x_n/C_k)$(since in Naive bayes classifier we assume features are independent) for all $k = 1, \ldots, K$.
- Once we have estimated the likelihood of all classes, given a new feature vector X we compute the posterior probability $P(C_k/X) = P(X/C_k) \times P(C_k)/P(X) \propto P(X/C_k) \times P(C_k)$ for all $k = 1, \ldots, K$.
- Finally we use MAP rule to classify the given new feature vector to the class C_{NB} which maximizes the posterior probability i.e, $C_{NB} = argmax_{C_k} P(X/C_k) \times P(C_k)$.

2.3.3 Observations

- Accuracy of the In built Multinomial Naive Bayes classifier was around 84
- Lower values of alpha(smoothing parameter) gave better accuracy.
- Accuracy of the Multinomial Naive Bayes classifier bulit from scratch was around 81

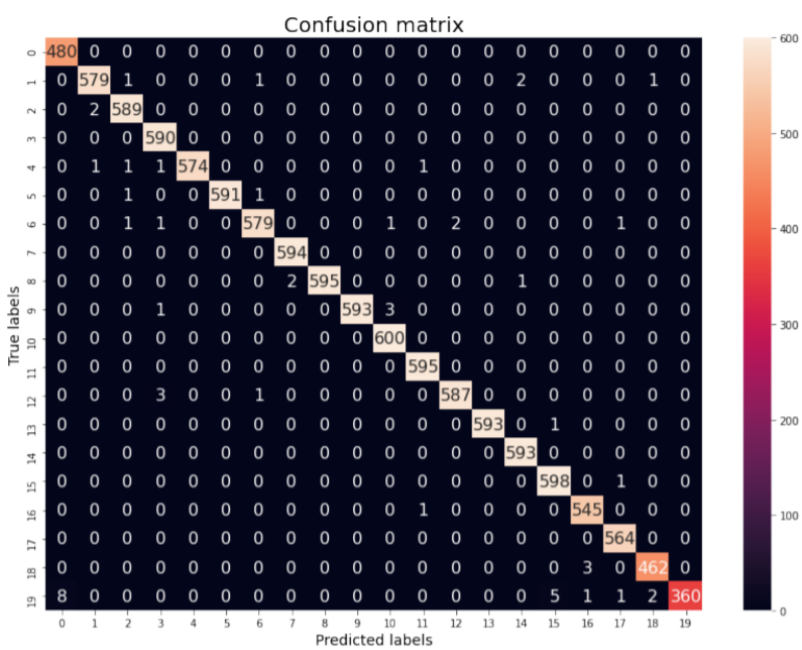


Figure 8: Confusion matrix obtained when Multinomial Naive Bayes classifier is bulit from scratch

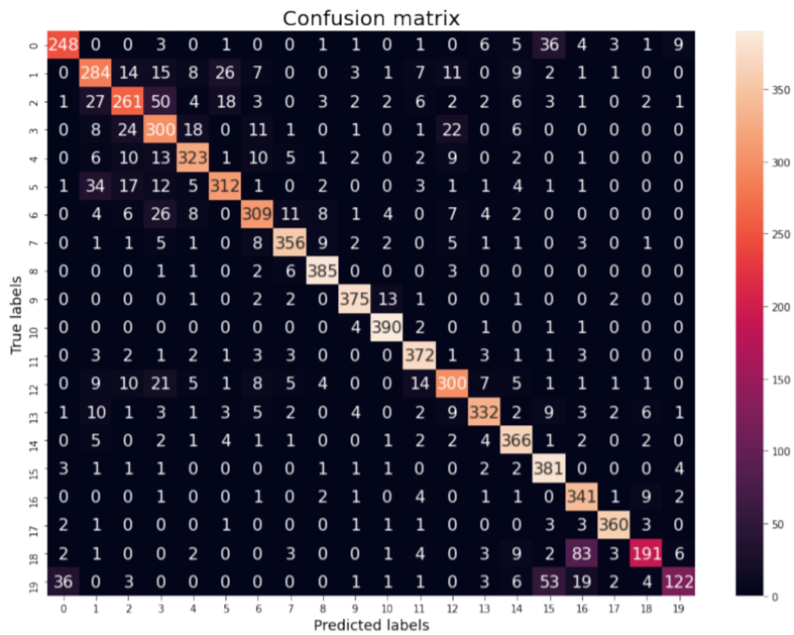


Figure 9: Confusion matrix obtained when inbuilt Multinomial Naive Bayes classifier is used

2.4 Observation for Question 3

- When Bayes classifier was implemented on Fashion MNIST Dataset by modeling the class conditonal densities as multivariate exponential distribution we got the below plots.

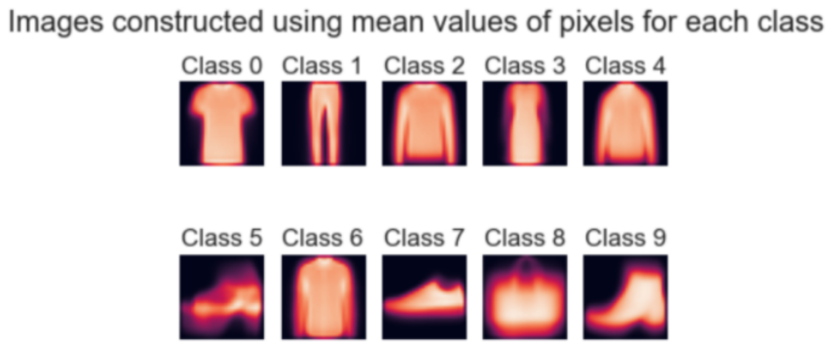
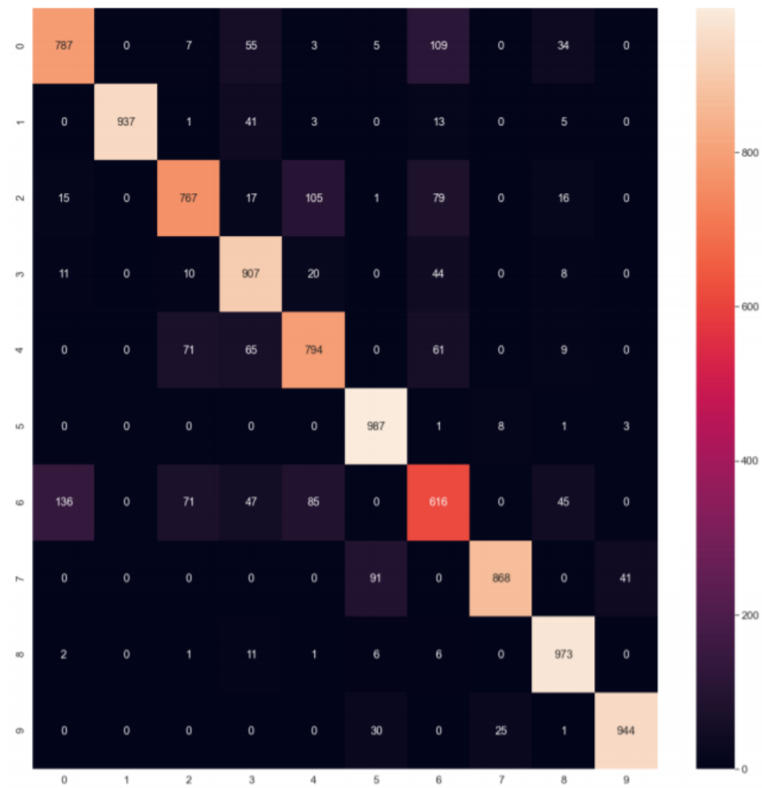


Figure 10: Images obtained from estimated parameters for each class by modeling class conditional densities as multivariate gaussian distribution




```
The % of misclassification is 14.2% for MLE estimation assuming the data is
sampled from Multivariate Gaussian Distribution
```

Figure 11: Confusion matrix obtained when class condition densities are modeled as multivariate gaussian distribution. x-axis represents predicted labels and y-axis represents true labels. The test accuracy is also given.

- When Bayes classifier was implemented on Fashion MNIST Dataset by modeling the class conditonal densities as multivariate exponential distribution we got the below plots.

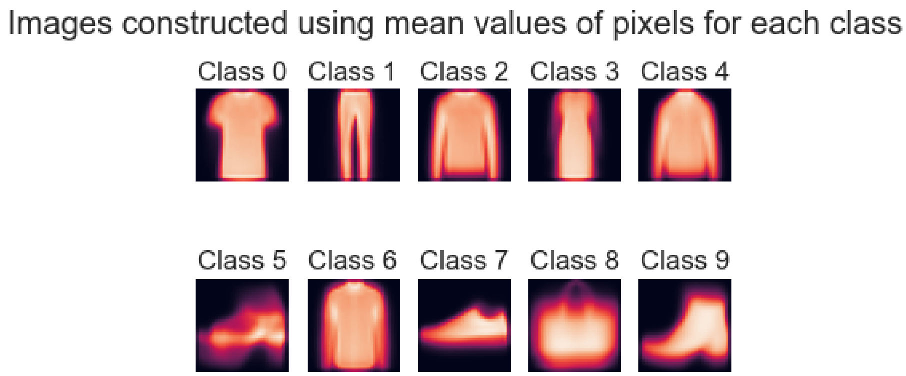


Figure 12: Images obtained from estimated parameters for each class by modeling class conditional densities as multivariate exponential distribution

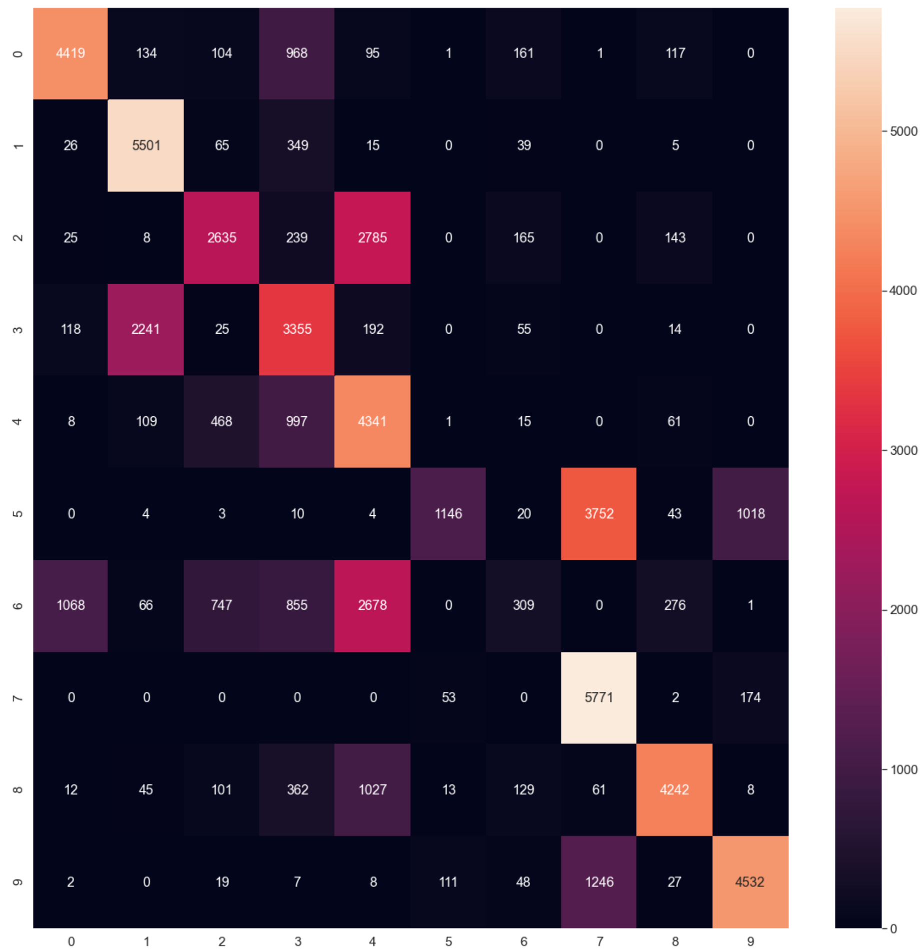


Figure 13: Confusion matrix for train data obtained when class condition densities are modeled as multivariate exponential distribution. x-axis represents predicted labels and y-axis represents true labels.

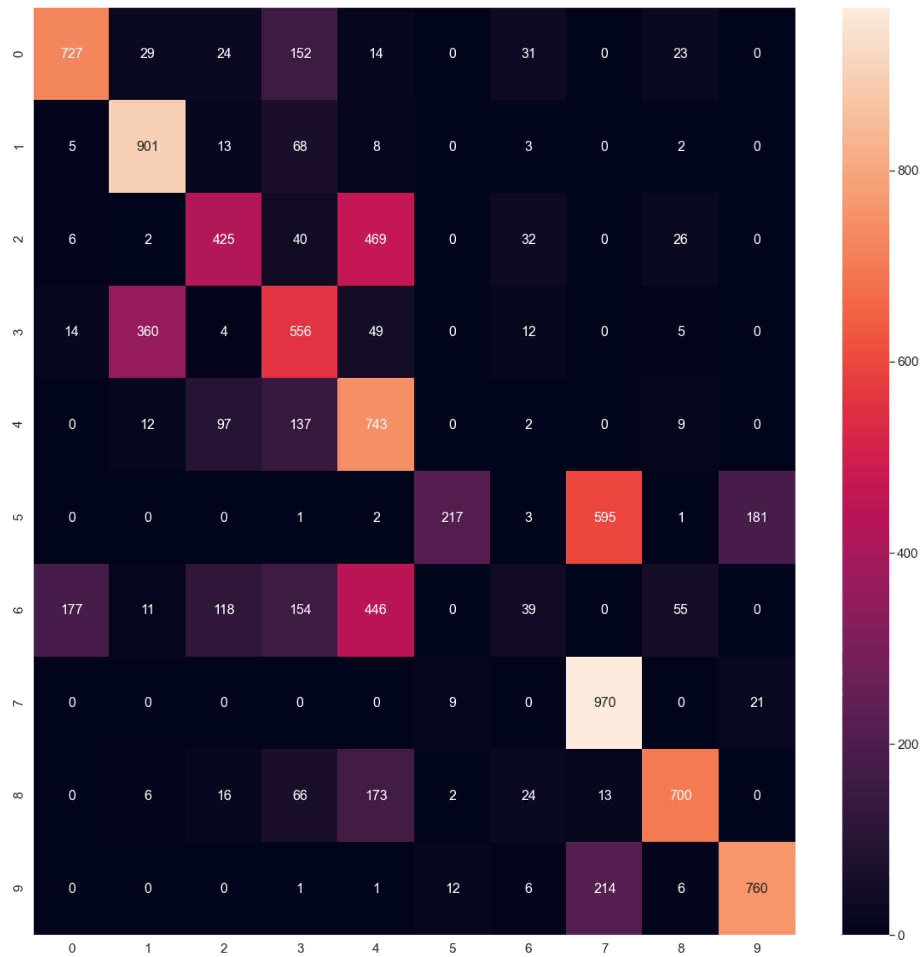


Figure 14: Confusion matrix for test data obtained when class condition densities are modeled as multivariate exponential distribution. x-axis represents predicted labels and y-axis represents true labels.

- The % of train accuracy is 60.4183
- The % of for test accuracy is 60.38