Machine Intelligence and Learning

Indian Institute of Technology Delhi

COURSE ASSIGNMENT 1

Mayank Mishra 2016EE30506 Chahat Chawla 2016MT10492 Gantavya Bhatt 2016EE10694

Medical Dataset

Data Visualization

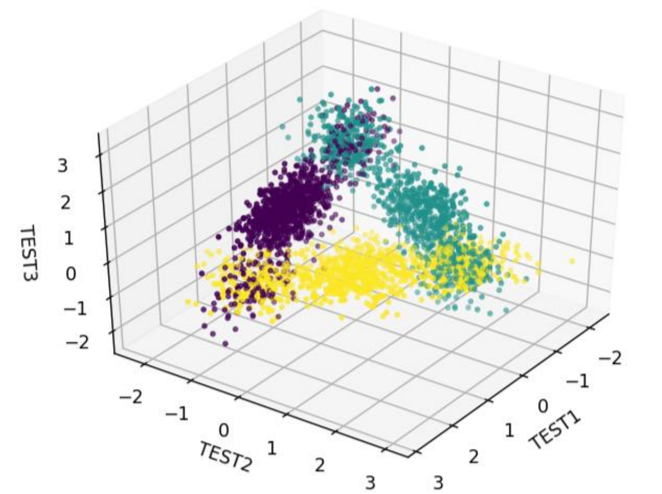

Fig. 1: True distribution of training data

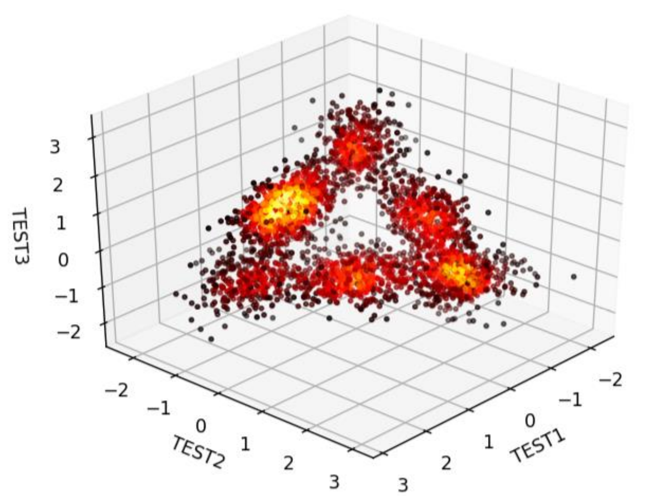

Fig. 2: Heat Map for analyzing the number of Gaussians (4D plot)

Figure 2 shows the density function f(x) where the 3 axes are the features and the color is the fourth dimension. Brighter color means a higher value. The same is shown in the following contour plots of Figure 3.

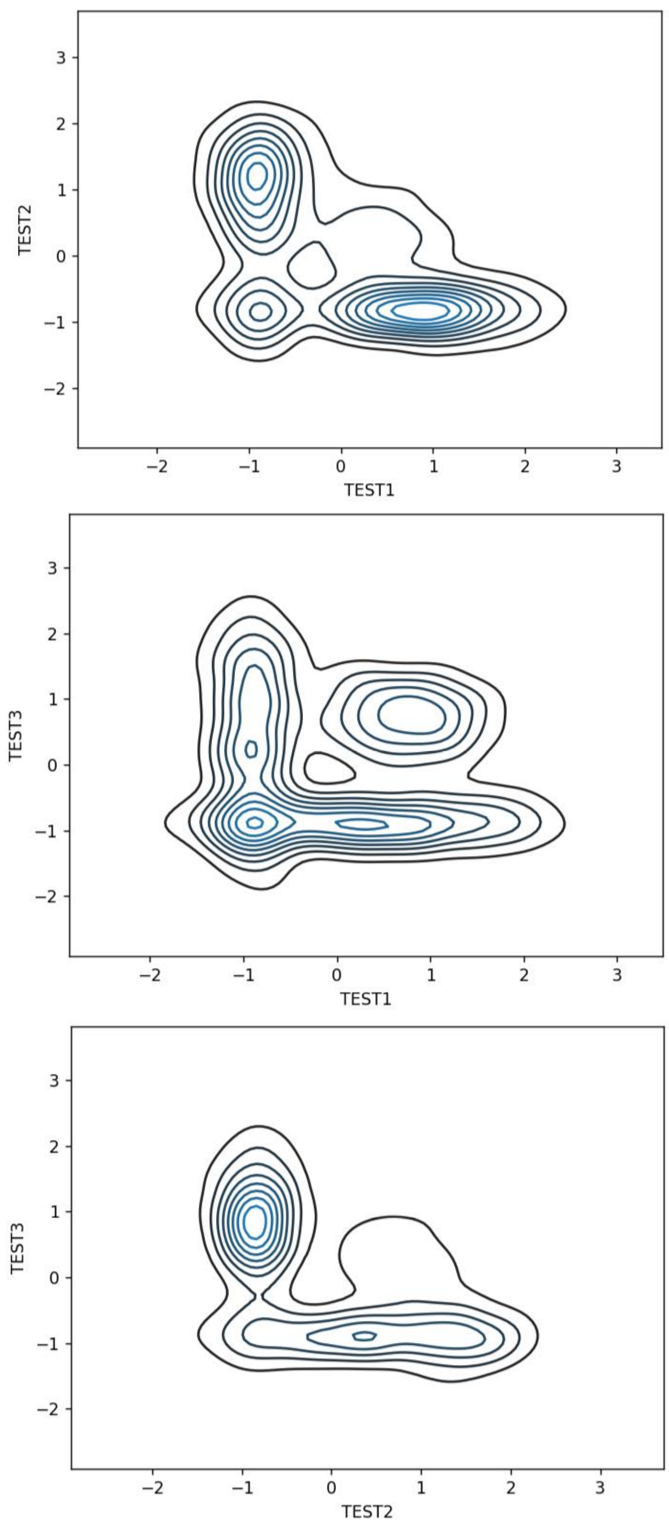

Fig. 3: Pairwise Contour Plots assuming Gaussian Kernel

Data being 3 dimensional, we employ class conditional plots and hyper tuning for selecting the optimal number of Gaussian mixtures for class conditionals. All combinations of classifiers with Naïve Bayes, Bayes and normal, GMM class conditional gives nearly the same performance.

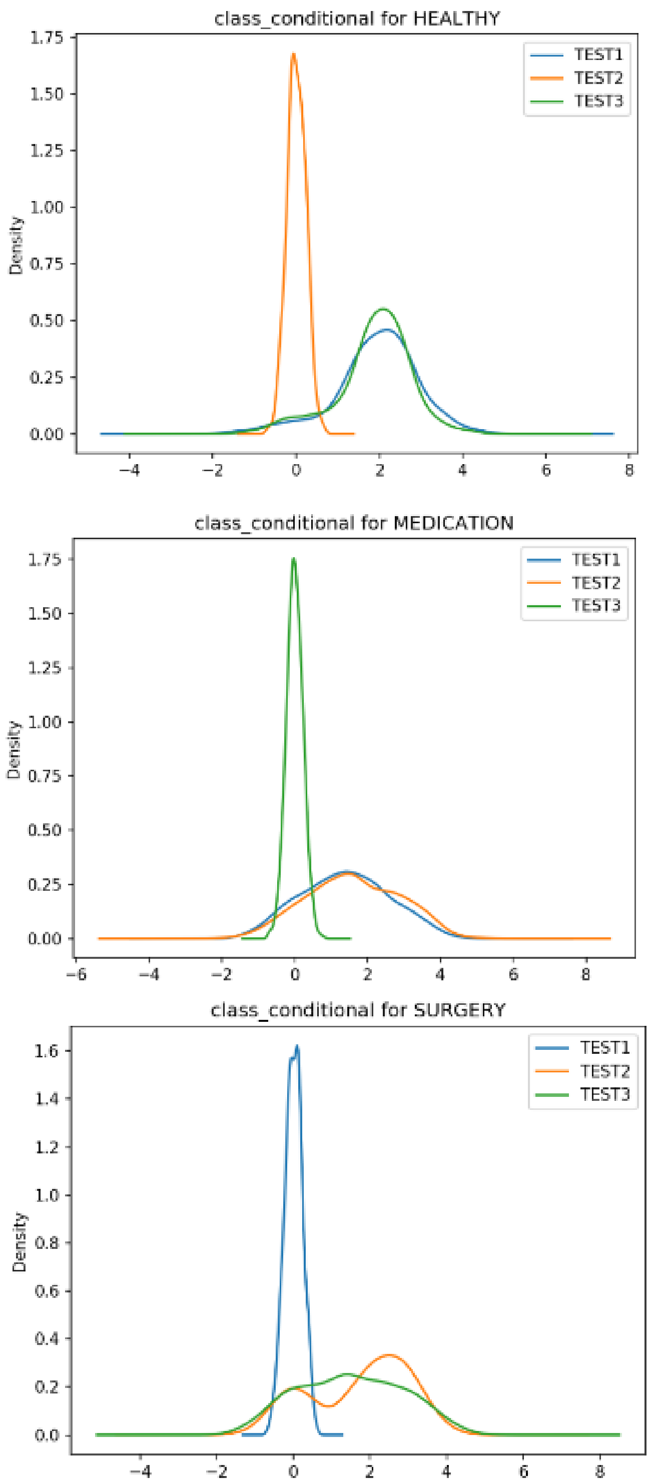

Fig. 4: Class Conditional Distribution of Features

Class\Feature	TEST1	TEST2	TEST3
Healthy	1	1	1
Surgery	1	2	2
Medication	1	2	1

Table 1: Optimal number of Gaussians for class conditional densities

Algorithm\Dataset	Train	Test
Naive Bayes MLE	90.66	89.86
Naive Bayes EM	90.96	90.16
Bayes MLE	90.33	89.76
Bayes EM	90.66	90.06
kNN (k = 11) Minkowski (n = 3) distance		89.73
Parzen window (hypercube) h = 0.65		89.3
Parzen window (Gaussian) h = 0.2		89.93

Table 2: Performance

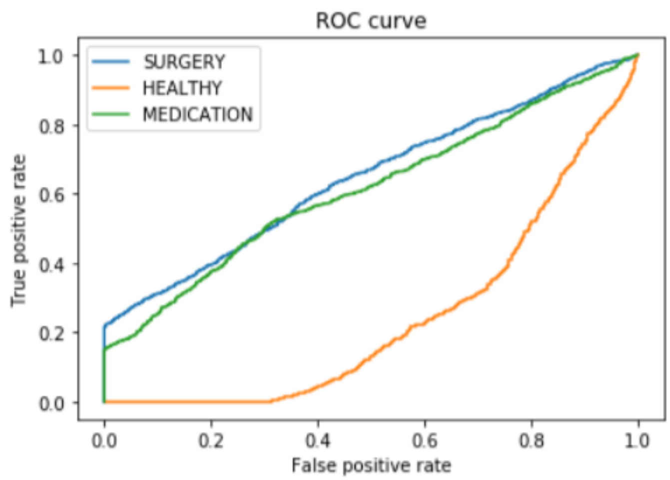

Fig. 5: ROC curves for Naive Bayes classifier

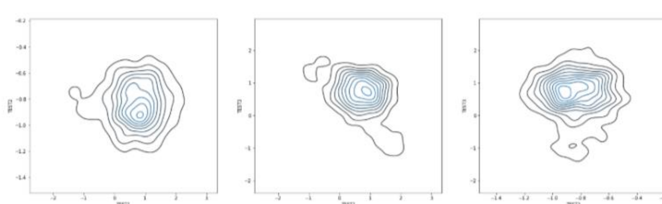

Fig. 6: Class conditional density for HEALTHY class (taking features pairwise)

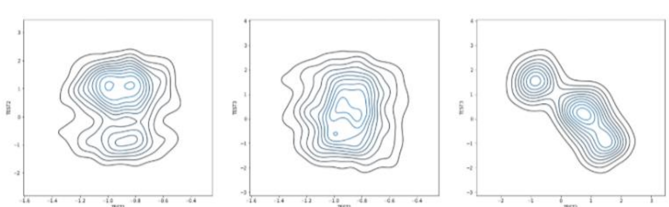

Fig. 7: Class conditional density for SURGERY class (taking features pairwise)

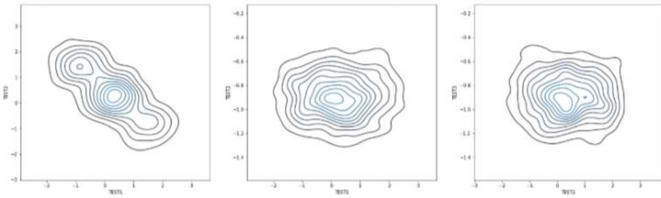

Fig. 8: Class conditional density for MEDICATION class (taking features pairwise)

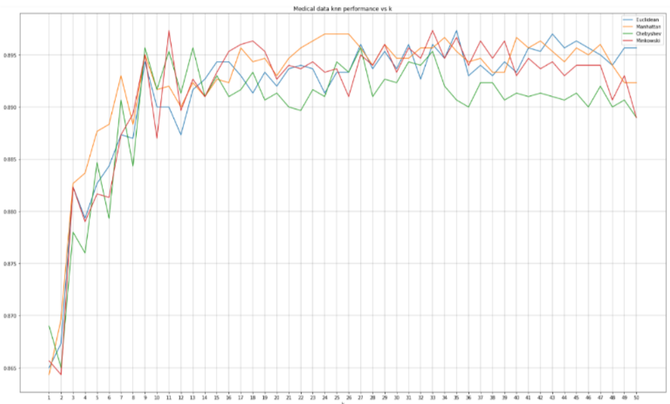

Fig. 9: Test accuracy vs k for kNN

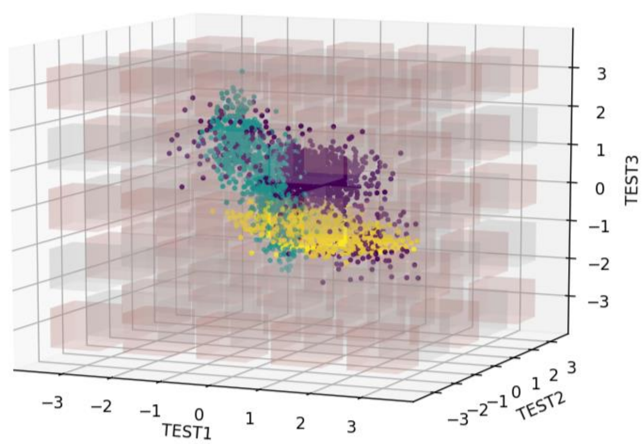

Fig. 10: Parzen Window Density Estimate Visualization (Hypercubes, and Density visualized)

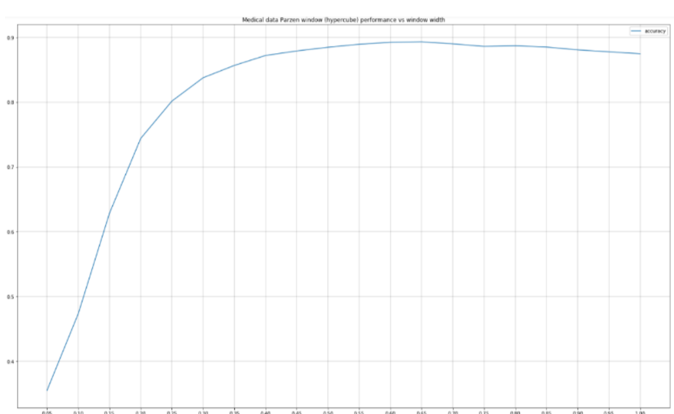

Fig. 11: Test accuracy vs h for Parzen window (hypercube window)

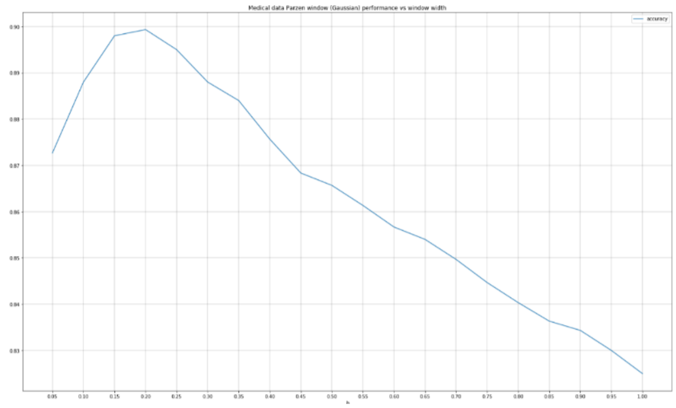

Fig. 12: Test accuracy vs h for Parzen window (Gaussian window)

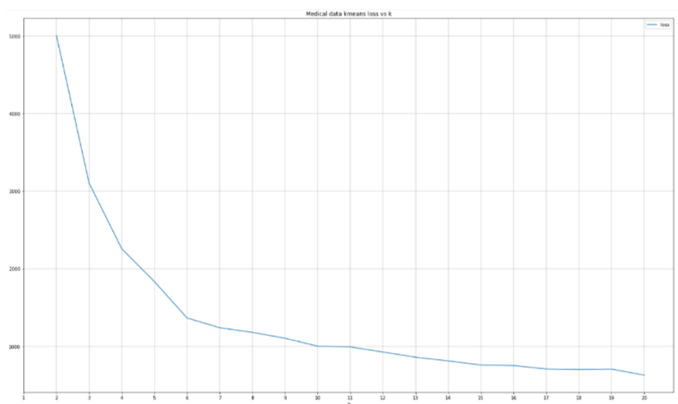

Fig. 13: Loss vs k for K-Means on CV data

For kmeans, knee-point occurs at k = 6. The clusters learnt failed to encompass the classification (healthy, surgery, medication).

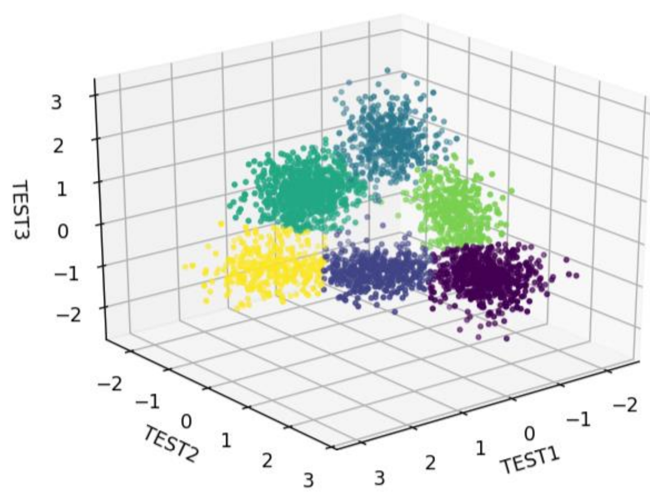

Fig. 14: Clusters learned by K Means for k=6

Railway Dataset

Data Visualization

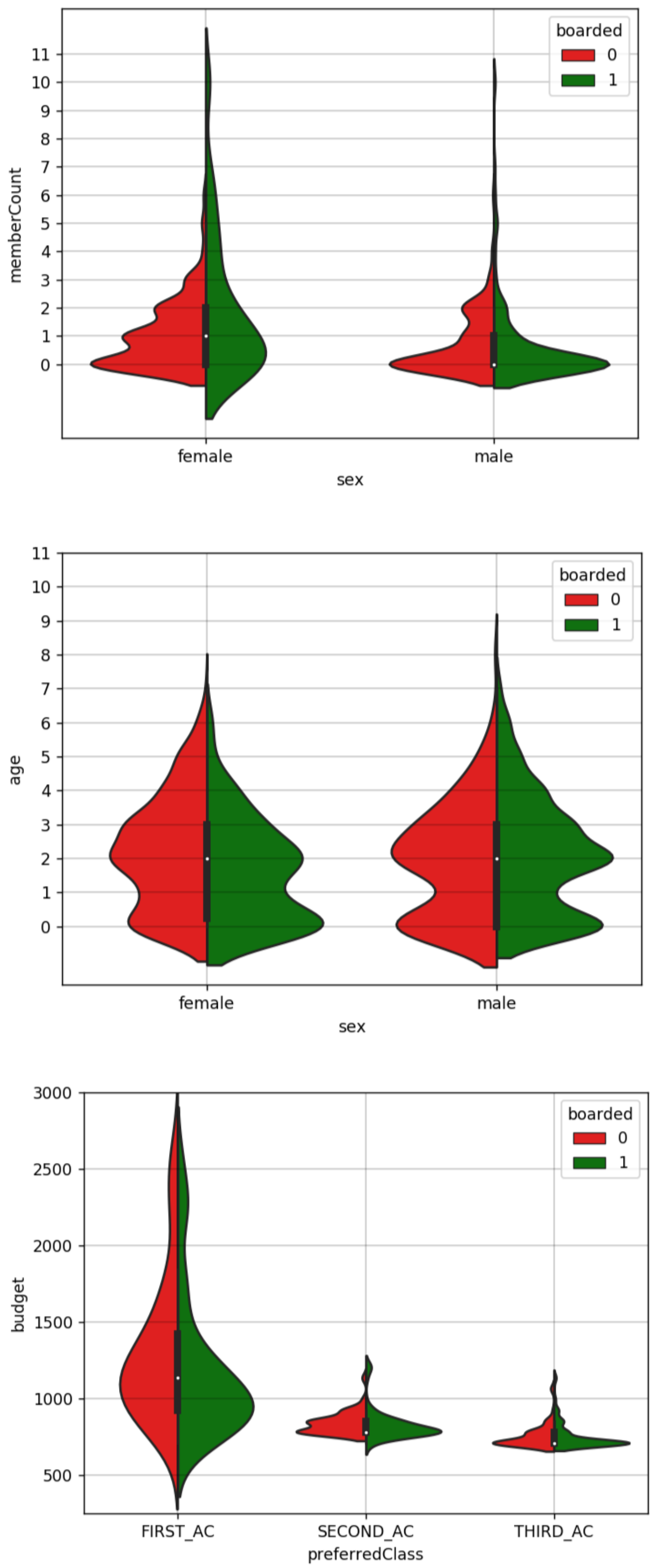

Fig. 15: Class Conditional Distribution of Features

Algorithm\Dataset	Train	Test
Naive Bayes MLE	77.92	83.96
Naive Bayes EM	40.06	21.37
Bayes MLE	77.75	83.96
Bayes EM	40.06	21.37
kNN (k = 23) Chebyshev distance		83.2
Parzen window (hypercube) h = 9		78.62
Parzen window (Gaussian) h = 2.84		78.62

Table 3: Performance

It is not possible to fit 2 or more Gaussians through the dataset because, the data is categorical with “sex” taking only 2 distinct values. If we try to apply Naive Bayes classifier with 2 Gaussians, we end up with the Gaussians converging to Dirac-deltas (σ converges to 0). If we try to apply Bayes classifier with 2 or more Gaussians, the covariance matrix (Σ) becomes singular.

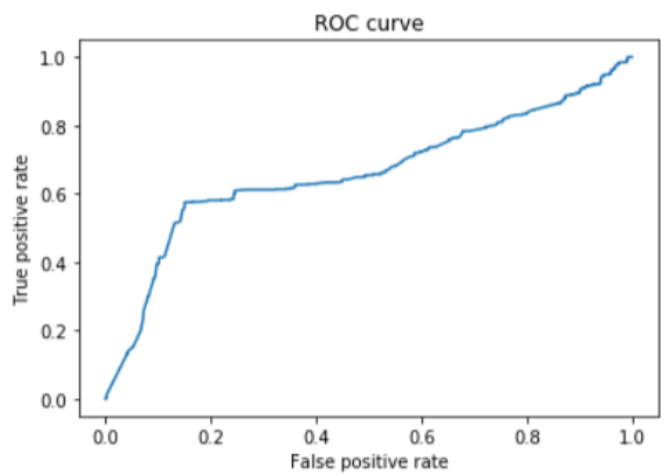

Fig. 16: ROC curves for Naive Bayes classifier

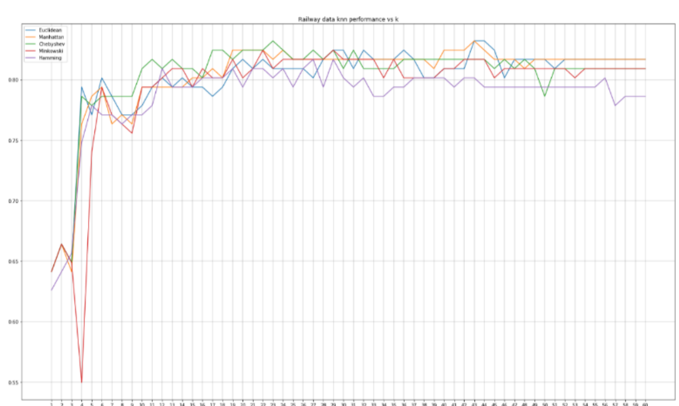

Fig. 17: (Test) Accuracy vs k for KNN

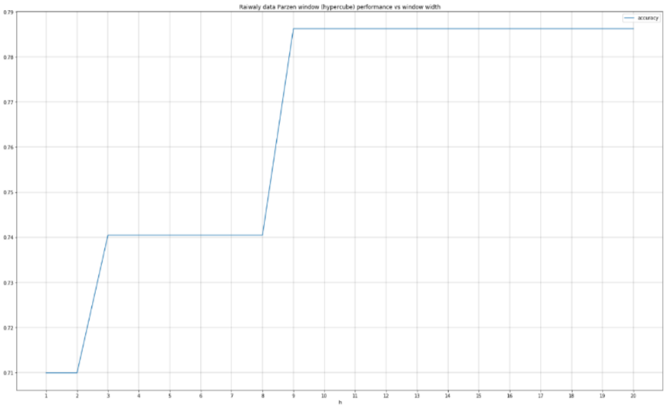

Fig. 18: Test accuracy vs h for Parzen window (hypercube window)

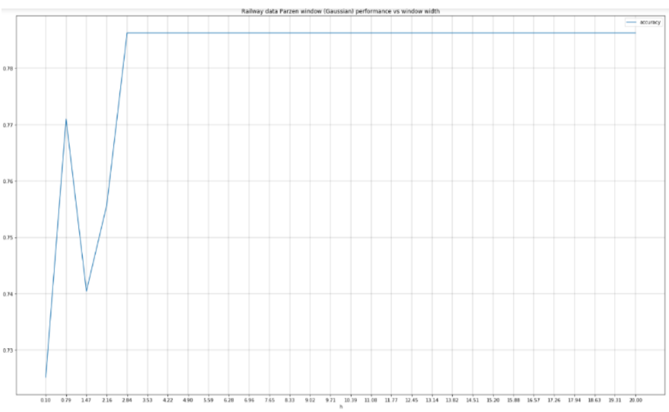

Fig. 19: Test accuracy vs h for Parzen window (Gaussian window)

The railway data is categorical, that is, on mapping numbers to the k distinct values taken by a particular feature, we can say that our dataset lies in k hyperplanes i.e the data will not be spread in the d dimensional space, but will have strict confinement on these hyperplanes. Thus, while applying Parzen window estimation technique, we see that the accuracy will have abrupt jumps and will finally achieve a steady state, because for larger values of window width the nearly the complete dataset is confined within the window and on increasing the window width further has no effect.

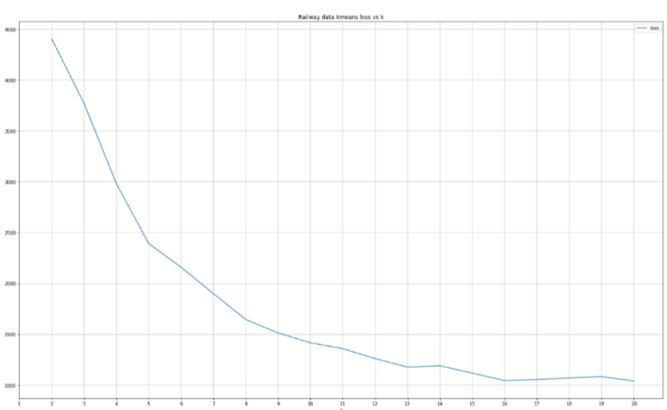

Fig. 20: Loss vs k for K-Means

For kmeans, knee-point occurs at k = 11.

Fashion MNIST Dataset

Data Visualization

Due to high dimensionality of data, we employ Principal Component Analysis to reduce dimensions from 784 to 186 retaining 95 % variance.

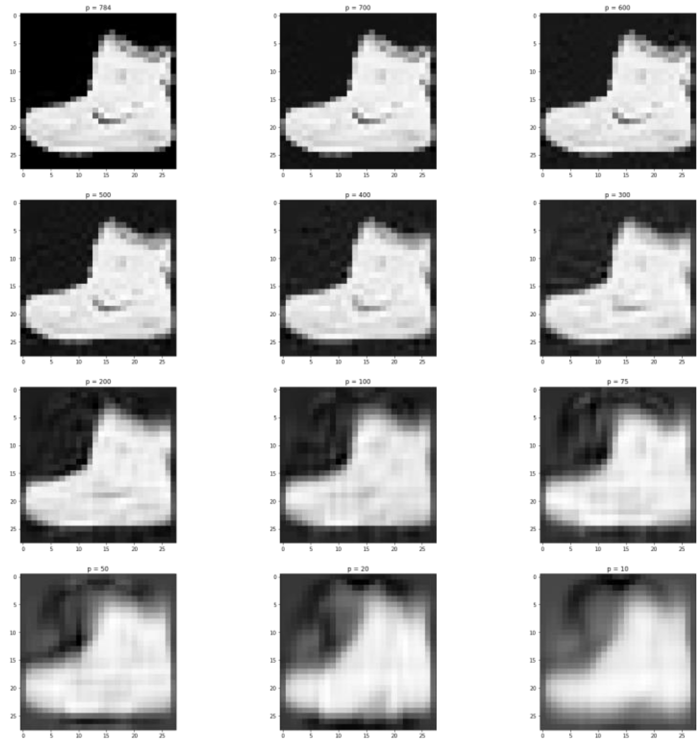

Fig. 21: PCA Reconstruction (for different p)

Here, Moore-Penrose pseudo inverse has been used to convert the data in the reduced dimensional-space to get data as close to the original as possible.

Algorithm\Dataset	Train	Test
Naive Bayes MLE	77.9	76.78
Bayes MLE	81.49	80.58
kNN (k = 9) Manhattan distance		86.15
Parzen window (hypercube) h = 720		84
Parzen window (Gaussian) h = 240		87

Table 4: Performance

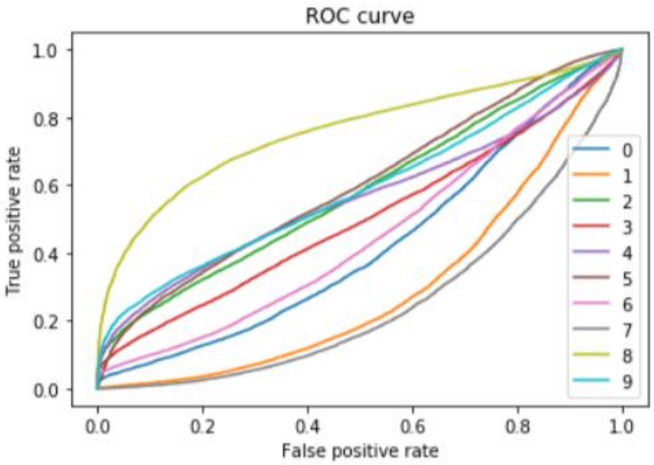

Fig. 22: ROC curves for Naive Bayes classifier

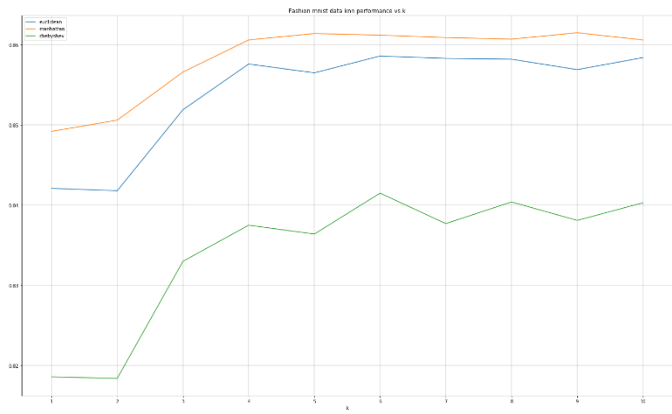

Fig. 23: Test accuracy vs k for KNN

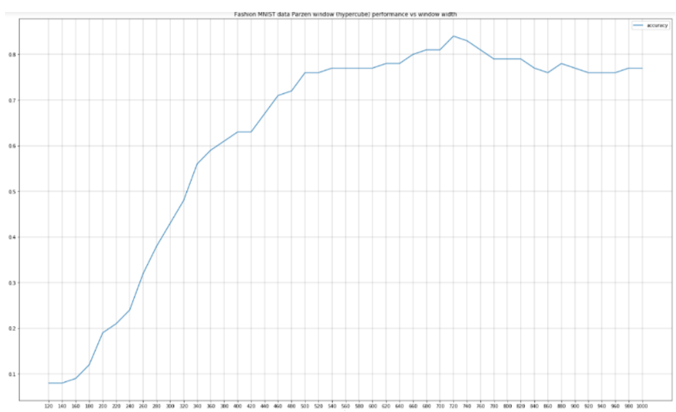

Fig. 24: Test accuracy vs h for Parzen window (hypercube window)

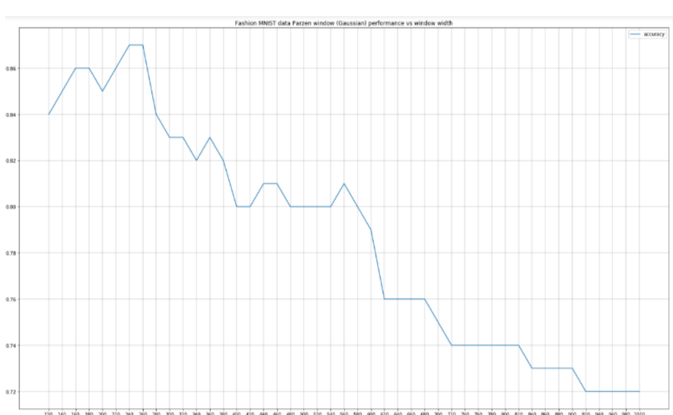

Fig. 25: Test accuracy vs h for Parzen window (Gaussian window)

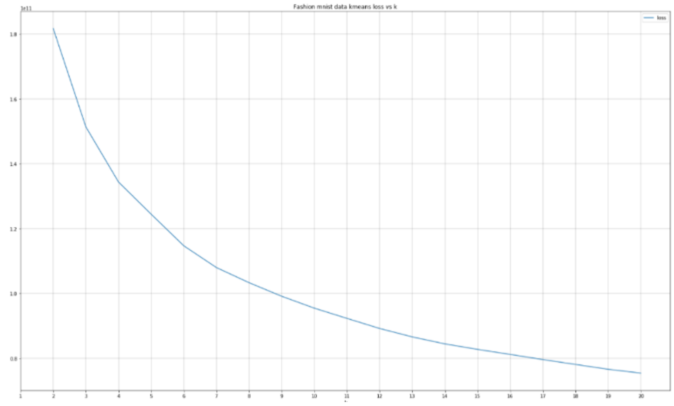

Fig. 26: Loss vs k for K-Means

Implementational Details

- We employed high dimensional tensors (wherever possible) for speedup by vectorized implementations.
- Tested Euclidean, Chebyshev, etc. metrics to optimize over indeterministic topology of the data space.
- Performed Z-scoring wherever it boosted performance.
- Detailed observations and plots submitted in the Jupyter Notebooks.

Discussion

- We learnt to deal with high dimensional data and catering to its high computational needs.
- Dealing with categorical data was one another interesting aspect of the assignment.
- We also learnt to deal with 3 and 4 dimensional tensors.