SML Assignment - 3

Report by Kaustav Vats (2016048)

Q1 DIMENSIONALITY REDUCTION

Evaluation Metrics for Dataset 1

Accuracy\Classifier	Gaussian Classifier	After PCA	After LDA
Train Accuracy	75.44	88.02	98.60
Test Accuracy	65.42	70.56	82.24

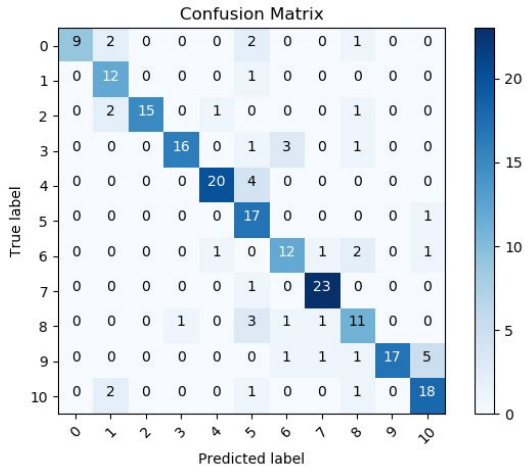
LDA + 5 Fold Cross Validation + Gaussian Classifier

Test Fold	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Test Accuracy	96.00	98.00	98.00	98.00	100.00

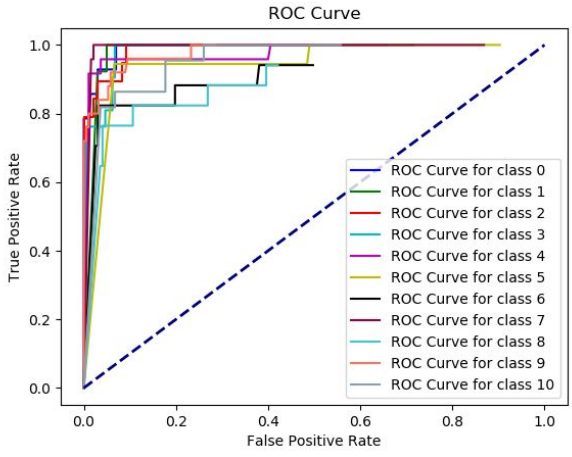
Data\Evaluation Metric	Mean Accuracy	Standard Deviation
Test Data	98.00	1.26

Confusion Matrix and Roc Curve For Best Model, Tested with original Split of the Test Data

Confusion Matrix



Roc Curve



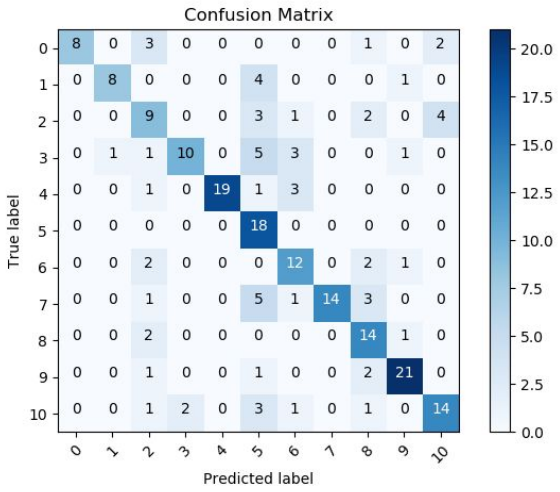
PCA[95] + 5 Fold Cross Validation + Gaussian Classifier

Test Fold	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Test Accuracy	83.00	83.00	77.00	80.00	78.22

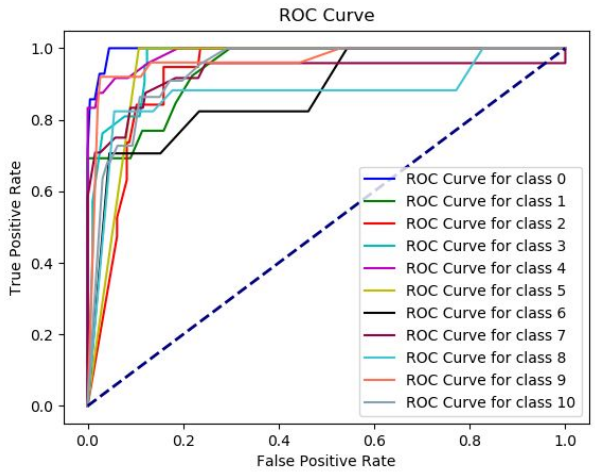
Data\Evaluation Metric	Mean Accuracy	Standard Deviation
Test Data	80.24	2.44

Confusion Matrix and Roc Curve For Best Model, Tested with original Split of the Test Data

Confusion Matrix



Roc Curve



PCA and LDA followed by 5 Fold Cross Validation Analysis

Performance	PCA	LDA
Mean Accuracy	68.03	78.79
Standard Deviation	1.24	1.58
Original Test data Accuracy	70.09	80.84

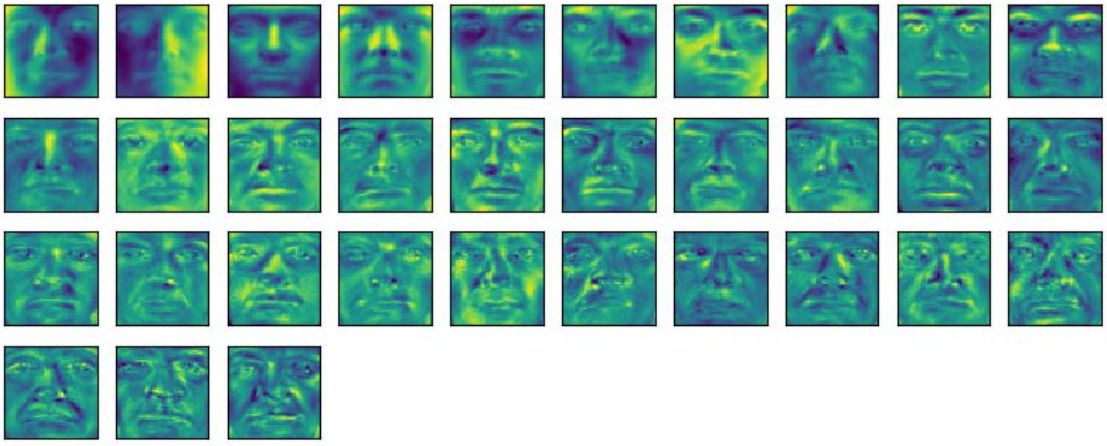
We can see from Roc Curves that LDA has found much better projection direction than PCA, therefore Accuracy with LDA is much better as compared to PCA. I also observed that number of features reduced for both were very different. I took 10 features in dataset 1 since there are only 11 classes and 33 features for PCA.

Eigenvectors Visualization

Top Eigenvectors with 95% Eigen Energy Conserved

Dataset - 1

We can see that eigen faces are very similar and changes are visible for each eigenvector. Face is clearly visible in first eigenvector. This produces dimension reduction by allowing the smaller set of basis images to represent the original training images



PCA with Different Eigen Energy Conserved

Test Accuracy	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Acc with 70% EE	13.00	12.00	9.00	8.00	9.90
Acc with 90% EE	67.00	71.00	61.00	70.00	71.29
Acc with 99% EE	74.00	84.00	78.00	84.00	83.17

Fold Test Data\Evaluation Metric	Mean Accuracy	Standard Deviation
PCA [70] Test Data	10.38	1.86
PCA [90] Test Data	68.06	3.84
PCA [99] Test Data	80.63	4.00

Result on Validation Test Data

Performance	PCA [70% EE]	PCA [90% EE]	PCA [99% EE]
Mean Accuracy	7.66	41.02	78.31
Standard Deviation	0.56	0.62	1.09
Original Test data Accuracy	8.41	42.06	79.44

Its clearly visible by the result that PCA performs better on taking more number of features, mean more eigen energy. Since its able to project to direction with more spread. On increasing eigen energy there's a drastic change in accuracy for the models.

Result for performing LDA after PCA & PCA after LDA

Taking PCA 99% EE

Accuracy\Combination	PCA then LDA	LDA then PCA
Train Accuracy	98.60	98.40
Test Accuracy	91.12	84.58

Both results are very close for Training Accuracy, whereas Test data accuracy is better for performing PCA followed by LDA. Comparing This result with LDA only, we can say that PCA + LDA is better combination than LDA followed by PCA and LDA alone for Dataset - 1

Dataset - 2 [CIFAR-10]

Evaluation Metrics for Dataset 2

Accuracy\Classifier	Gaussian Classifier	After PCA	After LDA
Train Accuracy	25.55	28.01	35.03
Test Accuracy	26.62	24.71	27.79

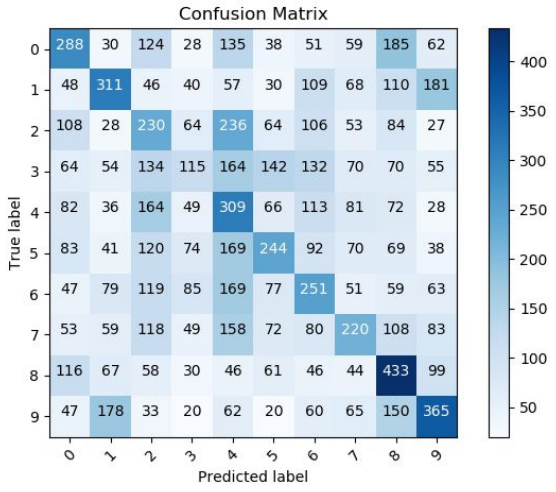
LDA + 5 Fold Cross Validation + Gaussian Classifier

Test Fold	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Test Accuracy	35.22	34.91	34.99	35.03	34.22

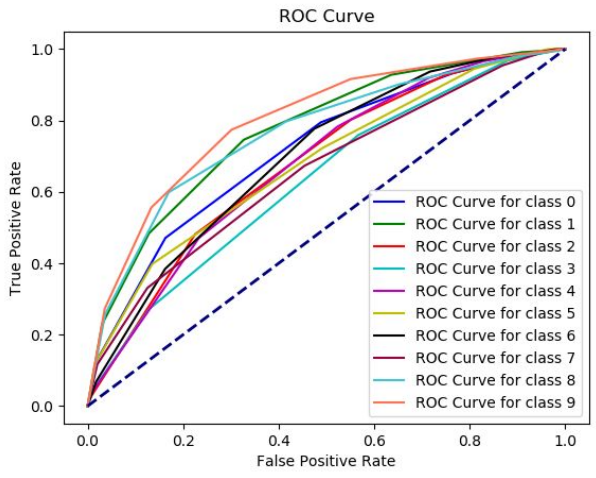
Data\Evaluation Metric	Mean Accuracy	Standard Deviation
Test Data	34.874	0.34

Confusion Matrix and Roc Curve For Best Model, Tested with original Split of the Test Data

Confusion Matrix



Roc Curve



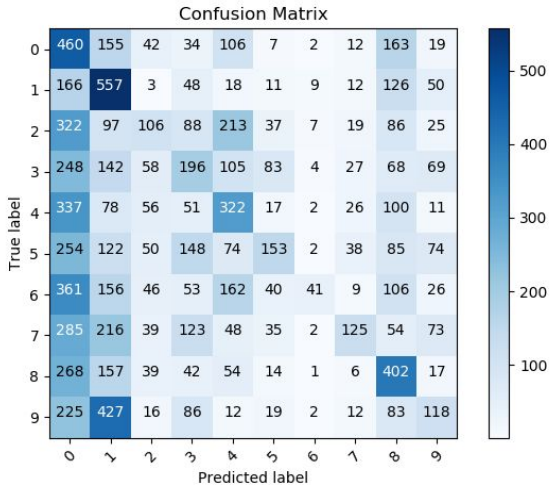
PCA[95] + 5 Fold Cross Validation + Gaussian Classifier

Test Fold	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Test Accuracy	26.82	27.33	27.62	27.22	27.16

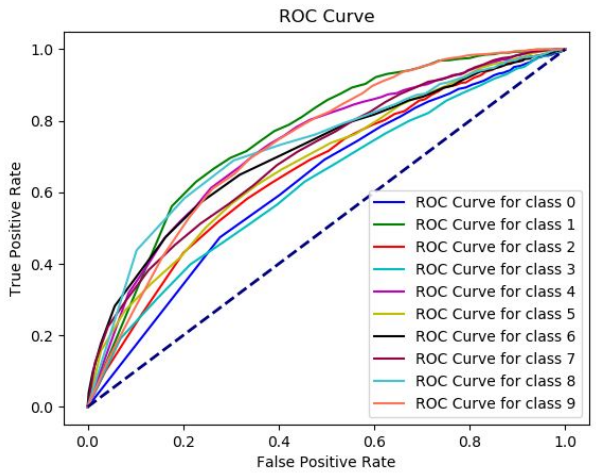
Data\Evaluation Metric	Mean Accuracy	Standard Deviation
Test Data	27.23	0.26

Confusion Matrix and Roc Curve For Best Model, Tested with original Split of the Test Data

Confusion Matrix



Roc Curve



PCA and LDA followed by 5 Fold Cross Validation Analysis

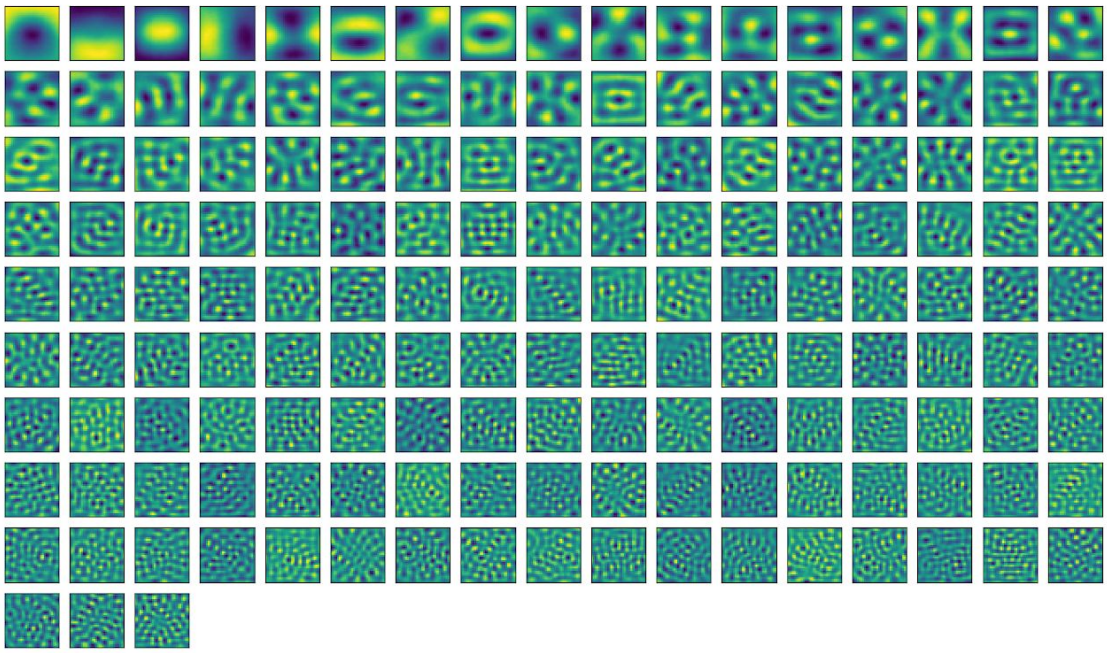
Performance	PCA	LDA
Mean Accuracy	24.72	27.74
Standard Deviation	0.14	0.09
Original Test data Accuracy	24.91	27.89

We can see from Roc Curves that LDA has found much better projection direction than PCA, therefore Accuracy with LDA is much better as compared to PCA. I also observed that number of features reduced for both were very different. I took 9 features in dataset 2 since there are only 10 classes and 156 features for PCA.

Eigenvectors Visualization

Top Eigenvectors with 95% Eigen Energy Conserved, Dataset - 2

We can see that eigen faces are very similar and changes are visible for each eigenvector. This produces dimension reduction by allowing the smaller set of basis images to represent the original training images



PCA with Different Eigen Energy Conserved

Test Accuracy	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Acc with 70% EE	24.49	29.30	29.83	30.15	29.87
Acc with 90% EE	29.12	29.50	28.05	29.29	28.23
Acc with 99% EE	27.70	27.46	28.06	26.34	28.23

Fold Test Data\Evaluation Metric	Mean Accuracy	Standard Deviation
PCA [70] Test Data	29.73	0.30
PCA [90] Test Data	28.84	0.59
PCA [99] Test Data	27.56	0.67

Result on Validation Test Data

Performance	PCA [70% EE]	PCA [90% EE]	PCA [99% EE]
Mean Accuracy	16.24	21.30	25.85
Standard Deviation	0.10	0.10	0.46
Original Test data Accuracy	16.41	21.43	26.45

Its clearly visible by the result that PCA performs better on taking more number of features, mean more eigen energy. Since its able to project to direction with more spread. On increasing eigen energy there's not much change in the accuracy, Accuracies are slightly increased on increasing number of projection vectors.

Result for performing LDA after PCA & PCA after LDA

Taking PCA 99% EE

Accuracy\Combination	PCA then LDA	LDA then PCA
Train Accuracy	32.86	35.03
Test Accuracy	18.91	18.86

Analysis for LDA + 5 Fold Cross Validation For both datasets

LDA has performed better in Dataset 1 as compared to dataset 2. It seems that data was still not separable after performing LDA. It seems that no of dimensions need to be increased for Dataset - 2 instead of reducing no of features and finding best projection. For Dataset 1, It was much better separable than Dataset2. It concluded that, LDA doesn't work always, sometime we need to add more features which make it more separable than before.

Analysis for PCA + 5 Fold Cross Validation For both datasets

PCA has performed better for Dataset 1 as compared to Dataset 2. Since PCA was able to project much better in dataset 1. We can see that mean accuracy for test data are different. PCA has shown better result than LDA in 2nd dataset.

PCA then LDA or LDA then PCA

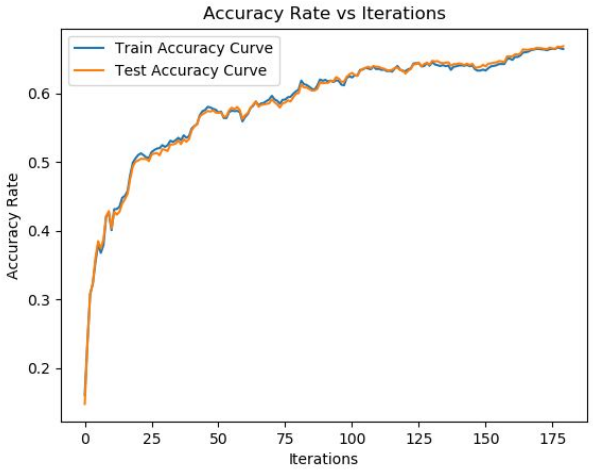
PCA + LDA was better for dataset 1, where as First LDA then PCA has show increase in train accuracy but much better accuracy achieved in PCA + LDA. It concluded that PCA then LDA is much better approach for both the dataset. It's not always true, sometimes PCA can perform better than LDA because LDA make assumption about normally distributed classes and equal class covariance. For some dataset, above orders can be affected, thus affecting overall accuracy of the model.

Q2 ENSEMBLE LEARNING

Evaluation Metrics for Dataset 1

Accuracy\Classifier	Decision Tree	AdaBoosting	Bagging
Train Accuracy	16.10	66.51	-
Test Accuracy	14.78	66.92	27.1

AdaBoost on Letter Recognition Dataset



Above image is result for optimal number of iteration. Model was overfitting on increasing number of iterations

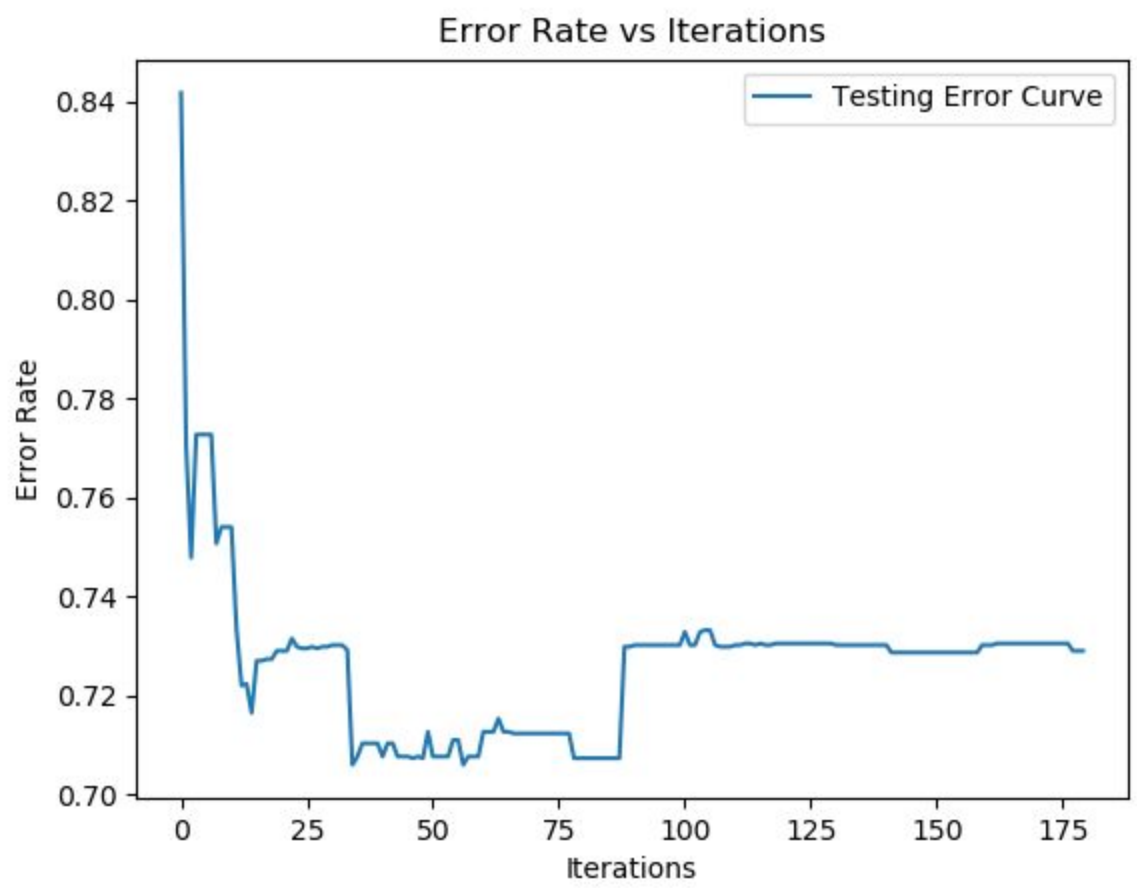
Test Accuracy	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Test Fold Acc	59.50	67.29	64.92	63.64	65.17

Fold Test Data\Evaluation Metric	Mean Accuracy	Standard Deviation
Test Fold Acc	64.11	2.58

Result on Validation Test Data

Performance	AdaBoosting
Mean Accuracy	63.95
Standard Deviation	2.34
Original Test data Accuracy	67.28

Bagging on Letter Recognition Dataset



In the above image you can see that after iteration 75, model overfits thus Error rate increases.

Test Accuracy	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Test Fold Acc	29.14	29.39	25.43	24.54	26.71

Fold Test Data\Evaluation Metric	Mean Accuracy	Standard Deviation
Test Fold Acc	27.04	1.95

Result on Validation Test Data

Performance	Bagging
Mean Accuracy	28.23
Standard Deviation	1.69
Original Test data Accuracy	29.75

Score Normalization

Score Normalization	Equal Weight Max Vote	Min-Max	Z - Score	Tanh
Test Data Accuracy	29.75	28.60	28.47	28.53

Z - Score: It consider mean and variance of raw data. It assumes data distribution to be normal. Since the result that were obtained were from different distribution, that's why z-score is less than score obtained in equal weight max vote. Data doesn't really follow normal distribution
Tanh: It's usually used to improve performance by giving some importance some data point for which classes that are close to be classified.
Min-Max: This is basically used to normalize score by subtracting it with minimum value of score and dividing by difference between max value and min value. This normalize score in range [0, 1]
Accuracy for all normalization techniques are close to ensemble learned model.

Accuracy for Decision Tree is very less for both train and test data, whereas after applying Adaboost on the original data. I observed a large change in Train and Test accuracy for optimal number of model combined for adaboosting. It means that data was **biased**. Using Adaboosting on biased data, i was able to increase accuracy and remove bias from the data. If we perform bagging, we can clearly see that data had some variance which was handled by bagging but bias was much higher. We can also create a combined model of bagging and adaboosting obviously with some weightage to each model.