Machine Learning for Computer Vision (EE5177)
Programming Assignment 1 : Fitting Probability Distribution to Data
Problem #2

Akshit Kumar
EE14B127

7th February 2017

Contents

1 Problem

We are given a data set consisting of face and non-face images. The task is to classify a new image as a face or a non face.

1.1 Goal

For a given test data, we are required to classify the test data to belong to either a face class or a non face class.

1.2 Challenges

Learning a model and doing classification on all 784 dimensions (28x28 greyscale images) is computationally expensive.

1.3 Approach

To classify a test data to either class of faces or non faces, we make use of Generative Classification Technique by learning the class conditional probability distributions for the face class and non face class. We train the data to fit Multivariate Normal Distribution for face and non face models using Maximum Likelihood Estimation. Depending on which model gives a better loglikelihood score, we classify the test data into the respective class.
To handle the computational infeasibility, we use Principal Component Analysis (PCA) for dimensionality reduction to say some K dimensions ($K < 784$). To do this, we do eigen value decomposition of the covariance matrix of the training data then order the eigen values from highest to lowest. Now we select the eigen vectors corresponding to the top K eigen values and project the data along these selected eigen vectors which give project the data along the plnae of the highest variance giving the final dimension reduced data. For the test data also, eigen vectors of the corresponding training data are used for reduction.
Note : In this implementation, I used the eigen vectors of the face training data to project all my other train and test data sets to that of the eigen vectors of the face.
For getting the best model to choose from, the training set of faces and non faces is divided into different sets and different models are learnt on these training sets of different sizes. The classification results of these models are plotted and the model giving the highest classification accuracy is used.
We then visualise the likelihood score given by the best model to the test data sets and decide threshold for using the face model alone versus non face model alone but plotting the ROC plots.

2 Results

2.1 Classification Accuracy

To find the best model in terms of number of faces and non faces, we plot the classification accuracy(A)which is defined as

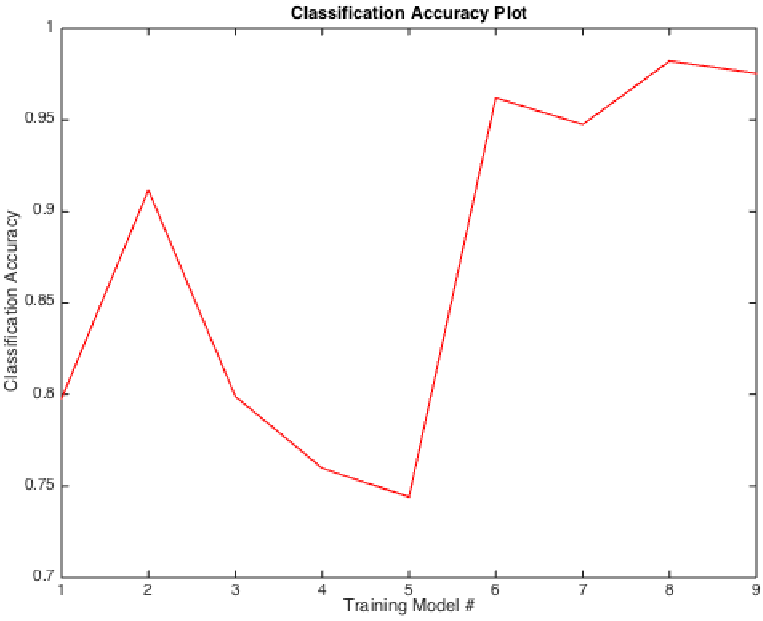
$$A = \frac{tp + tn}{tp + tn + fn + fp}$$

where the confusion matrix is defined as follows:

	Faces	Non Faces
Faces	True Positive (tp)	False Negative(fn)
Non Faces	False Positive (fp)	True Negative (tn)

Tabulation of the confusion matrix is as follows :

(Faces,Non-Faces)	tp	fn	fp	tn	A
(150,150)	493	7	174	221	0.7977
(150,200)	444	56	23	372	0.9117
(200,200)	497	3	177	218	0.7988
(250,200)	500	0	215	180	0.7597
(300,200)	499	1	288	167	0.7441
(300,300)	492	8	26	369	0.9620
(400,300)	493	7	40	355	0.9474
(400,400)	493	7	40	355	0.9821
(600,400)	489	11	11	384	0.9754



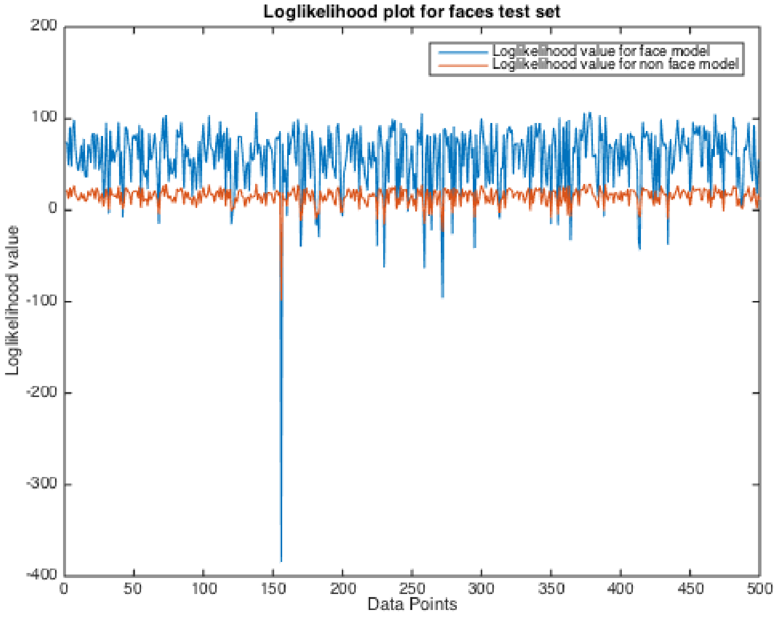
From this plot we see that the second last model (# of faces = 400 and # of non faces = 400) gives the best classification accuracy amongst all the models.

2.2 Visualization of Loglikelihood Scores

Given that the best model according to the classication accuracy plot is model #8. We visualise the loglikehood scores given by this model for faces and non faces data set.

2.2.1 Loglikelihood Scores On Faces Test Set

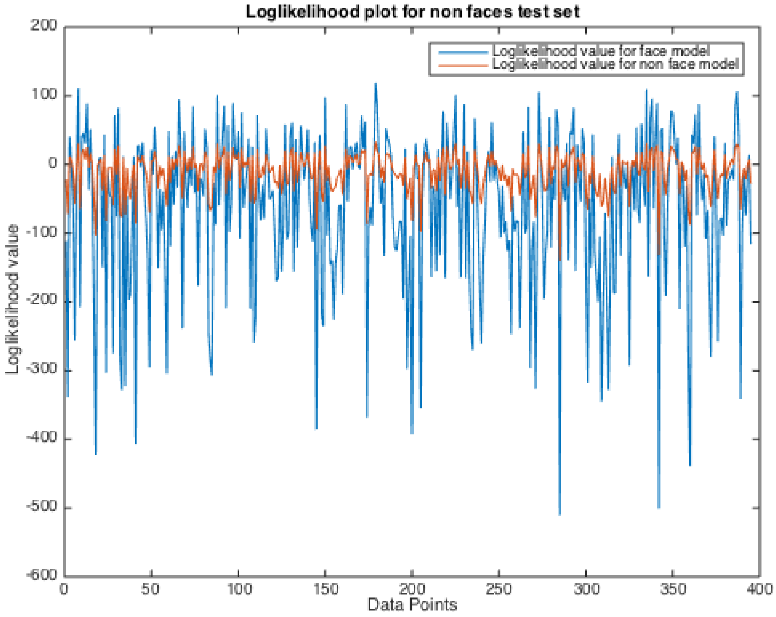
For the model learnt on the training set of faces and non-faces and then tested on face test sets, gives the following plot for loglikelihood scores.



To get the threshold for, we set the lower threshold to -50 and increase the threshold to +50 in steps of +1.

2.2.2 Loglikelihood Scores On Non Faces Test Set

For the model learnt on the training set of faces and non faces and then tested non-face test sets, gives the following plot for loglikelihood scores.



To get the threshold for, we set the lower threshold to -100 and increase the threshold to +100 in steps of +1.

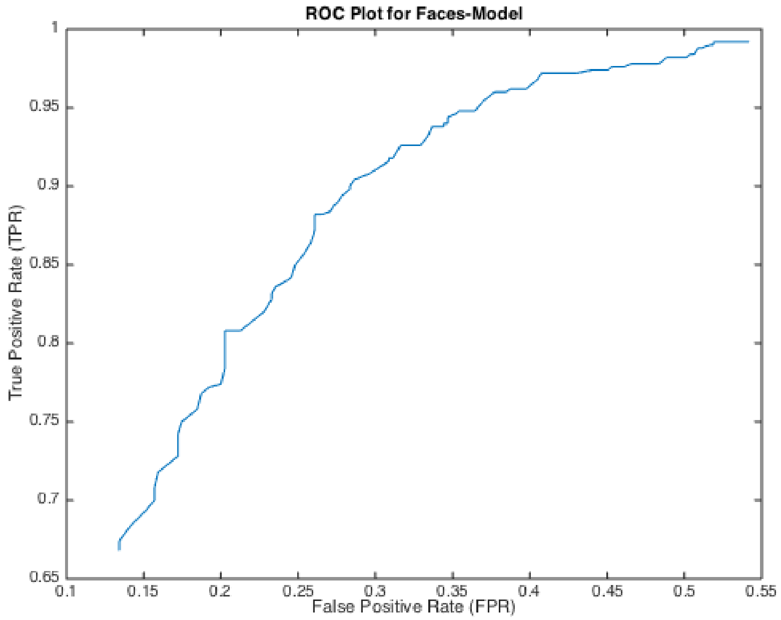
2.3 ROC Curve Plots for Face and Non-Face Model

By iterating through the different values of the threshold, we get different tp and tn and hence we plot the True Positive Rate (TPR) vs False Positive Rate (FPR) which is defined as follows :

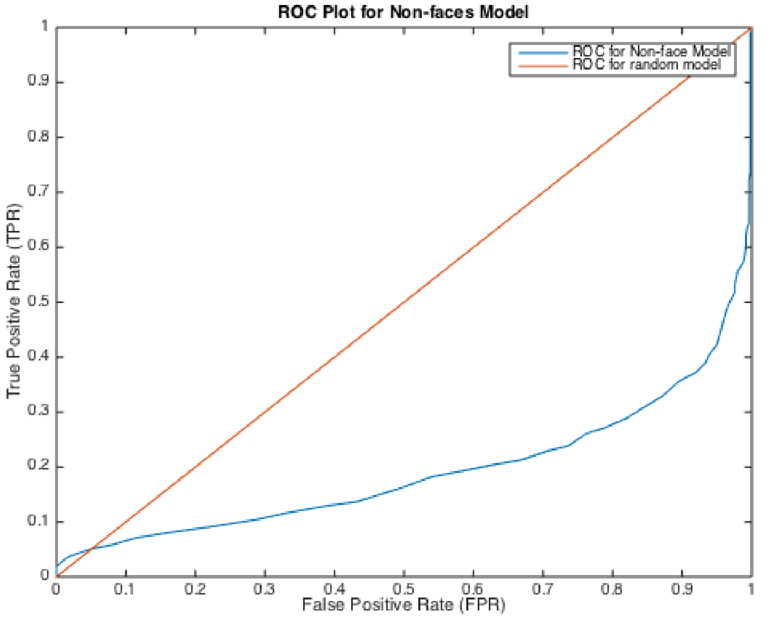
$$TPR = \frac{tp}{tp + fn}$$

$$FPR = \frac{fp}{fp + tn}$$

2.3.1 ROC curve using face model as predictor



2.3.2 ROC curve usinf non face model as predictor



2.4 Threshold for FPR ≤ 0.2

We find that the loglikelihood score threshold for the faces model is 38 and for the non faces model is 24 such that FPR ≤ 0.2

3 Inferences

3.1 Suitable model for classification task

By performing this experiment, we realise that using the faces model is better suited than using the non faces model for classification tasks because the ROC curve for faces model is always above the ROC curve for non-faces model. Also we notice that the ROC curve for the non-faces model is below the $y = x$ line which is essentially the ROC curve we get by randomly classifying the images as either face or non face. Hence the non-face model performs worse than a random classifier.
Also, intuitively speaking using the faces-model as a classifier also makes sense because faces will have a lot of features common amongst themselves which is not the case with non-faces.