North Carolina State University

ECE 763 Computer Vision

Project Report:

Face Image Classification

Author: Abhishek (200205655)

Instructor: Dr. Tianfu Wu

Objective:

In face detection, we attempt to infer a discrete label indicating whether a face is present or not based ε {0, } ω 1 on observed image data. Through this project I aim to achieve face image classification using Gaussian model, Mixture of Gaussian model, t-distribution, Mixture of t-distribution, and Factor Analysis. For each classification model the results are reported as follows:

1. Visualization of the estimated mean(s) and covariance matrix for face and non-face respectively.
2. Evaluation of the learned model on the testing images using 0.5 as threshold for the posterior by computing the false positive rate, false negative rate and the misclassification rate.
3. Plot of the Receiver Operating Characteristic Curve (ROC Curve) to evaluate the diagnostic ability of the binary classifier.

Dataset Preparation:

The Image Database used for this project is a subset of the Face Detection and Data Set Benchmark (FDDB) http://vis-www.cs.umass.edu/fddb/. The Database consists of RGB images with annotations for the face bounding boxes. I extracted 1000 training images and 100 testing images from the dataset through careful filtering to avoid anomalies such as spectacles. The annotations provided in the form of rectangular coordinates are used for the extraction of the face boundary from the images. The non-face image data set is created using the same extracted images by carefully cropping the background while ensuring no or minimal overlap exists between the annotated face and background section. All the testing and training images are resized to a 20 x 20 dimension. It is ensured that training face images and testing face images are separate, that is, no face testing images are from the same person in the training set of face images. The resulting dataset consists of 2000 training and 200 test images. A sampling of the dataset is depicted below. To recognize if a provided image patch contains a face or not face we first concatenate the RGB values to form a 1 x 1200 vector x.

Sample:

Figure 1. a) Sample Image from FDDB Database b) Extracted image patch from annotation.

Model 1: Single Gaussian Model:

The single Gaussian model also known as the multivariate gaussian mode is a generative approach to face detection in which we calculate the probability of the data x and parameterize this by the world state ω . The data is described with a multivariate normal distribution as given in equation (1).

$$Pr(x/\omega) = Norm_x[\mu, \Sigma] \qquad (1)$$

The parameters of this model $\Theta = \{ \mu_0 , \Sigma_0 , \mu_1 , \Sigma_1 \}$ include the mean and covariance for the face image and non face image normal distributions. Since the full covariance matrix contains $D(D + 1)/2$ parameters which we cannot uniquely specify with a training data set of just 1000 images, we use the diagonal form of the covariance matrix.

We use the maximum likelihood approach as given in equation (2) to obtain the mean μ_0 and covariance Σ_0 concerned with the background regions from the subset of training data with non-face images (S_0). Similarly. μ_1 and covariance Σ_1 are concerned exclusively with faces and can be learned from the subset of training data which contains faces (S_1) using equation (3). Figure

2 shows the maximum likelihood estimates of the parameters with the diagonal form of the covariance matrix.

$$\widehat{\mu_0}, \ \widehat{\Sigma_0} \ = armax \, [\prod_{i \varepsilon s_0} Pr(x_i/\mu_{0,}\Sigma_0)] \qquad (2)$$

$$\widehat{\mu_1}, \ \widehat{\Sigma_1} \ = armax \, [\prod_{i \varepsilon s_1} Pr(x_i/\mu_{1,}\Sigma_1)] \qquad (3)$$

The mean of the face model clearly captures classified information. The covariance of the face is larger at the edges of the image, which usually contains hair or background. On evaluating the learned model on the testing images using 0.5 as the threshold for the posterior we obtain the false positive rate, the false negative rate and the misclassification rate as given in Table 1.
I was able to implement this model without running into computation hiccups such as Singular Matrix and Determinant calculation opposed to the later models. Hence the single gaussian model was tested on RGB images whereas the latter models were tested on images converted to the grey scale.

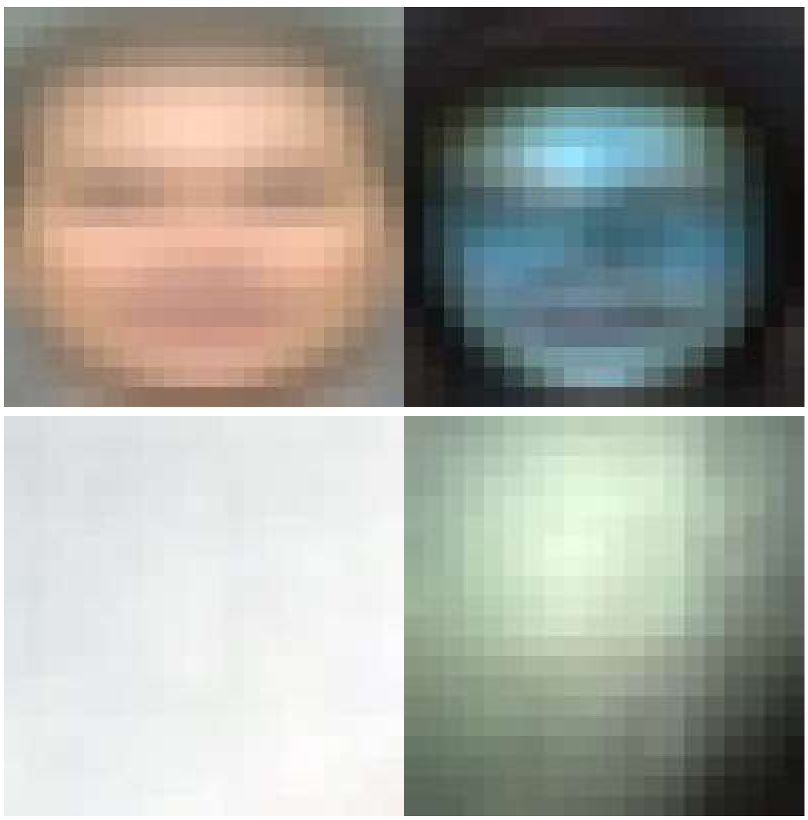
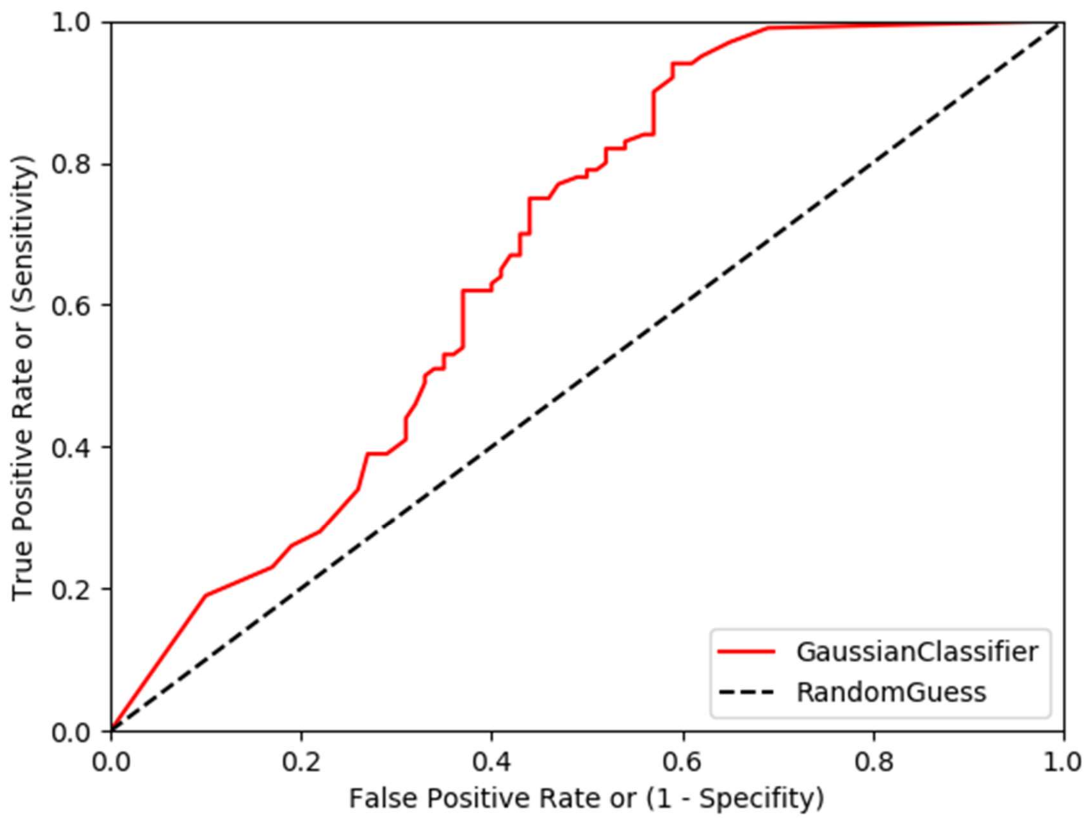


Figure 2. Maximum Likelihood fit based on 1000 training examples per class. a) Mean from face data b) Covariance from face data c) Mean from non-face data d) Covariance from non-face data. The background model has little structure: The mean is uniform and the variance is high everywhere.

Evaluation criteria	
Misclassification Rate	**0.35**
False Negative Rate	**0.37**
False Positive Rate	**0.47**

The Receiver Operating Characteristic Curve (ROC Curve) is a graphical plot that illustrates the diagnostic ability of a binary classifier system.The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The true-positive rate is also known as sensitivity. The false positive rate, also known as fall-out can be calculated as (1- specificity). The ROC curve for the simple gaussian model is given Figure 3.

Figure 3. ROC Curve for the Single Gaussian Model

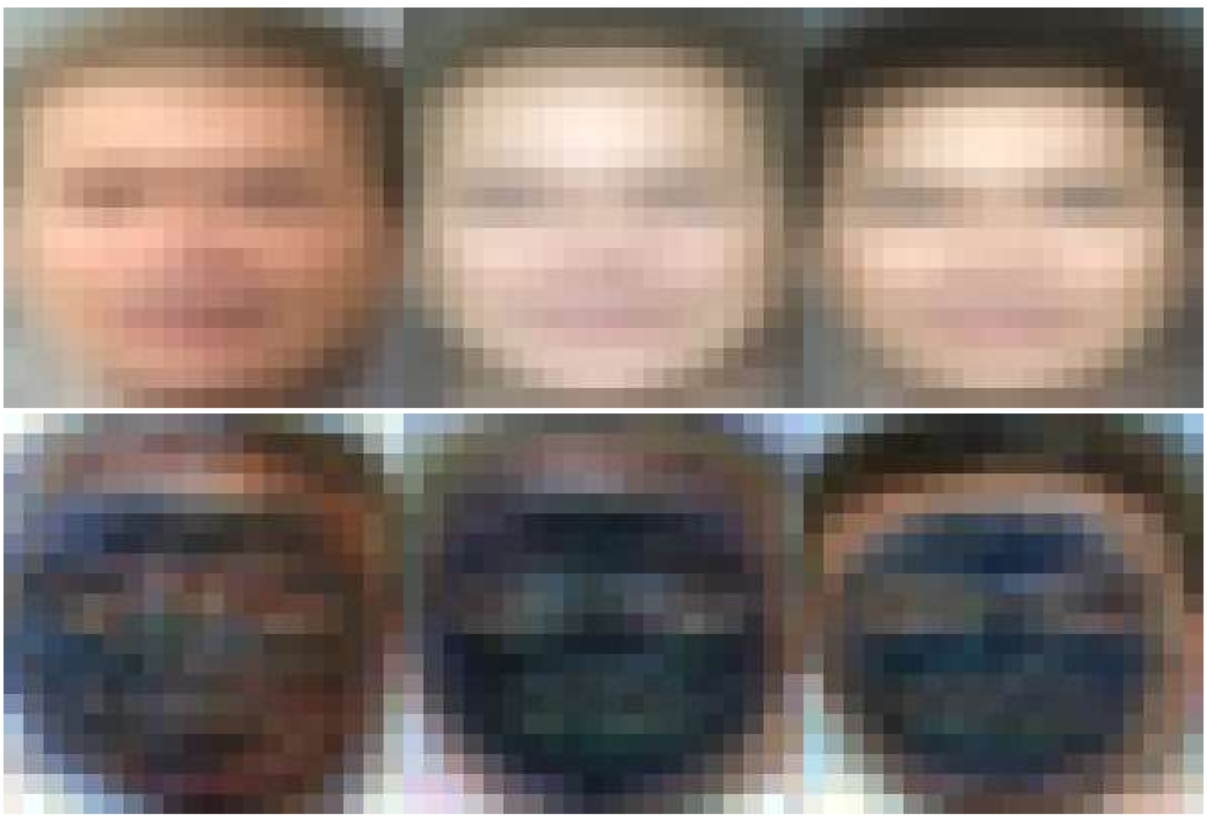


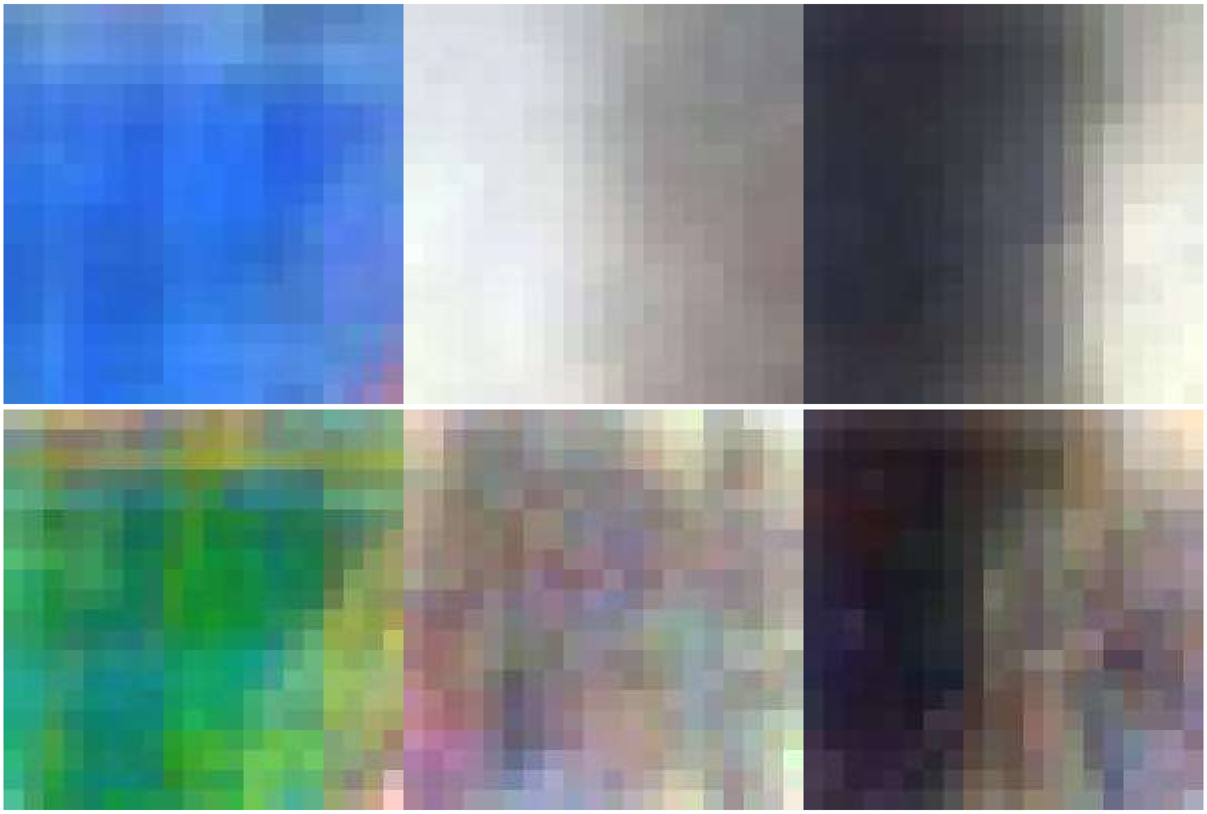
Model 2: Mixture of Gaussian Model:

The normal distribution is unimodal and neither the faces nor the background regions are well represented by a pdf with a single peek. To make the density multimodal we introduce mixture models. In the Mixture of Gaussians (MoG) model the dataset is described as a weighted sum of K normal distributions.

$$Pr(x/\omega) = \sum_{k=1}^{K} \lambda_k Norm_x[\mu_k , \Sigma_k] \qquad (4)$$

It is not possible to learn the parameters of the mode using the maximum μ , , } $for\ k$ 1...K Θ = { $_k$ Σ_k λ_k = likelihood approach since we cannot solve the resulting equations in a closed form. Instead, we express the observed density as a marginalization and use the EM algorithm to learn the parameters. In the E-step we maximize the bound with respect to the posterior probability distribution of each hidden variable give the observation and current parameter setting. We compute the probability that the normal $r(h /x ,) P_i = k_i \theta^{(t)} k^{th}$ distribution was responsible for the data point. In the M -step we maximize the bound with respect to the i^{th} model parameters. The E am M steps are alternated until the bound on the data no longer increases and the parameters no longer change. On applying the MoG model on the face data we obtain the mean vectors as shown in Figure 4. The weight for each component is indicated by the number. λ_k





The model captures variation in orientation of the face (pose) as well as the mean luminance. On evaluating the learned model on the testing images using 0.5 as the threshold for the posterior we obtain the false positive rate, the false negative rate and the misclassification rate as given in Table 2.

Evaluation criteria	
Misclassification Rate	**0.015**
False Negative Rate	**0.03**
False Positive Rate	**0.0**

Table 2. Evaluation of Mixture of Gaussian Model

We notice that the misclassification rate for the MoG model is much lower in comparison to the single gaussian model. The ROC curve for the simple gaussian model is given Figure 5.

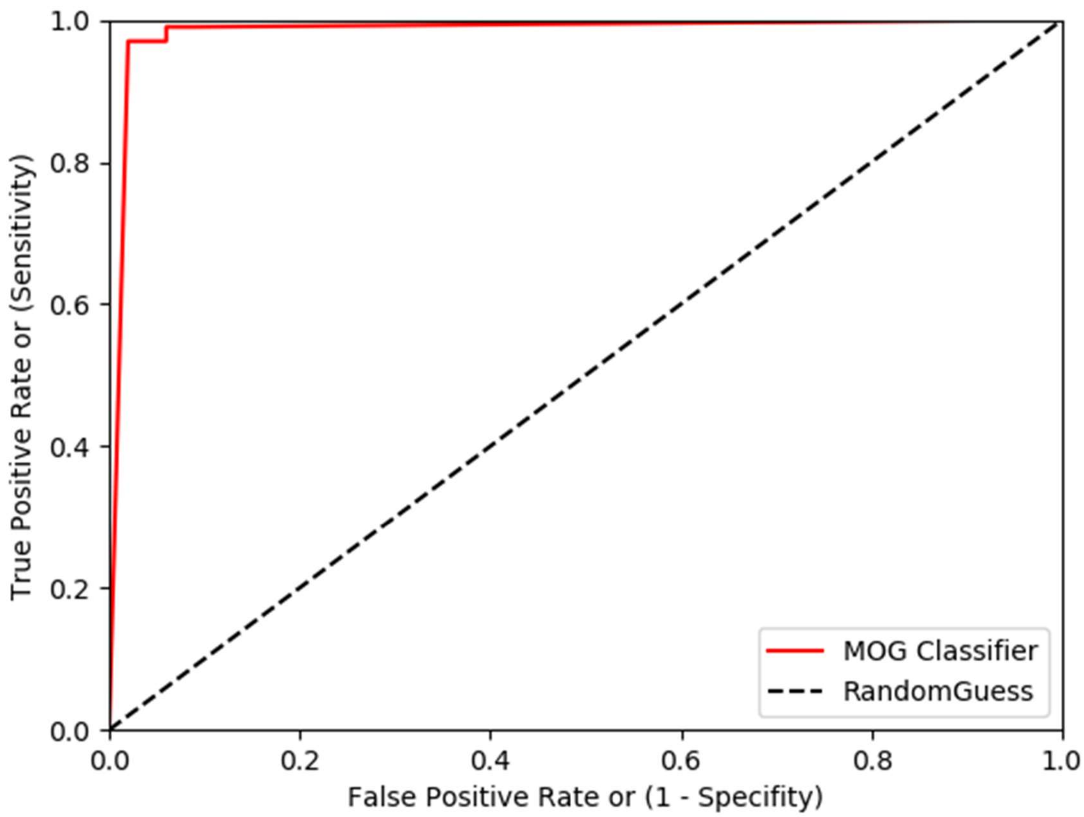


Figure 5. ROC Curve for the Mixture of Gaussian Model

Model 3: T - Distribution Model:

The normal distribution is not robust. The height of the normal pdf falls off very rapidly as we move into the tail due to which a single outlier can dramatically affect the estimates of the mean and the covariance of the classifier model. To make the density robust, we replace the normal distribution with the t distribution since the length of the tail is parameterized. The data is described using the probability density function of a univariate t distribution which has the parameters where the degree of freedom controls the length of the tails: when μ , , } Θ = { Σ ν ν ν is small there is considerable weight in the tails. As tends to infinity the distribution approximates a normal ν more closely and there is less weight in the tails.

$$Pr(x/\omega) = Stud_x[\mu, \Sigma, \nu] \qquad (5)$$

The t distribution is the marginalisation of the joint distribution between the observed variable x and a $r(x,)$ P h hidden variable h. The prior distribution over the hidden variable h has a gamma distribution. The conditional distribution with a variance that depends on h. So the t distribution can be considered as an infinite $r(x/h)$ P weighted sum of normal distributions with variance determined by the gamma prior. We use the EM algorithm to learn the parameters from the training data. In the E step we treat each data point as if it were generated from x_i one of the normals in the mixture where the hidden variable determines which normal. h_i

The E-step computes a distribution over which in turn determines which normal created the data. Outliers in h_i the data set will be explained best by the normal distribution with large covariances: for these distributions is h small. There is no closed form solution for the degree of freedom and hence we perform a one dimensional line ν search to maximize the bound.

When we fit the face data set to the t distribution with a diagonal scale matrix, the mean and covariance look visually similar to those for the normal model as shown in figure 6.

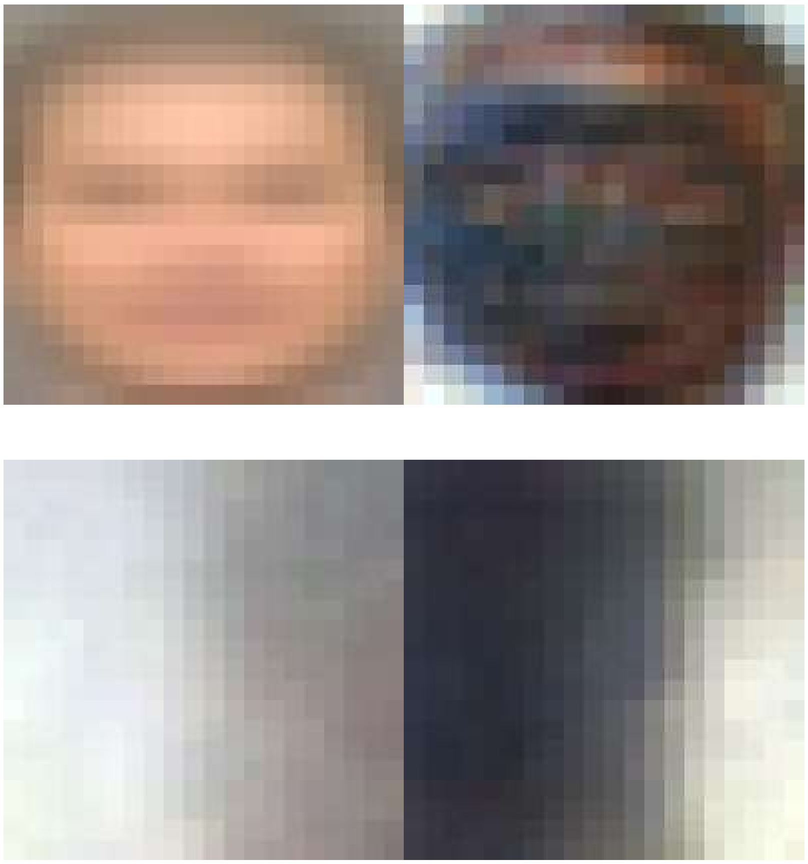


Figure 6. T distribution model for face dataset fitted on 1000 training examples per class. a) Mean from face data b) Covariance from face data c) Mean from non-face data d)Cov from non facedata

On evaluating the learned model on the testing images for face classification using 0.5 as the threshold for the posterior we obtain the false positive rate, the false negative rate and the misclassification rate as given in Table 3. The ROC curve for the simple gaussian model is given Figure 7.

Evaluation criteria	
Misclassification Rate	**0.22**
False Negative Rate	**0.32**
False Positive Rate	**0.475**

Table 3. Evaluation of T - Distribution Model

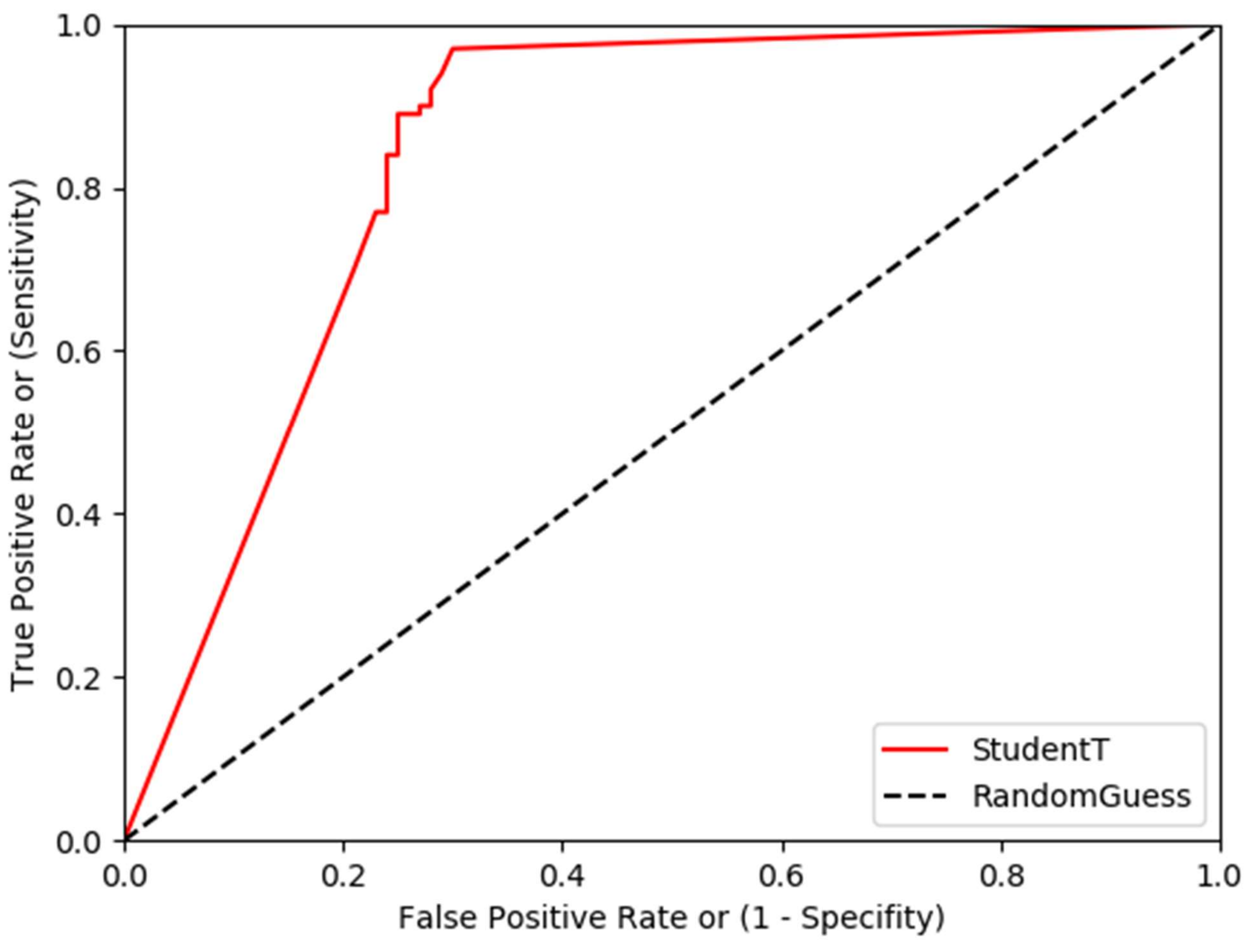


Figure 7. ROC Curve for the T-Distribution Model

Model 4: Factor Analysis Model:

Visual data are often very high dimensional. In the face detection task, data comes in the form of a 60 x 60 x 3 = 10800 dimensional vector. To model the data with a full multivariate normal distribution we require the covariance matrix of dimension 10800 x 10800 for which we would need a very large number of training examples to get a good estimate of all these parameters in the absence of prior information. Memory for storing the parameters as well as inversion of the full covariance matrix are additional problems. Factor analysis provides a compromise in which the covariance matrix is structured so that it contains fewer parameters than the full matrix but more than the diagonal form. It describes a linear subspace with a full covariance model. The data is given a probability density given by equation (6).

$$Pr(x/\omega) = Norm_x[\mu, \phi\phi^T + \Sigma] \qquad (7)$$

The covariance matrix contains a sum of two terms where describes a full covariance model over $\varphi \varphi^T + \Sigma \varphi \varphi^T$ the subspace and is a diagonal matrix that accounts for all remaining variations. Replacing the MoG model with Σ a infinite sum over a continuous family of Gaussians, each of which is determined by a certain value of h. If we choose the prior over the hidden variable to be a normal we obtain the equation for factor analyser. We describe the variance in a set of directions in a high dimensional space. The different factors encode the $\varphi \varphi \} \varphi = \{ _{1....k}$ different modes of the variation of the data set which has real world interpretations such as changes in pose or lighting.

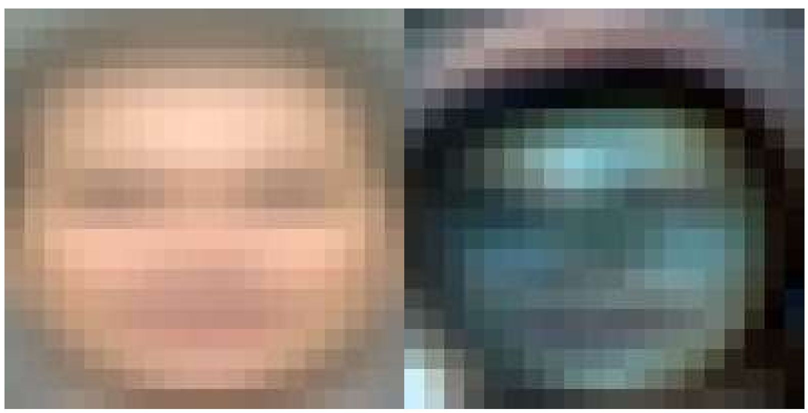


Figure 10. a) Mean for face model. b) Covariance component for face model.

Factor analyzer is an efficient model for capturing the covariance in high dimensional data. Figure 11 represents the mean for face model and the diagonal covariance component for face model. To visualize the effect of μ Σ the first factor we add or subtract or a multiple of it from the mean φ_1 . μ

On evaluating the learned model on the testing images for face classification using 0.5 as the threshold for the posterior we obtain the false positive rate, the false negative rate and the misclassification rate as given in Table 5. The ROC curve for the simple gaussian model is given Figure 12.

Evaluation criteria	
Misclassification Rate	**0.325**
False Negative Rate	**0.42**
False Positive Rate	**0.046**

Table 5. Evaluation of Factor Analysis Model.

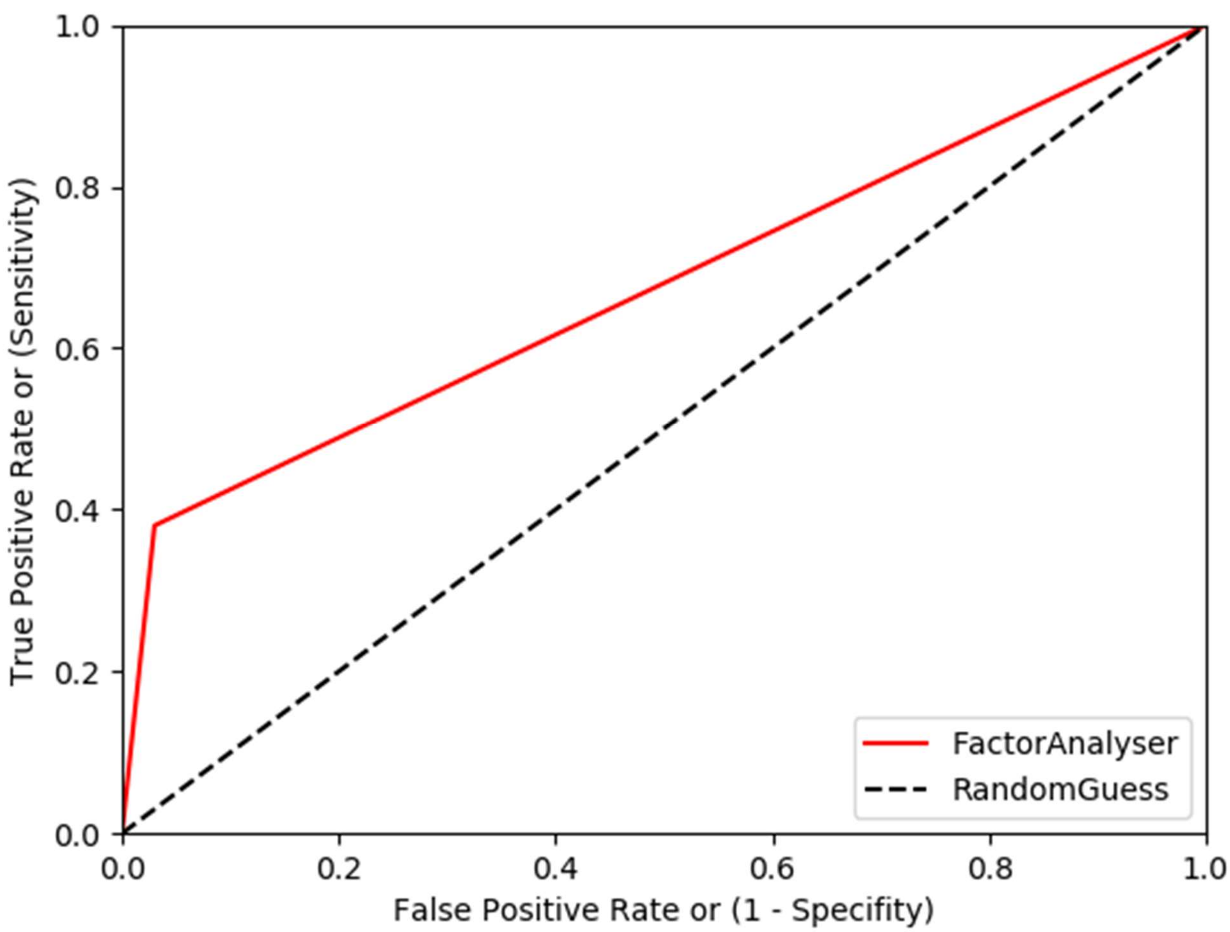


Figure 12. ROC Curve for the Factor Analysis Model

Appendix:

To reproduce the results, run experiment.py (inside Arsingh3_code/Experiment) in the command prompt with the following arguments.

1. '—input' to make sure the data is loaded properly
2. '--traingaussian' to reproduce the results of gaussian model
3. '—trainMOG' to reproduce the results of Mixture of Gaussian
4. '--trainT' to reproduce the result of T
5. '--trainFA' to reproduce the result of Factor Analysis

In addition to basic python libraries the code requires numpy, scipy, Pandas and OpenCv for execution.