ECE763 Project1

Yuhan Chen

February 28, 2019

The purpose for this project is to do classification in face and non-face. We use FDDB face data set to train and test the performance of our models. We built four models in this project: Single Gaussian model, Mixture of Gaussian model, t-distribution model and factor analyzer. Each model has their own advantages and disadvantages.

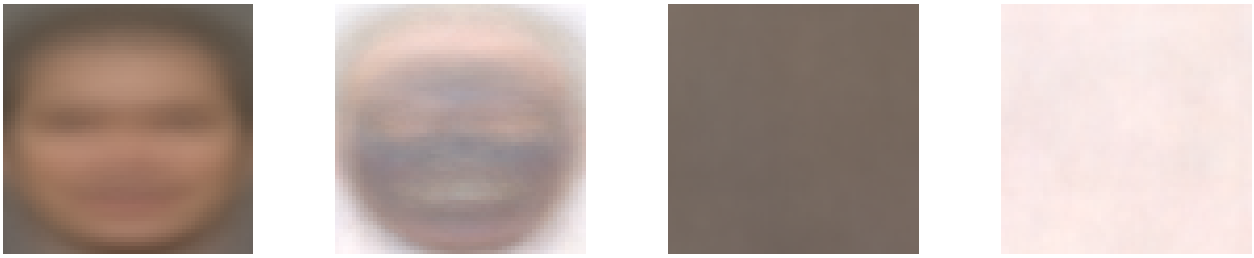
1 Single Gaussian Model

For single Gaussian model, it treats every class as single gaussian model. To use MLE, the formula of μ and σ in function to calculate the MLE, the fomula is shown below.

$$\hat{\mu} = \sum_{i}^{N} x_i / N$$

$$\hat{\sigma}^2 = \sum_{i}^{N} (x_i - \hat{\mu})^2 / N$$

The result is shown in Fig. 1 and ROC curve is shwon in Fig. 2. From the ROC curve, we can see many score is concentrated in a range, so the curve is straight on the both side and zigzag in the middle. The NP(negatives being classified as faces), FN(positives being classified as non-face) and Misclassification is shown in Table. 1. From the table, we can see even the total result is good enough. There are still many faces cannot be recognized.



(a) (b) (c) (d)

Figure 1: Single Gaussian model (a)μ of face, (b)σ of face, (c)μ of non-face, (d)σ of non-face.

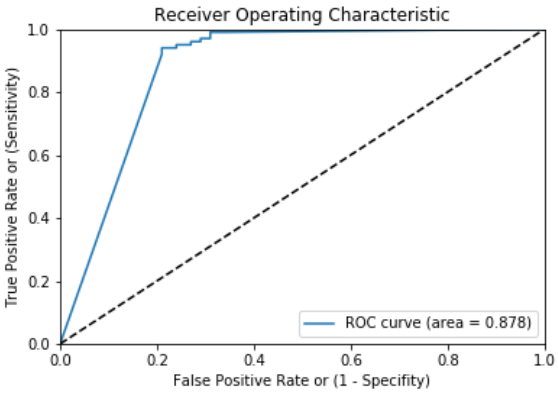


Figure 2: ROC curve of sigle Gaussian model

Table 1: Single Gaussian Test Results

FP	FN	Misclassification
0.24	0.14	0.19

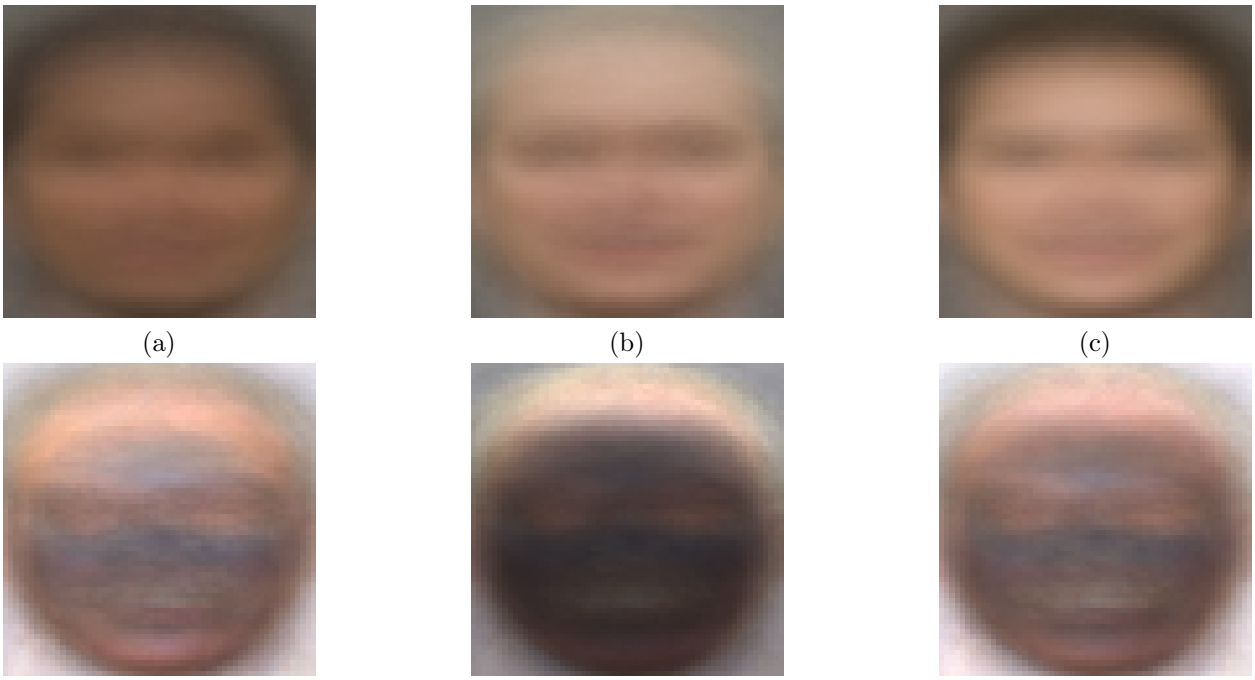
2 Mixture of Gaussian

In face and non-face, there are many different types. For example, in face set, we have different race. In non-face set, we have white or black backgroud. Therefore, we use Gaussian to test face and non-face. Here we use one-fold cross-validation to choose the number of Gaussian we should use. I choose model k=[2,3,4] to do cross-validation(data set:validation set = 9:1) and the result is shown in Table. 2

Table 2: Cross Validation Test in MoG

Fk	P	FN	Misclassification
2	0.42	0.03	0.225
3	0.54	0.01	0.275
4	0.59	0	0.295

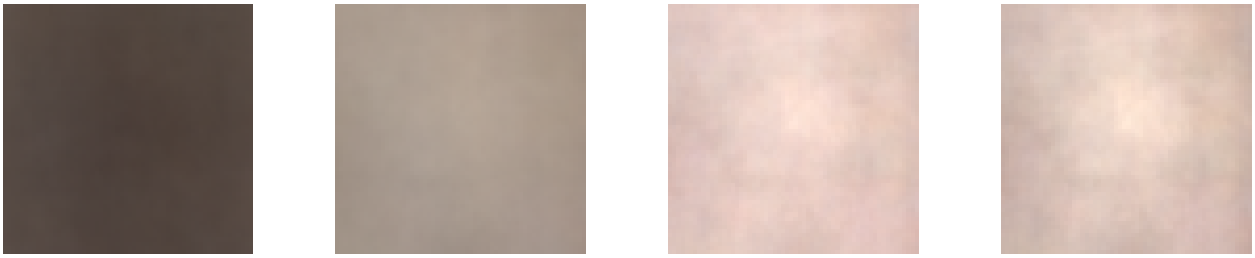
From the table, we can find that when use more classes in face set, more faces can be recognized. However, for background model, we can only classify it as two model in Gaussian model, so more non-faces picture be misclassfied. After using cross-validation, I decide to use k=3 to do the final test. The face results are shown in Fig. 3 and the non-face results are shown in Fig. 4. The ROC curve is show in Fig. 5. Since the FN is very low, the ROC curve is very sharp.



(a) (b) (c)

(d) (e) (f)

Figure 3: Mixture of Gaussian (a)-(c)μ of face model1-3, (d)-(f)σ of face model1-3.



(a) (b) (c) (d)

Figure 4: Mixture of Gaussian (a)-(b)μ of non-face model1-2, (c)-(d)σ of non-face model1-2.

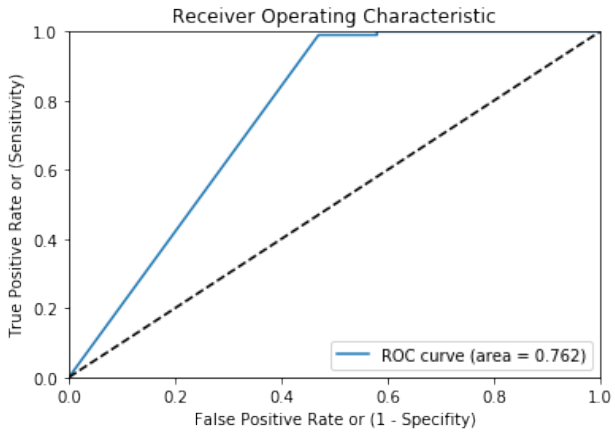
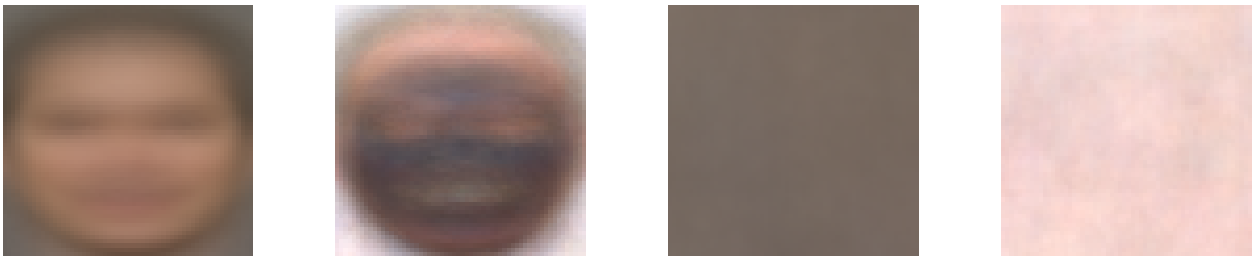


Figure 5: ROC curve of Mixture of Gaussian

3 T-distribution Model

For Gaussian, it is too sensetive for outlier. Therefore, we introduce t-distribution to slove this problem. T-distribution can select outlier. For the only T-distribution model, the results are shown in Fig. 6 and Table. 3. The ROC cure is shown in Fig. 10.The single t-distribution model has a little higher performance than single Gaussian model.



(a) (b) (c) (d)

Figure 6: Mixture of Gaussian (a)-(b)μ and σ for face set, (c)-(d)μ and σ for non-face set.

Table 3: Single T-distribution model

FP	FN	Misclassification
0.23	0.13	0.18

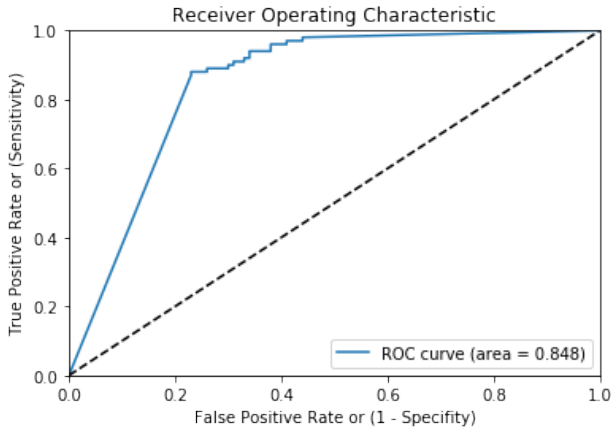


Figure 7: ROC curve of sigle T-distribution model

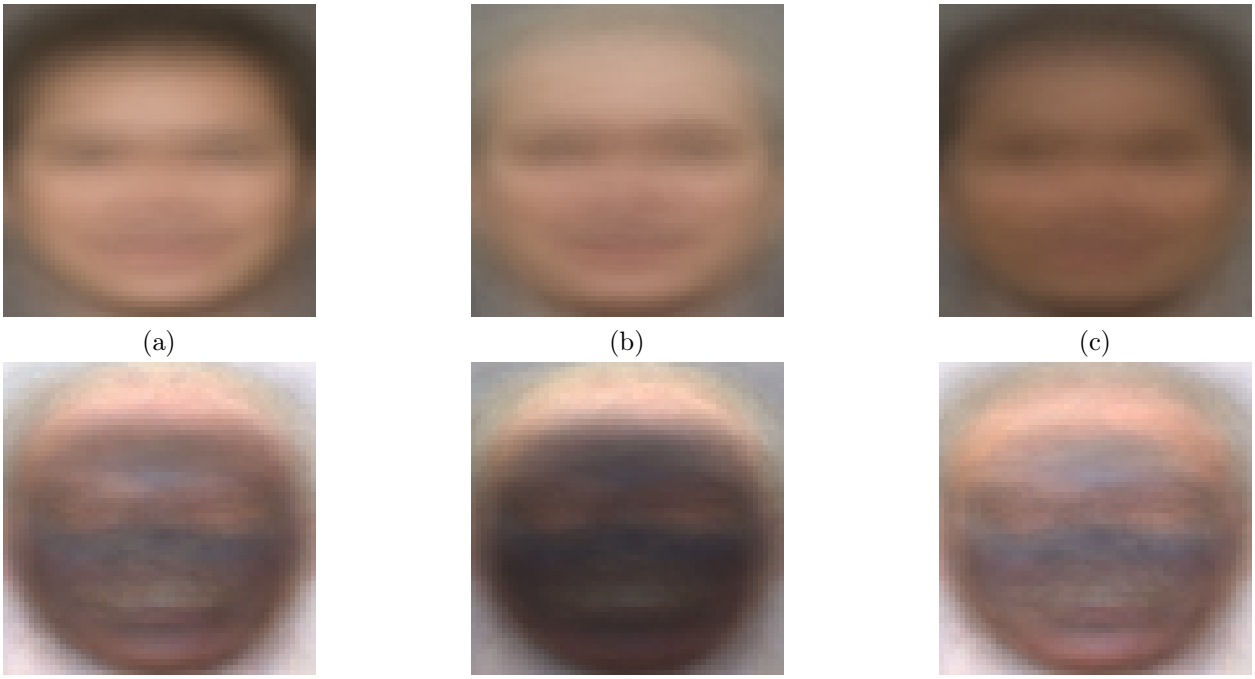
4 Mixture of T-distribution Model

In this case, I still use one-fold cross-validation(data set:validation set = 9:1). The test result is shown in Table. 4. In the table, we can find Mixture of T-distribution improves the performance a lot. Because it is not that sensetive to outlier, the covariacne can show the feature more obviously.

Table 4: Cross Validation Test in Mixture of T-distribution

Fk	P	FN	Misclassification
2	0.292	0.21	0.225
3	0.15	0.08	0.145
4	0.18	0.1	0.148

In the test result, I use k=3 to do final test. The result is shown in Fig. 8 and Fig. 9. The final result is shown in Table. 5 and Fig. **??** is the ROC curve of Mixture of T-distribution.



(a) (b) (c)

(d) (e) (f)

Figure 8: Mixture of T-distribution (a)-(c)μ of face model1-3, (d)-(f)σ of face model1-3.

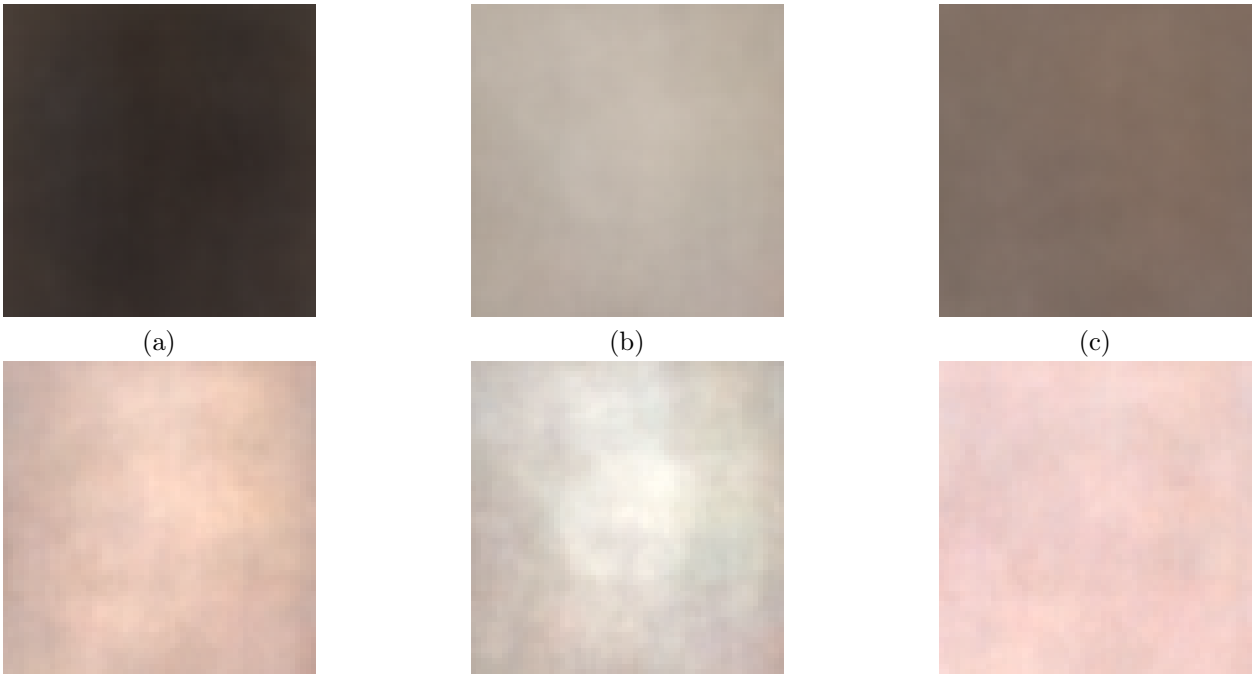


Figure 9: Mixture of T-distribution (a)-(c)μ of face model1-3, (d)-(f)σ of face model1-3.

Table 5: The Test of Final Mixture of T-distribution model

FP	FN	Misclassification
0.18	0.06	0.12

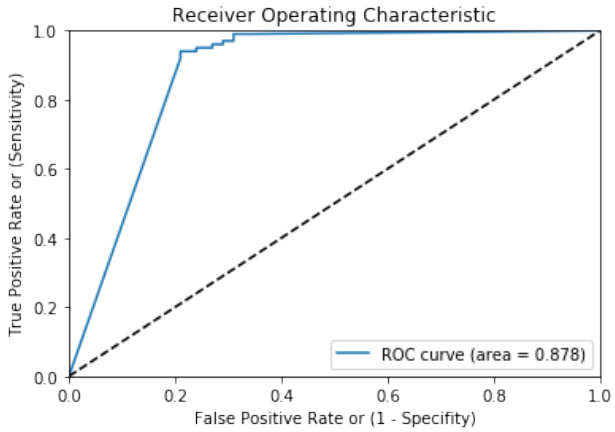
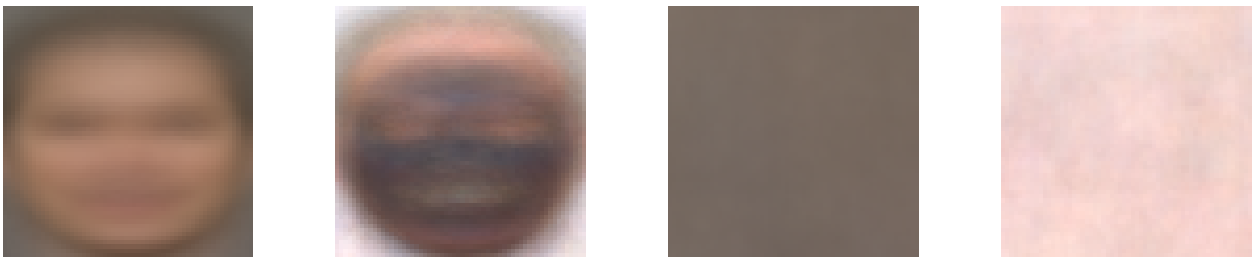


Figure 10: ROC curve of k=3 Mixture of T-distribution model

5 Factor Analyzer

The different between Factor Analyzer and single Gaussian is that the sigma. For Factor Analyzer, we add some factor into covariance matrix to capture the covariance in high-dimensional data. Since Factor Analyzer is very time comsuming, I just use Factor(k=2) here, so the result is not that well. The result figure is shown in Fig. 11 and Table. 6 and Fig. **??** is ROC curve.



(a) (b) (c) (d)

Figure 11: Mixture of Gaussian (a)-(b)μ and σ for face set, (c)-(d)μ and σ for non-face set.

Table 6: The Test of Final Mixture of T-distribution model

FP	FN	Misclassification
0.31	0.22	0.265