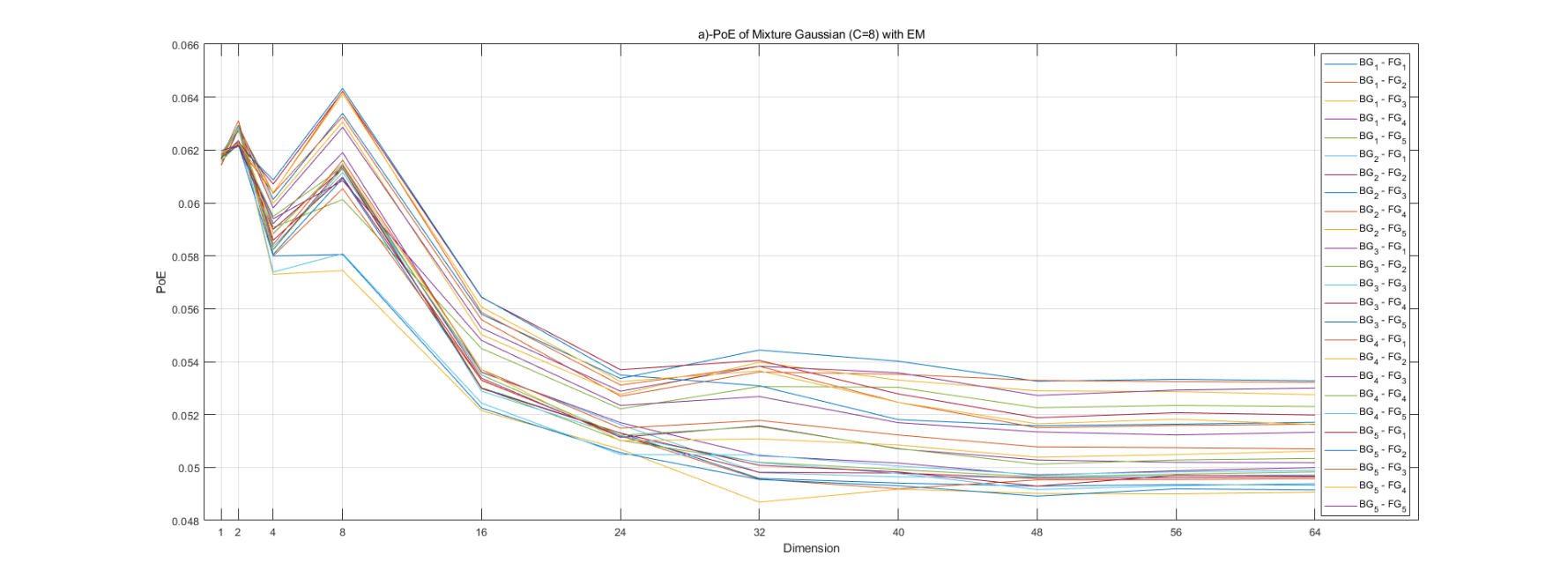
[ECE 271A] Homework5 Report

Yang Yue | A53301503

Learning 5 Gaussian mixtures with C = 8 components

1. Plot of PoE vs dimension
2. Comments to the result

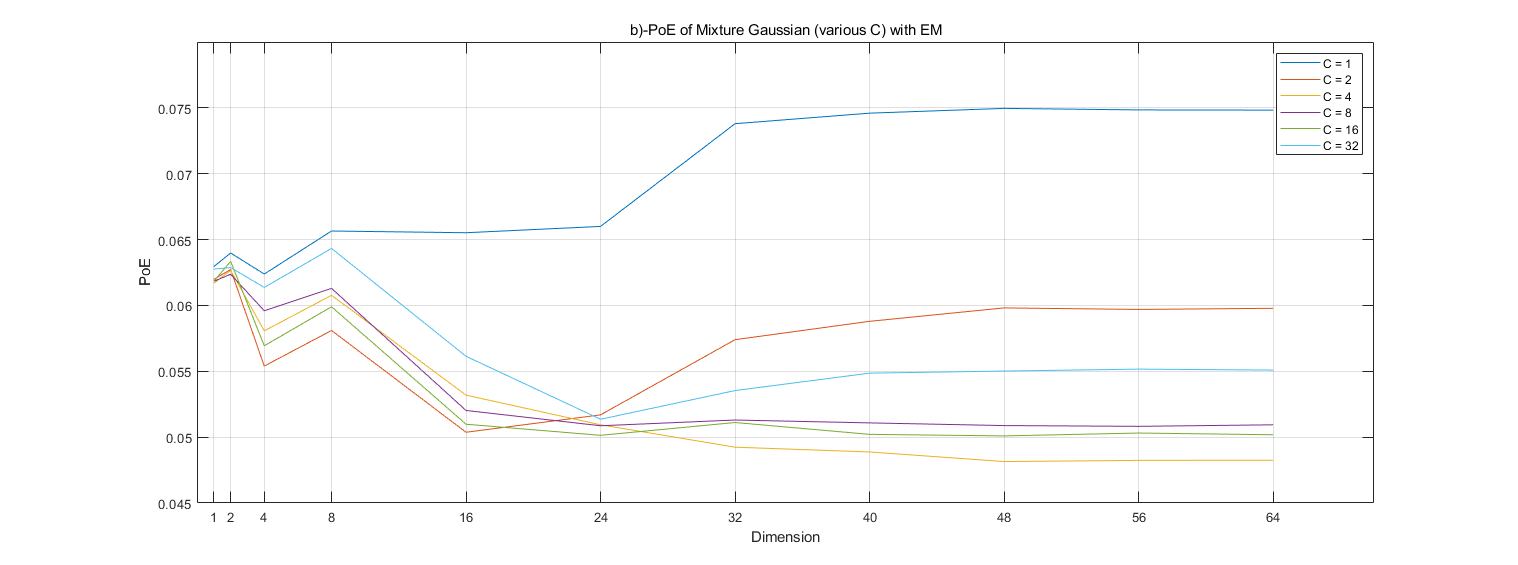
The plot of the error rate of multi-gaussian classifier with various random initializations vs. the dimension of test data is shown above. The plot implies a generally decreasing trend of error rate with the increasing dimensions.

As 1 dimension is used, the PoEs with different initials are close, which implies that the random-initialized parameters of 1-D multi-gaussian of training set is converged to the same values.

With more dimensions of training data involved in parameter estimation, it is intuitively correct that the PoE curve will decline as the dimension increases. While the result shows that there exist some disturbances of the decreasing trend of the curve. This uniform disturbance occurs when dimension equals 2 and 8. And since the parameters are trained by all 64 dimensions of training data, it may be caused by the information from the 2nd and the 5th to 8th dimensions are not “helpful” for us to distinguish where the test data belong to. Then the disturbance effect is weakened because the status of each dimension is no more outstanding, which causes the decreasing more stable.

Also, it can be seen that the distance between each PoE curve is getting large as the dimension increases. This may be due to the random initialized parameters falls to different local optimized areas. The mix-gaussian model is getting complicated will the data dimension increase, so that the local optimized parameter values for every initialized parameter vary even though there exist only tiny difference between each parameter.

Learning Gaussian mixtures with various Cs

1. Plot of PoE vs dimensions
2. Explanation of the effect of number of mixture components

The plot of PoE of mixture gaussian vs different numbers of mixture components is shown above. It is noticed that the trend of curve of number of mixture component is obviously different from others. Since when number of C is too small, it may lead to bias of estimation. Thus, it provides a poor performance of estimating the classification.

We can also observe that the PoE raises again when , this may due to the overfit of training data.

For other cases, the classifier should perform better, and the probability distribution can be fitted more accurately. However, the parameters are randomly initialized at first, the classification result may be disturbed by that fact. It may be concluded that the best number of components may be in the range of .