Brian Wilcox

EM for Gaussian Mixtures Report

a) Probability of Error vs Dimension for 25 classifiers (mixture pairs of 5 randomly initialized mixtures of background and foreground each)**.**

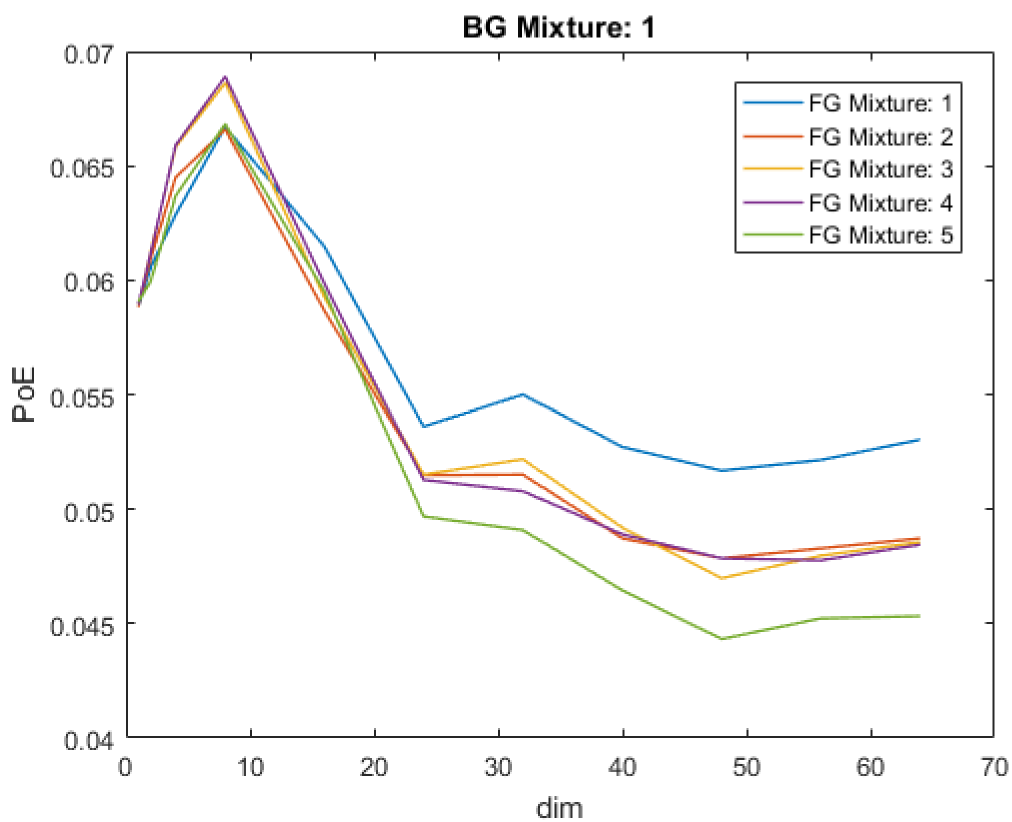
Note: the randomization of mixture parameters were as follows:

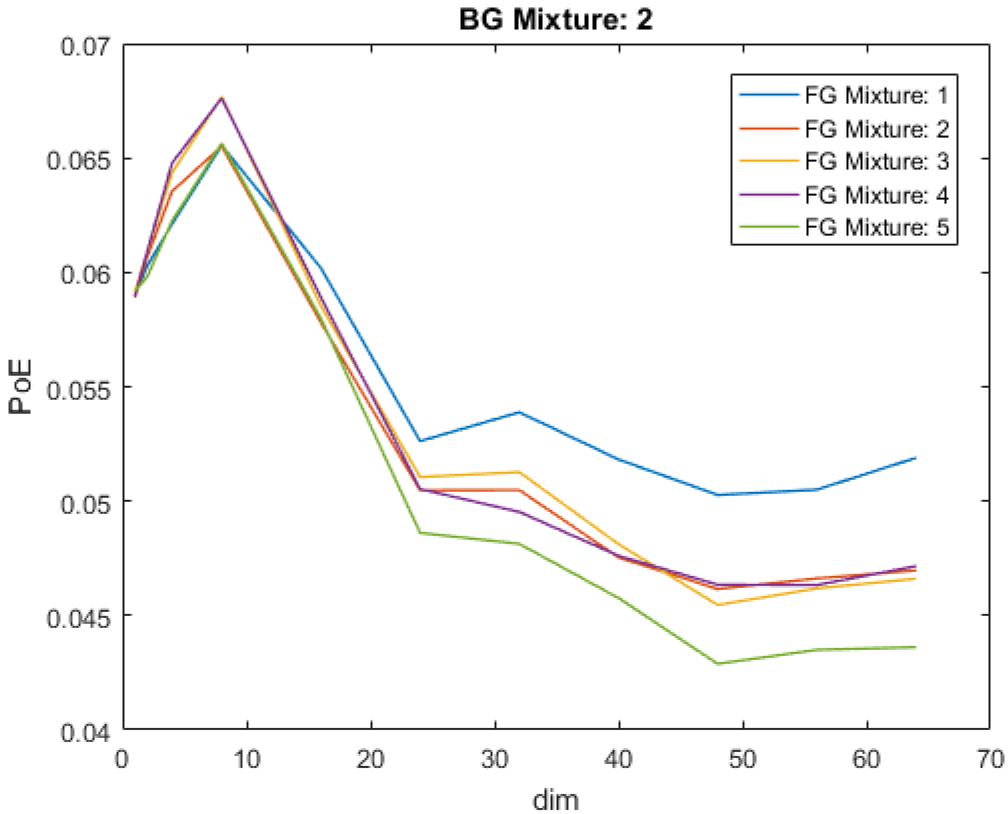
Alpha weights α and initial means μ were sampled from a uniform distribution in the interval [0 1]. The alpha weights were of course normalized.

The diagonal entries of the initial covariance Σ parameters were pulled from a normal distribution with mean of 10 and standard deviation of 2.

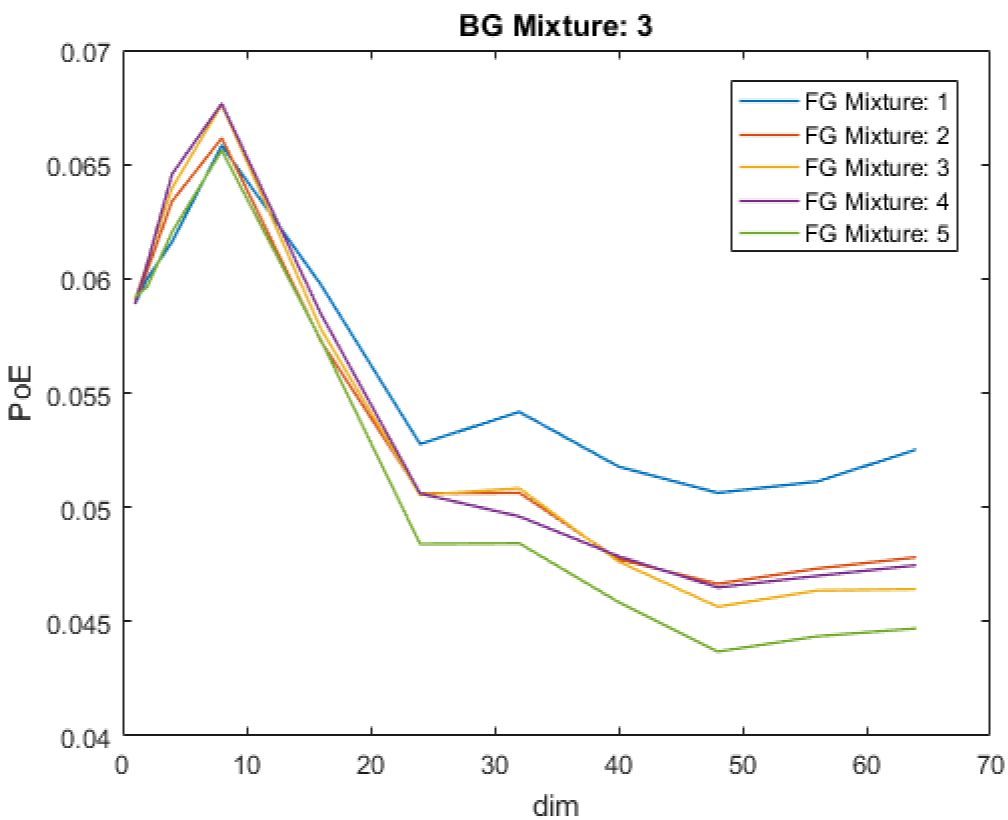
Note 2: The dimensions taken are the first n dimensions in the feature vector

Comments on problem a) will follow after all the graphs

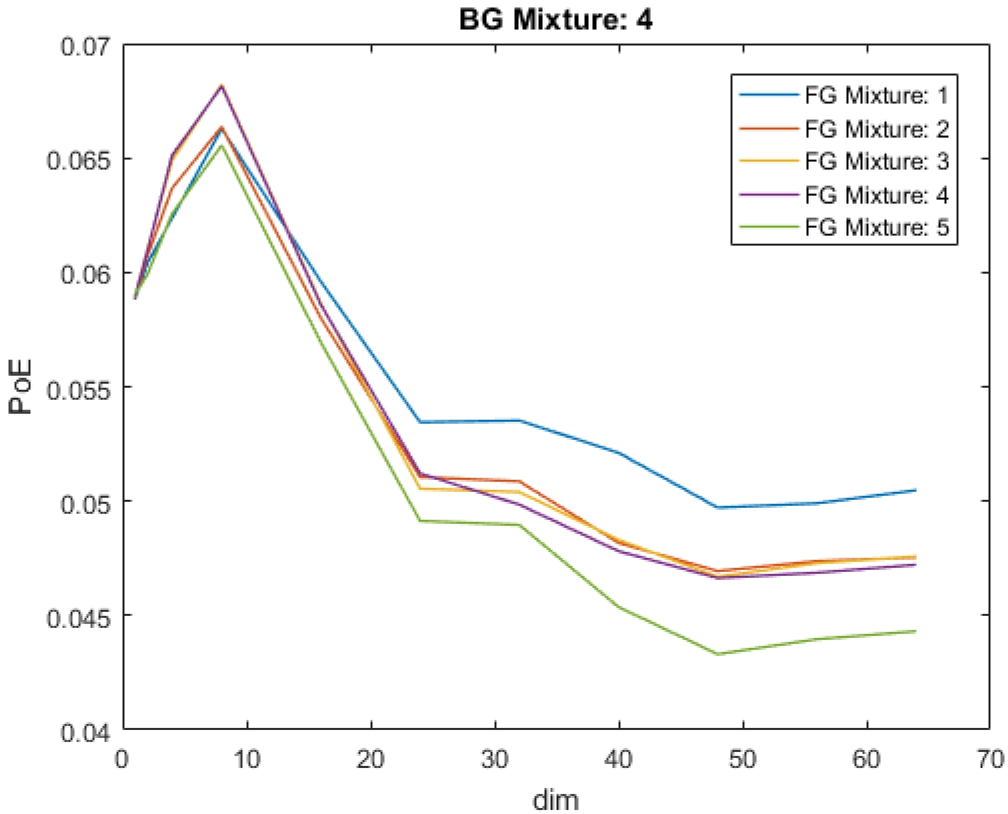


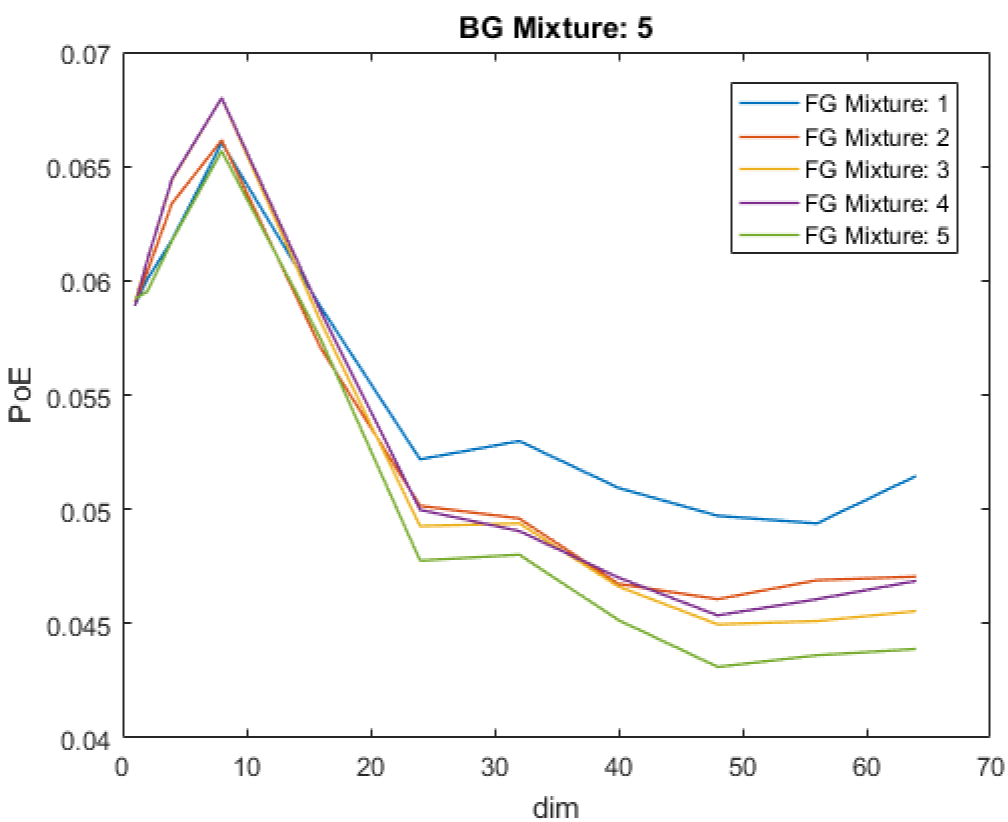


BG Mixture: 2
PoE
dim
0.07
0.065
0.06
0.055
0.05
0.045
0.04
0
10
20
30
40
50
60
70
FG Mixture: 1
FG Mixture: 2
FG Mixture: 3
FG Mixture: 4
FG Mixture: 5



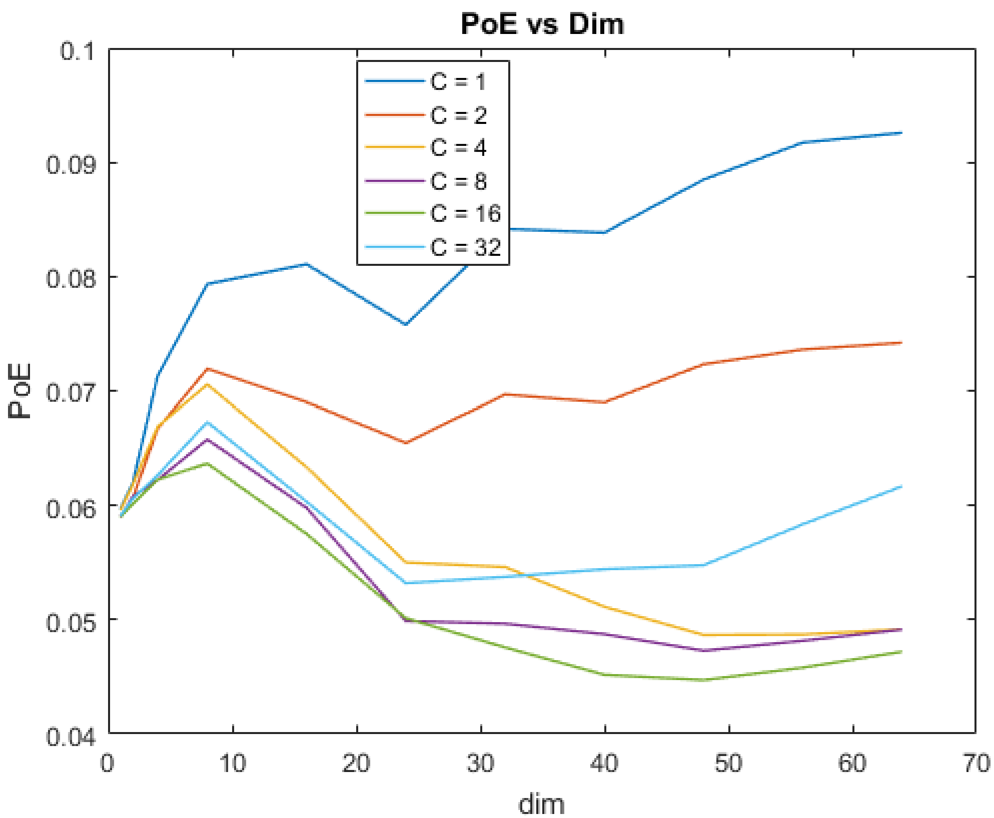
BG Mixture: 3
PoE
dim
0.07
0.065
0.06
0.055
0.05
0.045
0.04
0
10
20
30
40
50
60
70
FG Mixture: 1
FG Mixture: 2
FG Mixture: 3
FG Mixture: 4
FG Mixture: 5


BG Mixture: 4
FG Mixture: 1
FG Mixture: 2
FG Mixture: 3
FG Mixture: 4
FG Mixture: 5
PoE
dim
0.07
0.065
0.06
0.055
0.05
0.045
0.04
0
10
20
30
40
50
60
70


BG Mixture: 5
FG Mixture: 1
FG Mixture: 2
FG Mixture: 3
FG Mixture: 4
FG Mixture: 5
PoE
dim
0.07
0.065
0.06
0.055
0.05
0.045
0.04
0
10
20
30
40
50
60
70

From the 5 graphs above, the PoE generally decreased with more dimensions in the mixture model. The highest error, however, occurred with around 8 dimensions and the lowest PoE with 48 dimensions. This implies the first 48 dimensions does the best job in the mixture model to represent the data, and thus the lowest PoE. There is very little difference across the graphs with only slight variations line to line. Within each graph, there is very prominent variability with the best mixture performance (lowest PoE) for all graphs coming from the FG Mixture 5 (green line). The difference in performance across all the mixtures per each graph shows how the initialization is important for the EM mixture model, particularly since likelihoods can converge to different local maxima.

b) Probability of Error vs dimension for mixture components C = {1,2,4,8,16,32}



In general, the more mixture components led to a lower PoE, however after 16 components the 32 component mixture model performed worse at higher dimensions. The number of mixture components affected the relationship between PoE vs dim seen in part a. With low number of components the PoE generally increased with dimension whereas the best components (8,16) decreased until the best performance with 48 dimensions (as the case in part a). These results indicate that the best number of components to represent the n dimensions was 16. This may be due to the 16 mixtures better capturing the modes/distribution of the data since it is obviously not a simple gaussian distribution. Likewise, representing the data with not enough components would do a worse job particularly as the dimensionality increases and the distribution is more complex. In the case of 32 mixtures, perhaps too many components leads to overfitting.