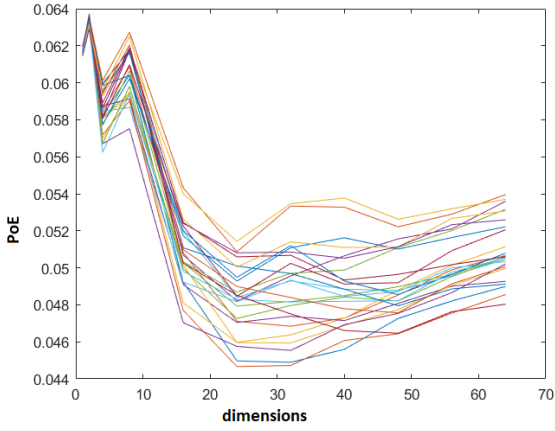


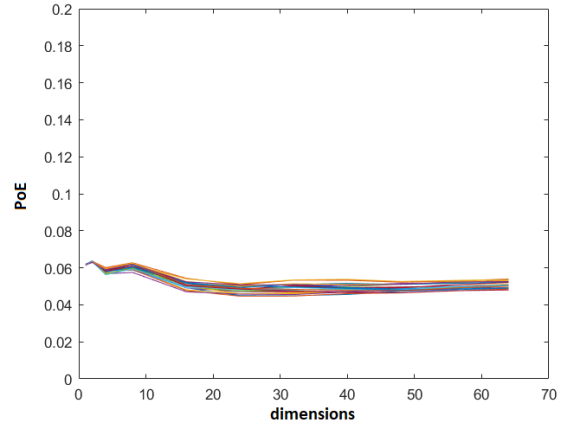
Saurabh Mirani

December 2019

1. Dependence of the probability of error on the initialization

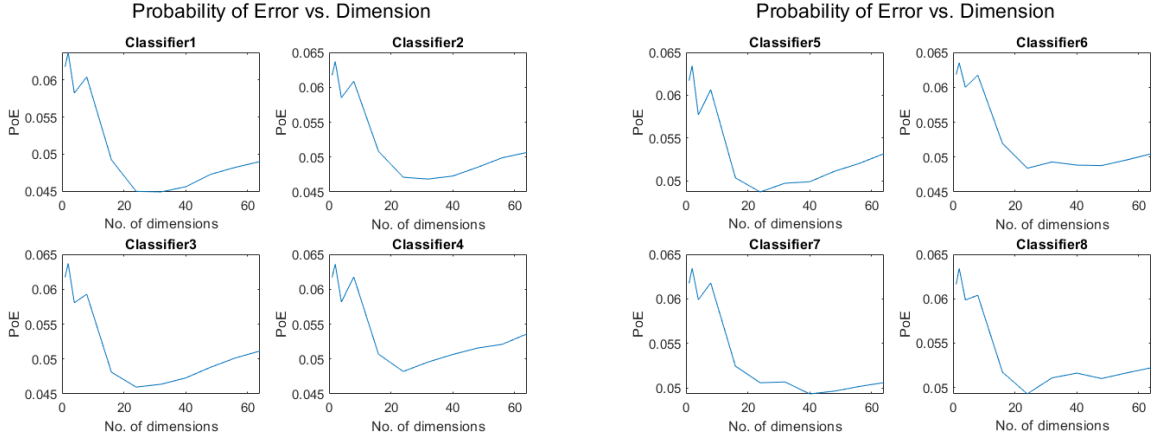


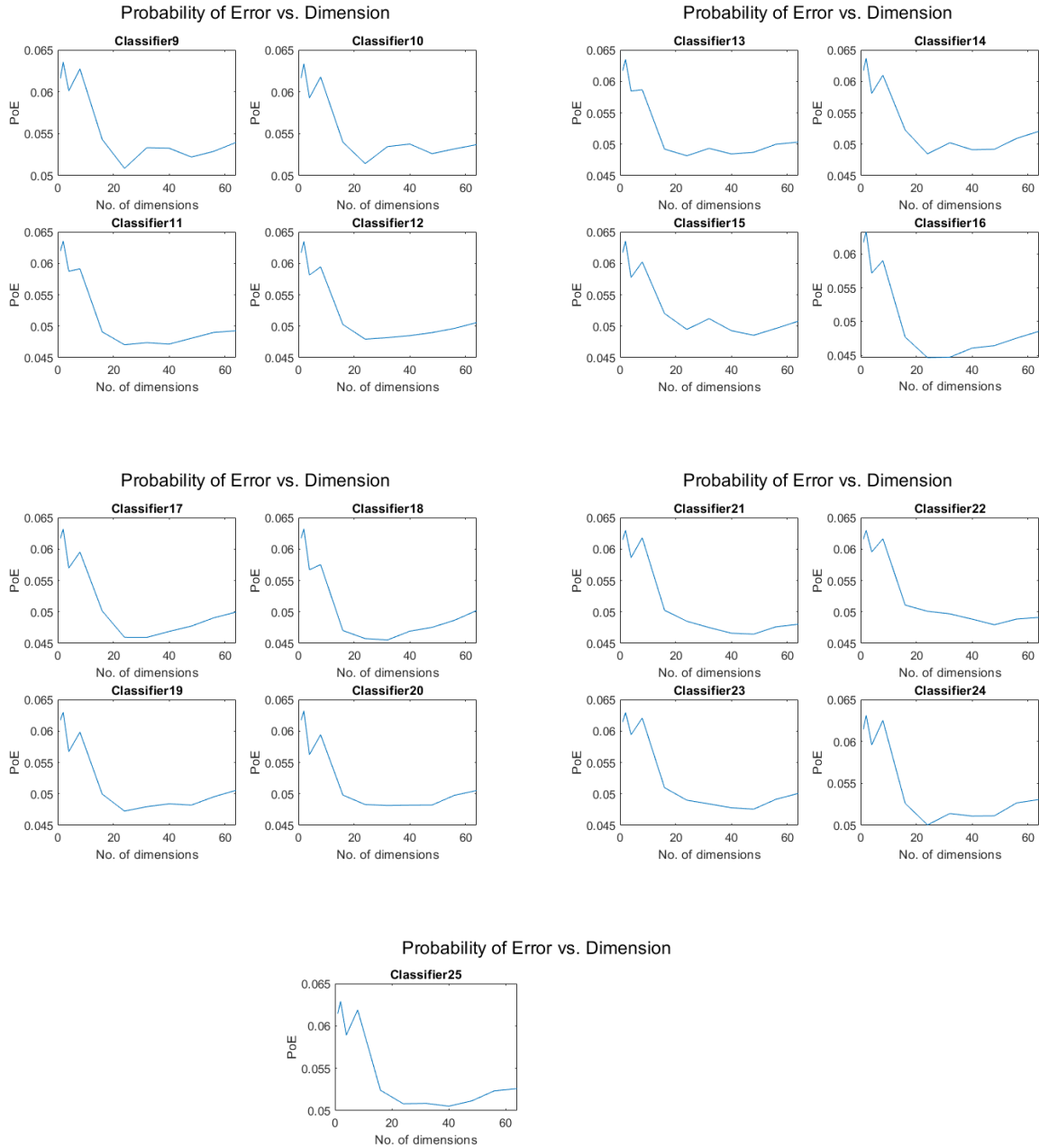
(a) Y - Range : [0.044,0.064]



(b) Y - Range : [0,0.2]

Figure 1: All 25 Classifiers





We have already seen in Homework-2 that 8 dimensions performed better than 64 dimensions. That is because there are some features which are quite bad and increase the error. Here also the same logic applies. We can see that 64 dimensions is never the best. Now we can see a good amount of error for less than 20 dimensions for almost all graphs. This can be explained by, say if we consider only 1 dimension we are losing a lot of information. As we add one more dimension say it is 2D not, we see that error

increases. This is because 2nd feature is not a good feature, as seen in Homework-2. Hence the error increases. However if we keep on adding feature the error then starts decreasing. upto 20 features. Then it again starts increasing due to many bad features being included in our training set.

In general, it can be said that we will obtain minimum PoE, somewhere in between minimum dimensions i.e. 1D and maximum dimensions i.e. 64D here.

Figure 1 (a) & (b) are same just the Y-axis limits are different, (a) to analyze the variation (b) to see the variation with respect to larger scale, i.e. how much significant is the variation. From Figure 1 (a) & (b), we can clearly see that initialization slightly affects the PoE but not so huge that can result in random results. Just a slight variation. It can be considered negligible for all practical purposes. Also initialization affects PoE of higher dimensions much more than that of lower dimensions. Almost all the graph lines are co-incident for lower dimensions. For higher dimensions we can see variation.

2. What is the effect of the number of mixture components on the probability of error?

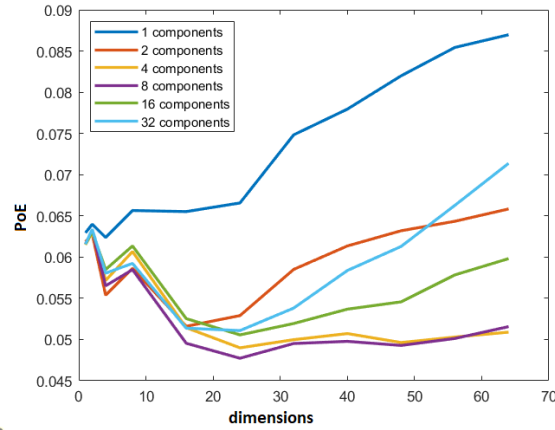


Figure 6: PoE vs No. of dimensions for different no. of mixture components

It can be clearly seen from Figure 6 that PoE decreases initially with increase in number of mixture components. It decreases up to 8 components and then it starts increasing. This can be understood by the fact that 32 components are trying to over-fit the data. Hence we get more error. The noise might have also been fitted in our model which must have been avoided. Over-fitting is almost always bad. 4 and 8 mixture components seem to be fit the data best, not too much and not too less like 1 mixture component. 1 mixture component and 2 mixture components are not able to fit the data properly.

Hence, in general we can say that PoE initially decreases and then increases with increase in number of components.