Nisarg Dabhi, Joshua D'Arcy, Niral Shah
Group 6

Problem 1:

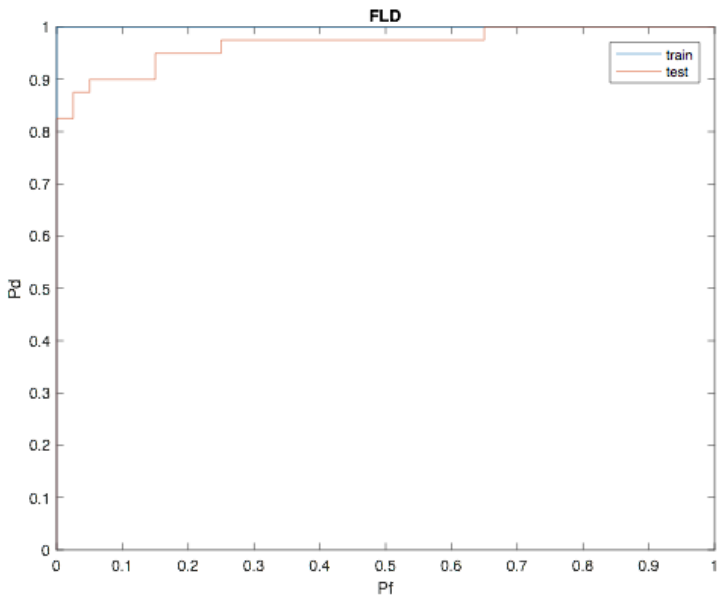

Figure 1 (Indicating Overfitting)

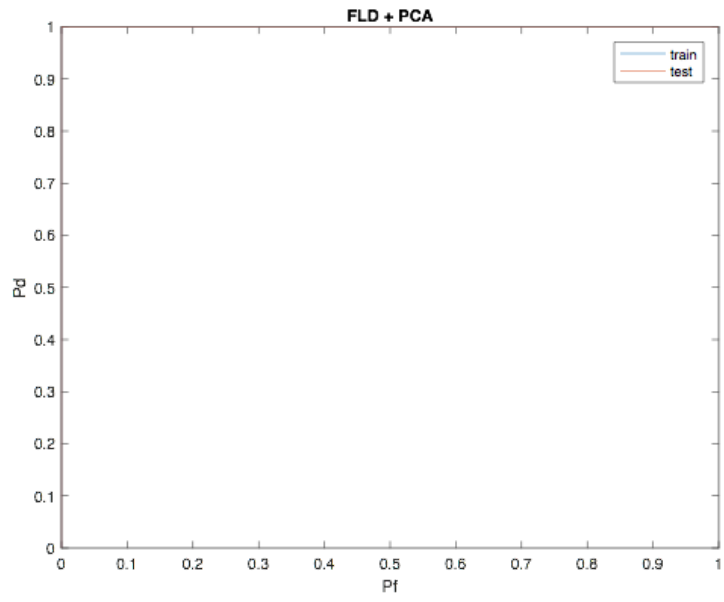

Figure 2 (PCA reduces overfitting)

Ranking:

SBS (2) **=** SFS (2) **<** FLD **<** FLD + PCA

FLD overfits due to the high dimension/low observation ratio. PCA improves on this by reducing overfitting (dimensionality reduction). SBS = SFS, since each dimension has the same variance/difference in means, SBS and SFS with FLD will both choose 2 random features.

Problem 2:

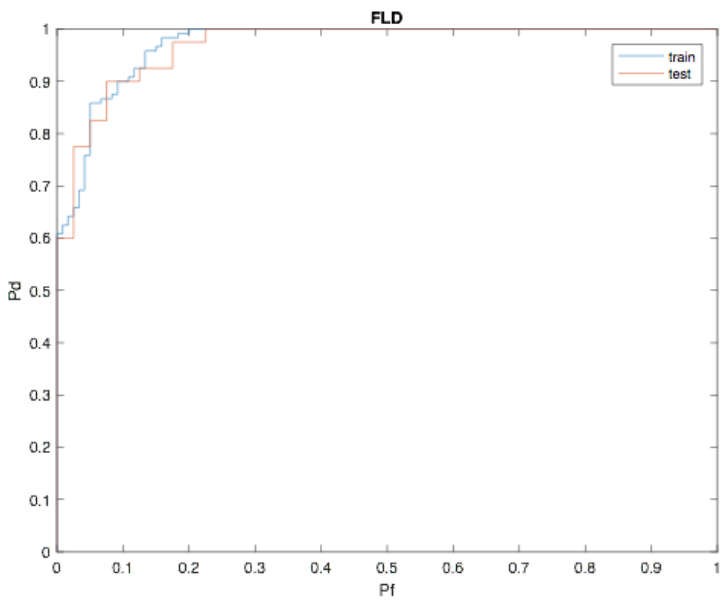

Figure 3 (FLD performs very well)

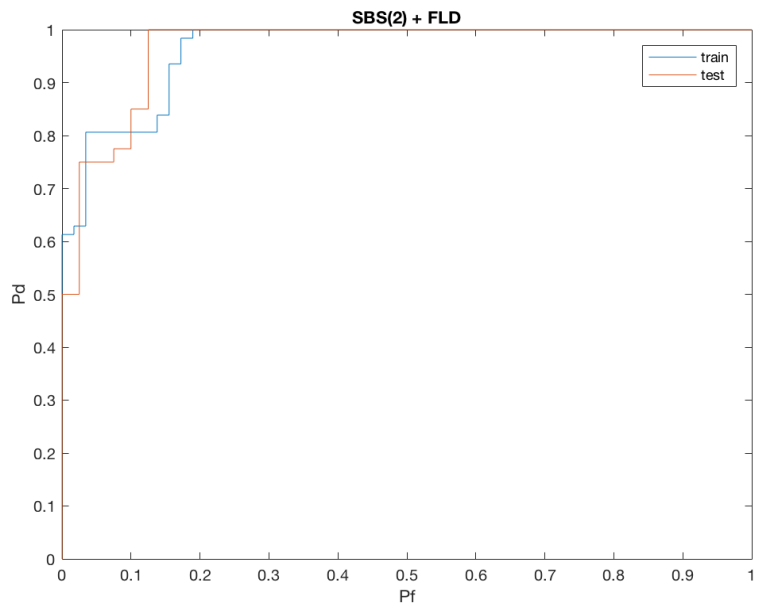

Figure 4 (SBS is the next best)

Ranking:
PCA (1) < SFS (2) < SBS (2) < FLD

In this dataset the first and second dimensions (X1 & X2) are correlated with a correlation of -0.9. This suggests that in terms of feature reduction that features X1 and X2 should be taken together, because otherwise information would be lost. On the other hand this data is already separable as the means in every dimension are different. So as the number of data points increases FLD will perform very well on the dataset. Following that, SBS works better on the test data than SFS because starting with all 3 features, it eliminates 1 (X3) to get the best performance. In doing it this way, it gets the relationship between X1 and X2 (the two variables are correlated). SFS, in contrast, starts with no features, and adds X3 as it performs best as an individual (missing out on the relationship between X1 and X2 for higher combined performance). It then adds a random selection from X1 or X2, since they both appear the same. PCA completely loses the relationship between X1 and X2 when reduced to one dimension.

Problem 3:

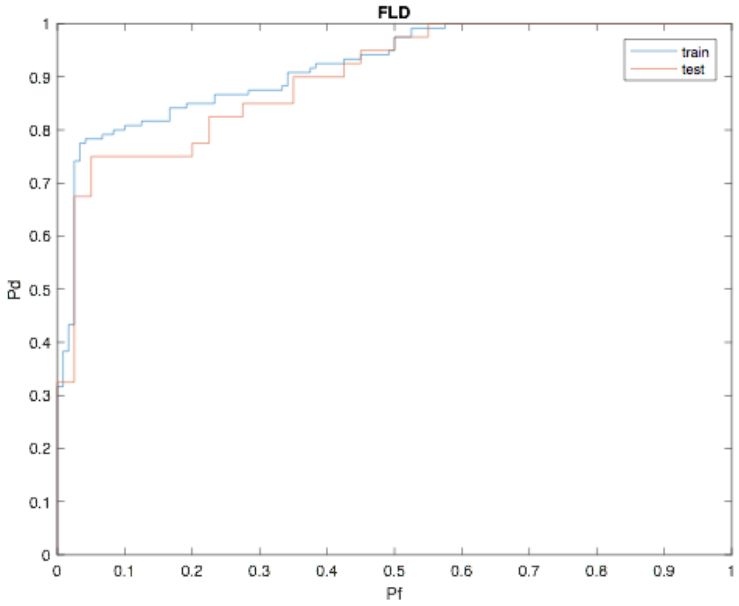

Figure 5 (FLD performs well)

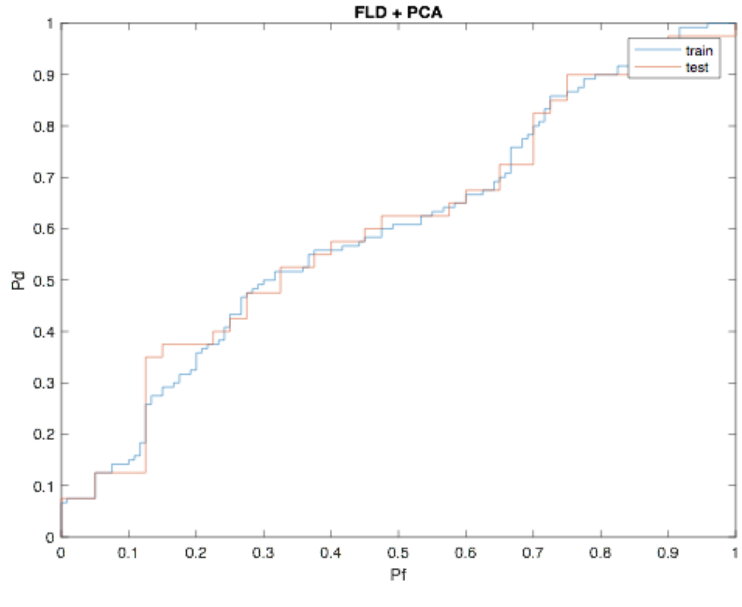

Figure 6 (PCA throws away key info)

Ranking:
PCA (1) < SBS (2) = SFS (2) < FLD

SFS and SBS will perform the same, since they will both choose some random combination of 2 of X1,X2, or X3. X4 has no difference between means. PCA will perform very poorly since the X4 will have high variability, but the data sets for X4 have no difference in mean/separability. FLD will perform well for features X1, X2, and X3, and will ultimately perform better than PCA (which will solely be based on X4).