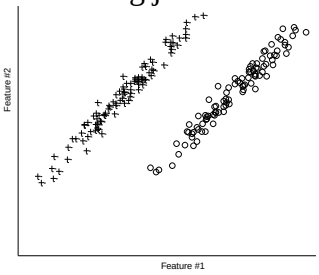
Advantage of Feature Correlations in Pattern Recognition

Bruce R. Linnell, PhD

2005

When it comes to removing features from a dataset, many people have assumed that the best feature sets consist of features which are highly independent of each other. Both the correlation between a feature and the class label, and the correlation between features have been considered as metrics to remove redundant features.

However, features which are correlated to each other are not necessarily redundant – as can be seen in the following figure, both features are highly correlated to each other within each class, yet both features are needed for the best possible separation between classes – using just one feature, the classes are highly overlapped.



The next figure shows the correlations between the features for four different e-nose datasets, after extensive feature selection has found those features which result in the highest estimated classification success rate.

Because only two features are required for the Bacteria dataset, the actual data is shown. The per-class correlations between the two features for the Bacteria datasets are 0.97 and 0.94 for the first and third class. As can be seen in the graph, the second class is bimodal, and the correlations for each group are 1.00 and 0.96.

The other graphs are histograms showing the distributions of the correlations between features. Not only do these results support the arguments that the best features need not be uncorrelated, but show that **many of the best features are actually highly correlated**.