

Tutorial 8

The library “mlbench” contains a set of benchmark datasets for learning problems. Install this package and load the lonosphere dataset from this package by typing “data(lonosphere)”.

- (1) Partition the dataset into training and test sets (60% and 40%).
- (2) Using the absolute fold change to filter the data for classification. Plot the number of features included against the classifier performance of kNN and logistic regression on test set.
- (3) Repeat the above but using t-statistics to rank and filter the features this time.
- (4) Compare the results from using absolute fold change and t-statistics. How many features should be included according to each method? Does your initial partition of training and test sets have anything to do with the selected features?