1. Consider the other parameters that can be used to control the recursive partitioning process. Read the documentation for them in the *rpart.control()* documentation. Also, carry out an Internet search for more information on how to tweak the *rpart()* tuning parameters. Experiment with values for these parameters. Do the trees that result make sense with your understanding of how the parameters are used? Can you improve the prediction using them?
2. In [the section called “Classifying New Messages”](http://www.rdatasciencecases.org/exercises.html#sec:SPAM_ClassifyNew) we used the test set that we had put aside to both select , the threshold for the log odds, and to evaluate the Type I and II errors incurred when we use this threshold. Ideally, we choose from another set of messages that is both independent of our training data and our test data. The method of cross-validation is designed to use the training set for training and validating the model. Implement 5-fold cross-validation to choose and assess the error rate with our training data. To do this, follow the steps:
3. Use the *sample()* function to permute the indices of the training set, and organize these permuted indices into 5 equal-size sets, called folds.
4. For each fold, take the corresponding subset from the training data to use as a 'test' set. Use the remaining messages in the training data as the training set. Apply the functions developed in [the section called “Implementing the Naïve Bayes Classifier”](http://www.rdatasciencecases.org/exercises.html#sec:SPAM_ImplementLLR) to estimate the probabilities that a word occurs in a message given it is spam or ham, and use these probabilities to compute the log likelihood ratio for the messages in the training set.
5. Pool all of the LLR values from the messages in all of the folds, i.e., from all of the training data, and use these values and the *typeIErrorRate()* function to select a threshold that achieves a 1% Type I error.
6. Apply this threshold to our original/real test set and find its Type I and Type II errors.