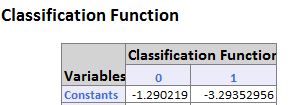
Assignment #10 Discriminant Analysis by Joshua Troup

C. Total spam messages found were 729. 203 of the emails were misclassified as non-spam. The error rate is 27.84% which is moderately high. The misclassification error for training data was 31.5%. Validation error was more accurate with a misclassification error of 27.84% meaning less emails were rejected as being spam.



D. If the actual proportion of spam messages in the email accounts was 10% with the records being considered the “most probable spam” then it result would capture 2.2 times more “spam” messages than selecting 10% of the records at random based on the Decile Lift chart. The existing classification constants are Spam: -3.2935 Nonspam: -1.2902 If we change the proportion from the existing 40% to 10% then the results would be the following:

Spam: -4.2935

Nonspam: -1.3359

E. -1.2902 + log (20) = 0.01083