

DEPARTMENT OF STATISTICS

STATS 784 Data Mining

Assignment 3

Due Monday September 12. Hand in or mail to Alan by 5pm.

Your task in this assignment will be to develop a classification rule for the handwritten digits using the Zip code data set discussed in class. In doing this, you should try all the classification methods we have discussed so far, tuning them as necessary. Calculate estimates of the prediction error of your rule.

Then, using the feature data in the test set provided, predict the corresponding digits using your classification rule. Your answer should include a document in pdf format describing how you chose your predictor, plus a csv file containing your predictions. A sample csv file is on the web page, please copy the format exactly. (i.e. with a single column, with each cell containing one of the digits 0-9, and with no header. There should be 2007 rows in your csv file. Provide enough code so that I can verify your predictions if necessary.

Marks will be awarded for the thoroughness of your search for a prediction rule, and on the number of correct predictions. There will be a prize for the best predictor.

Please include your name in all filenames.

Note: The first variable is the digit, and the remaining variables the 256 greyscale values.