

MATH 567: Mathematical Techniques in Data
Science
Lab 10

Dominique Guillot

Departments of Mathematical Sciences
University of Delaware

April 26, 2017

Exercise 1: Random Forests

- 1 Download the red wine quality dataset at <http://archive.ics.uci.edu/ml/datasets/Wine+Quality>
Read the documentation, and load the dataset into R.
Note: Convert the `quality` variable to a factor.
- 2 Use a random forest to predict the quality variable using the other features. Compute the test error of your model.
- 3 Use the function `varImpPlot` to plot the *importance* of each predictor.

Exercise 2: Deep learning with the zip dataset

- 1 Train a feed-forward neural network to predict whether or not a digit in the zip dataset is a “0” or not.
- 2 Compute the test error of your model.

Note: If training the neural network is too slow, you can work with a subset of the images (e.g. the first 1000 images).

Examine the effect of:

- Using different numbers of neurons on a single layer.
- Using multiple layers.
- Replacing the `tanh` activation function by the `RectifierWithDropout` function.
- Changing the `input_dropout_ratio`.
- Changing the number of epochs.