Statistics and Data Analysis in MATLAB
Kendrick Kay, kendrick.kay@wustl.edu

MATLAB In-Class Exercises 4 (covering Statistics Lecture 7)

Exercise 1. Generate some data using `y = rand(1,80)>.3;`. These data consist of 0s and 1s, representing class assignments. Now suppose we have a model that gives the following predictions for the data: `ypredict = rand(size(y))>.5;`. (Yes, the model is just a random coin-flipper.) What is the percent correct achieved by the model? Of the data points assigned to class 0, what is the percent correct achieved by the model?

Exercise 2. Suppose we measure the height and weight of two groups, males and females. Generate some hypothetical data as follows:

```
heightM = 50 + 10*randn(1,100);
heightF = 45 + 9*randn(1,60);
weightM = 50 + 2*heightM + 10*randn(size(heightM));
weightF = 40 + 3*heightF + 10*randn(size(heightF));
```

Visualize these data using a scatter plot. Then perform nearest-prototype classification to see how well we can predict gender based on measurement of height and weight. What percentage of data points in our sample are correctly classified? Draw the decision boundary of the classifier on the plot.