

EE435 Spring 2012 PS02 (Problem Set 02)

Due: Wednesday 2/1/2012

Download the file from the course website called PS01-MidData.xlsx. This is a compilation of all the data I received from you provided from problem set 1, as it was sent to me. Using this data, answer the following questions:

1. Why will some values of PRT run time cause a problem? What can be done about it?
3. Why will some entries for “Overall PRT Score” cause a problem? What can be done about it?
4. Create and printout a well-labeled histogram for males and females for “Run Time”, after fixing any issues in the data. Who runs faster? Where would you set a threshold to decide male or female for a new sample that you don’t know if it was a male or female? Using this threshold, how many errors would you have?
5. Create and printout a well-labeled scatter plot for “# Situps” (feature #1) vs. “# Pushups” (feature # 2). Choose a suitable linear boundary as a decision boundary, and add this line to your plot (using MATLAB code). Using this boundary, write MATLAB code that will automatically determine the number of errors you have. Record your decision rule, based on your boundary, below.

Decision rule: