Department of Statistics

STATS 784: Data Mining

Assignment 1 2016

Question 1

Consult the references in Lecture 1 and identify and describe two applications of data mining technology. Write about a page for each one.

Question 2

For the Boston Housing data, construct a linear predictor using a subset (possibly all) of the features in the data set Boston (which you can find in the MASS package).

Calculate estimates of prediction error for your predictor using the R330 functions **cross.val** and **err.boot**, as well as the functions **crossval** and **bootpred** in the bootstrap package.

How accurate do you think your estimates are?

Note: the R330 function **cross.val** produces only printed output, it does not return a value. Code for a modified version returning a value for the estimate and its standard error is in the code handout for Lecture 2 on the web page.

Email answers to me by Friday July 29.