

NAME (Please Print): _____

HONOR PLEDGE (Please Sign): _____

Statistics 218

Homework 2

You are allowed to discuss problems with other students, but the final answers must be your own work. Please attach this page to your solutions.

YOU MUST ATTACH SEPARATE PAGES, NEATLY WRITTEN, THAT SHOW YOUR WORK. You may include computer printout.

Report all numerical answers to at least two correct decimal places.

DUE DATE: IN CLASS ON WEDNESDAY, SEPTEMBER 23.

1. Fit a multiple linear regression model, a stepwise linear regression model, an additive model, a generalized additive model, a projection pursuit regression model, a neural network model, an ACE model, an AVAS model, a regression tree model, and a MARS model to the Los Angeles Ozone data posted on our website's homework section. For the GAM fit, first transform the response value (groundlevel ozone concentration) using the transformation found from ACE. Comment briefly on the comparative fits.
2. Exercise 3.2 on page 94.
3. Efron famously introduced the bootstrap by setting a confidence interval on the correlation coefficient between the average LSAT scores and the average GPAs for entering classes at a sample of law schools. Download the population data (use law82 in R, or go to our class website) and use the bootstrap to set the confidence interval for 5 random samples of size 15 at level 95%. Compare your results to Efron's confidence interval (e.g., in the SIAM monograph) and comment.
4. Consider a nearest-neighbor regression algorithm in \mathbb{R}^1 that puts equal weight on the response for the second and third nearest x values, but no weight upon the clos-

est value. Describe the “smoothing” matrix, indicate why it is unusual, check the fine print about smoothing matrices, and diagnose the problem. Also, describe the implications of this procedure for leave-one-out cross-validation.