

## STAT 621 HOMEWORK 6

Due: Monday October 28

1. Spline Regression: The data set `ozone.txt` contains daily measurements of temperature and ozone in New York from May to September 1973. The objective here is to model  $Y = \text{Ozone}$  as a function of  $X = \text{Temp}$ .
  - (a) Fit models using the natural cubic spline basis. Use the `df=` option to control the smoothness of the estimate. Plot the estimated function for several different choices of `df`. Choose a model that you think fits best (just eyeball, this is a subjective choice).
  - (b) Now estimate the function using smoothing splines, with smoothing parameter estimated by Generalized Cross Validation. Plot this estimate along with choice of model from part a. Comment.
2. Generalized Additive Models: The data set `HeartDisease.txt` contains measurements from a case-control study of coronary heart disease. Here we will model  $Y = \text{systolic blood pressure sbp}$  as a function of one or more predictors: no. cigarettes per day (`tobacco`), LDL cholesterol (`ldl`), body fat (`adiposity`), family history (`famhist`), stress (`typea`), obesity (`obesity`), alcohol use (`alcohol`), and age (`age`).

Model the data using GAMs, assuming a normal response. Find the best model you can using any combination of smooth terms and linear terms that is appropriate. Give a brief summary of your modeling strategy, and justify your final choice of model including results of model comparisons (e.g., F-tests or AIC) and an evaluation of model assumptions. Discuss your final model and what it tells you about the relationships among the variables. Provide marginal effects plots.
3. Project Proposal: Write up a description (shoot for 1 page) of your plan for the final project in this class. Provide as much detail as you can.