Math676: Statistical Learning II

Homework Assignment #1

Due 03/05/2009

1. P296 Ex 9.2 Let $\mathbf{A}$ be a known $k \times k$ matrix, $\mathbf{b}$ be a known k-vector, and $\mathbf{z}$ be an unknown k-vector. A Gauss-Seidel algorithm for solving the linear system of equations $\mathbf{Az} = \mathbf{b}$ works by successively solving for element z_j in the jth equation, fixing all other z_j's at their current guesses. This process is repeated for $j = 1, 2, ..., k, ...$, until convergence (Golub and Van Loan, 1983).

   (a) Consider an additive model with N observations and p terms, with the jth term to be fit by a linear smoother $\mathbf{S}_j$. Consider the following system of equations:

$$\begin{pmatrix} \mathbf{I} & \mathbf{S}_1 & \mathbf{S}_1 & ... & \mathbf{S}_1 \\ \mathbf{S}_2 & \mathbf{I} & \mathbf{S}_2 & ... & \mathbf{S}_2 \\ & ... & ... & & \\ \mathbf{S}_p & \mathbf{S}_p & \mathbf{S}_p & ... & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ ... \\ \mathbf{f}_p \end{pmatrix} = \begin{pmatrix} \mathbf{S}_1\mathbf{y} \\ \mathbf{S}_2\mathbf{y} \\ ... \\ \mathbf{S}_p\mathbf{y} \end{pmatrix}.$$

   Here each $\mathbf{f}_j$ is an N-vector of evaluations of the jth function at the data points, and $\mathbf{y}$ is an N-vector of the response values. Show that backfitting is a block-wise Gauss-Seidel algorithm for solving this system of equations.

   (b) The Gauss-Seidel procedure converges if the matrix $\mathbf{A}$ is positive definite. For a simple case $p = 2$, and assuming each $\mathbf{S}_j$ is symmetric with eigenvalues in $[0, 1)$, show that backfitting converges.

2. Consider the ozone data of Figure 6.9.

   (a) Fit an additive model to the cube root of ozone concentration as a function of temperature, wind speed, and radiation. Compare your results to those obtained via the trellis display in Figure 6.9. (Hint: gam in R package gam for additive models)

   (b) Fit trees, MARS and PRIM to the same data, and compare the results to those found in (a) and in Figure 6.9. (Hint: rpart in R package rpart for regression tree, prim in R package prim for PRIM, earth in R package earth for MARS)