Statistical Methods - Nonparametric Regression

Lecture 8

Mattias Villani

Sveriges Riksbank and Stockholm University

May 18, 2010

Overview Lecture 8

- Shrinkage and the LASSO
- Natural cubic splines
- Interpretation

Ridge regression / Normal prior

- Ridge regression

$$\min_{\beta} (y - X\beta)'(y - X\beta) + \lambda\beta'\beta$$

  gives the estimator

$$\hat{\beta}_{\lambda} = (X'X + \lambda)^{-1}X'y$$

  which has the effect of shrinking the usual OLS estimate $\hat{\beta}$ toward zero. See for example the case with orthogonal predictors $X'X = I$ where $\hat{\beta} = X'y$ and $\hat{\beta}_{\lambda} = \frac{1}{1+\lambda}\hat{\beta}$.
- It can be shown (see HTF) that $\hat{\beta}_{\lambda}$ shrinks more in the unstable directions of the covariate space (the directions given by the principal components with smallest variances).
- $\hat{\beta}_{\lambda}$ can be motivated as the posterior mode (=mean) under the prior $\beta \sim N(0, \lambda^{-1}I)$.

Lasso / Laplace prior

- The alternative loss function

$$\min_{\beta}(y - X\beta)'(y - X\beta) + \lambda \sum_{j=1}^{p} |\beta_j|$$

gives the Lasso estimator, $\hat{\beta}_t$, which cannot be written up in closed form.

- This problem can be re-written as

$$\min_{\beta}(y - X\beta)'(y - X\beta) \text{ subject to } \sum_{j=1}^{p} |\beta_j| \leq \text{t}.$$

- Continuous variable selection since some elements may be exactly zero.
- Lasso may be motivated as the posterior mode under a Laplace prior

$$p(\beta_j) \propto \exp\left(-\lambda\,|\beta_j|\right).$$

- Variable selection from the Lasso is somewhat artificial. Posterior mode$\neq$posterior mean, and the posterior mean under a Laplace prior does not give rise to variable selection.

The Lasso path and Least Angle Regression

- When $t > \sum_{j=1}^{p} |\hat{\beta}_j|$, $\hat{\beta}_t = \hat{\beta}$.
- It is common to look at the whole Lasso path $\hat{\beta}_L(\lambda)$ for all $\lambda > 0$, or even better $\hat{\beta}_L(s)$ where $s = \frac{t}{\sum_{j=1}^{p} |\hat{\beta}_j|}$ for all $s \in [0, 1]$.
- The entire Lasso path can be computed extremely efficiently using a modification of Least Angle Regression (LAR).
- Least Angle Regression is a more fair version of forward variable selection. LAR doesn't add **all** of a variable to the regression in a given step of the forward search. Instead, an increasing fraction of the variable's coefficient is added to the model until some other variable j has higher correlation with the residuals. We then add an increasing fraction of variable j to the model until some other variable has higher correlation with the residuals, and so on until we obtain the least squares estimate.
- The Lasso path is obtained by a modification of the LAR where a variable is set to zero once it crosses the zero line.

Natural cubic splines

- Polynomials can be problematic at the boundaries. Splines are like local polynomials and can therefore have even bigger problems at the boundaries.
- Solution: force the spline to be linear beyond the boundaries, i.e. impose the condition that second and third order derivatives are zero outside the boundaries.
- Starting from the truncated cubic splines, the linear-beyond-boundary condition gives us the reduced basis:

$$N_1(x) = 1, N_2(x) = x, \text{ and } N_{k+2}(x) = d_k(x) - d_{K-1}(x)$$

where

$$d_k(x) = \frac{(x-\xi_k)_+^3 - (x-\xi_K)_+^3}{\xi_K - \xi_k}$$