6242 Midterm

G43232719 Dandan Feng

3. Consider the elastic net problem

$$\min[||y - X\beta||^2 + \lambda\{\alpha||\beta||_2^2 + (1-\alpha)||\beta||_1\}], \qquad \lambda \geq 0, \qquad 0 \leq \alpha \leq 1.$$

Show how one can turn this into a lasso problem using an augmented version of X and Y.

When $\alpha = 0$, then the whole ridge regression penalty goes away, and we're left with just the lasso regression penalty. The elastic net problem becomes a lasso problem.

$$\arg\min[||y - X\beta||^2 + \lambda\{||\beta||_1\}], \qquad \lambda \geq 0$$

7. What is the degrees of freedom of a kernel regression fit with bandwidth h. Show how to obtain it in detail.

We are given $i.i.d.$ samples $(x_i, y_i), i = 1, \dots, n$ from the model

$$y_i = f(x_i) + \epsilon_i, \qquad i = 1, \dots, n$$

where the errors $\epsilon_i, i = 1, \dots n$ are uncorrelated with common variance σ^2, and we treat the predictor measurements $x_i, i = 1, \dots n$ as fixed. Assume for now that each $x_i \in \mathbb{R}$ (i.e., the predictors are 1-dimensional). Our goal is to estimate f with some function $\hat{f}$.

Kernel regression is a special case of linear smoothers since:

$$\hat{f}_i = \hat{y}(x_i) = \sum_{j=1}^{n} w(x_i, x_j) \cdot y_j$$

i.e., we can write

$$\hat{y} = Sy$$

for the matrix $S \in \mathbb{R}^{n\times n}$ defined as $S_{ij} = w(x_i, x_j)$.

Given a choice of kernel K, and a bandwidth h, kernel regression is defined by taking

$$w(x_i, x_j) = \frac{K(\frac{x_j - x_i}{h})}{\sum_{j=1}^{n} K(\frac{x_j - x_i}{h})}$$

The kernel regression estimator is

$$\hat{y}(x_i) = \frac{\sum_{j=1}^{n} K(\frac{x_j - x_i}{h}) \cdot y_j}{\sum_{j=1}^{n} K(\frac{x_j - x_i}{h})}$$

Degrees of freedom describes the effective number of parameters used by a fitting procedure and provides a quantitate measure of estimator complexity. We define the degrees of freedom of $\hat{f}$ in matrix notation as

$$df(\hat{y}) = \frac{1}{\sigma^2} tr(cov(\hat{y}, y))$$

where $y = (y_1, \dots, y_n) \in \mathbb{R}^n$ and $y = (\hat{y}_1, \dots, \hat{y}_n) \in \mathbb{R}^n$

$$df(\hat{y}) = \frac{1}{\sigma^2} tr\big(cov(Sy, y\,)\big) = \frac{1}{\sigma^2} tr\big(Scov(y, y\,)\big) = tr(S) = \sum_{i=1}^{n} w(x_i\,, x_i).$$

So, the degrees of freedom of a kernel regression fit with bandwidth h is:

$$\sum_{i=1}^{n} \frac{K(\frac{x_i - x_i}{h})}{\sum_{j=1}^{n} K(\frac{x_j - x_i}{h})} = \sum_{i=1}^{n} \frac{K(0)}{\sum_{j=1}^{n} K(\frac{x_j - x_i}{h})}$$