Regression Methods: Problems MATH-408
Anthony Davison

Problem 1 Let $y \sim (\mu, \sigma^2 I_n)$ and let H_λ denote the 'hat matrix' corresponding to a linear smoother (such as a ridge or smoothing spline fit), i.e., the fitted values are $\widehat{\mu} = H_\lambda y$.

(a) Show that $(\widehat{\mu} - \mu)^{\mathrm{T}}(\widehat{\mu} - \mu)$ has expectation $\|(I - H_\lambda)\mu\|_2^2 + \sigma^2 \mathrm{tr}(H_\lambda^{\mathrm{T}} H_\lambda)$.

(b) Show that $(y - \widehat{\mu})^{\mathrm{T}}(y - \widehat{\mu})$ has expectation $\sigma^2(n - 2\nu_1 + \nu_2) + \|(I - H_\lambda)\mu\|_2^2$, where $\nu_1 = \mathrm{tr}(H_\lambda)$ and $\nu_2 = \mathrm{tr}(H_\lambda^{\mathrm{T}} H_\lambda)$. Hence explain the use of

$$\widehat{\sigma}_\lambda^2 = \frac{(y - \widehat{\mu})^{\mathrm{T}}(y - \widehat{\mu})}{n - 2\mathrm{tr}(H_\lambda) + \mathrm{tr}(H_\lambda^{\mathrm{T}} H_\lambda)}$$

as an estimator of σ^2. Under what circumstances is this estimator unbiased? What does this give in the case of a standard (i.e., non-smoothed) linear model?

Problem 2 Consider lasso estimation when the design matrix X is orthogonal, i.e., $X^{\mathrm{T}} X = I_p$.

(a) Show that in this case the least squares estimator equals $\widehat{\beta} = X^{\mathrm{T}} y$ and deduce that the function to be minimised is of the form

$$L = \tfrac{1}{2}\left(y^{\mathrm{T}} y - 2\widehat{\beta}^{\mathrm{T}}\beta + \beta^{\mathrm{T}}\beta\right) + \lambda \sum_{r=1}^{p} |\beta_r|.$$

Explain why this is convex in each element of β.

(b) Show that the lasso estimator $\tilde{\beta}$ can be computed from $\widehat{\beta}$ using the soft thresholding function

$$\tilde{\beta}_r = \mathrm{sign}(\widehat{\beta}_r)(|\widehat{\beta}_r| - \lambda) I(|\widehat{\beta}_r| > \lambda), \quad r = 1, \ldots, p.$$