Regression Methods: Problems MATH-408
Anthony Davison

Problem 1 Let A and B be $q \times q$ matrices, and suppose that $(A + \alpha B)^{-1}$ exists for some $\alpha > 0$. If η is an eigenvalue of $(A + \alpha B)^{-1}A$, show that

(a) if B is invertible, then

$$(A + \alpha B)^{-1}A = B^{-1/2}(B^{-1/2}AB^{-1/2} + \alpha I_q)^{-1}B^{-1/2}A,$$

and deduce that $\eta = \eta'/(\alpha + \eta')$ where η' is an eigenvalue of $B^{-1/2}AB^{-1/2}$, and

(b) if A is invertible, then $\eta = 1/(1 + \alpha\eta'')$, where η'' is an eigenvalue of $A^{-1/2}BA^{-1/2}$.

Problem 2

(a) If X has $n > p$ and rank p, use its singular value decomposition to write the linear model $y \sim (X\beta, \sigma^2 I_n)$ as $y \sim (UD\gamma, \sigma^2 I_n)$, and give expressions for the least squares estimators $\widehat{\beta}$ and $\widehat{\gamma}$.

(b) Show that the squared Euclidean distance of $\widehat{\beta}$ from β, i.e., $Q = \|\widehat{\beta} - \beta\|_2^2$, can be written as $\|\widehat{\gamma} - \gamma\|_2^2$, where $\widehat{\gamma} = \mathrm{diag}(1/d_1, \ldots, 1/d_p, 0, \ldots, 0)U^{\mathrm{T}}y$. Under what circumstances will the variance of $\widehat{\gamma}$ be large?

(c) If y has a normal distribution, show that $Q \stackrel{\mathrm{D}}{=} \sum_{r=1}^{p} \sigma^2 Z_r^2/d_r^2$, where $Z_1, \ldots, Z_p \stackrel{\mathrm{iid}}{\sim} \mathcal{N}(0, 1)$, and hence find the mean and variance of Q.

Problem 3

(a) Show directly that $\widehat{\beta}_\lambda = (X^{\mathrm{T}}X + \lambda I_p)^{-1}X^{\mathrm{T}}y$ minimises the ridge regression sum of squares

$$(y - X\beta)^{\mathrm{T}}(y - X\beta) + \lambda\beta^{\mathrm{T}}\beta,$$

when the response y and the covariate matrix X are centred, so the model does not include a vector of ones.

(b) Use the singular value decomposition $X = UDV^{\mathrm{T}}$, in which $V = (v_1, \ldots, v_p)$ and $U = (u_1, \ldots, u_n)$ in terms of column vectors, to show that

$$\widehat{\beta}_\lambda = \sum_{d_j > 0} \frac{d_j}{d_j^2 + \lambda} u_j^{\mathrm{T}} y \times v_j.$$

Give a similar expression for the fitted value $\widehat{y}_\lambda = X\widehat{\beta}_\lambda$ and discuss the effect of increasing λ for different values of d_j.

(c) Find expressions for edf_λ and for the bias and variance of $\widehat{\beta}_\lambda$ in terms of the SVD.

Problem 4 (Centering and penalisation)

(a) In a linear model whose response vector y has mean $\beta_0 1_n + X\beta$, show that

$$y - \beta_0 1_n - X\beta = y_* - (\gamma - \overline{y})1_n - X_*\beta,$$

with a suitable choice of γ, where $1_n^{\mathrm{T}} y_* = 0$ and $1_n^{\mathrm{T}} X_* = 0$; this splits $\mathbb{R}^n$ into $\mathrm{span}(1_n)$ and its orthogonal subspace. Is the interpretation of β in this new parametrisation the same as in the original model?

(b) Deduce that both minimisation problems

$$\min_{\beta_0,\beta} \|y - \beta_0 1_n - X\beta\|_2^2 + \lambda p(\beta), \quad \min_{\beta} \|y_* - X_*\beta\|_2^2 + \lambda p(\beta)$$

give the same estimate $\widehat{\beta}_\lambda$, and conclude that ridge regression and lasso fits to centred response and design matrices need not include the intercept.

(c) Show that if β_0 is included in β and is penalised, then the problem is not invariant to response transformations of the form $y \mapsto ay + b1_n$ for constants $a > 0$ and b. Explain why invariance to such transformations is desirable.