Assignment sheet 12

December 13, 2017

Consider a matrix $Z_{n\times q}$ with centred columns ($Z^T\mathbf{1}_n = 0_q$). We are interested in estimating the parameter β in the model

$$y = X\beta + \epsilon = \beta_0\mathbf{1} + Z\gamma + \epsilon, \qquad X = [\mathbf{1}\ Z], \quad \beta_0 \in \mathbb{R}, \quad \gamma \in \mathbb{R}^q, \quad \beta^T = (\beta_0, \gamma^T) \in \mathbb{R}^{q+1}.$$

The parameter $\lambda > 0$ (sometimes one can consider the case $\lambda = 0$) is the penalisation parameter in ridge regression or in the lasso. Since the objective functions are convex in γ (in fact, in β as well), a local minimum is a global minimum.

Assignment 1. Observe that the Ridge estimator is a function of the smoothing parameter λ.

$$\widehat{\beta}_0 = \overline{y}, \quad \widehat{\gamma}_\lambda = (Z^tZ + \lambda I)^{-1}Z^ty.$$

(i). Using the SVD decomposition of $Z = U_{n\times n}\Sigma_{n\times q}V^t_{q\times q}$ with $\Sigma = \mathrm{diag}(\omega_1, \ldots, \omega_q)$, show that

$$\widehat{\gamma}_\lambda = V(\Sigma^t\Sigma + \lambda I)^{-1}\Sigma^tU^ty.$$

(ii). Conclude that for the fitted values of the Ridge regressions holds

$$\hat{y}_{\text{ridge}} = \overline{y}\mathbf{1} + \sum_{j=1}^{q} \frac{\omega_j^2}{\omega_j^2 + \lambda} u_j(u_j^Ty), \tag{1}$$

where u_j are the eigenvectors of ZZ^T.

Hint : You need to observe that a certain matrix is diagonal

(iii). Let $\lambda > 0$. What is the impact on $\widehat{y}_{\text{ridge}}$ of the ω_j which are close to 0 ?

Assignment 2. (i). Let $Z = U\Sigma V^t$ the SVD decomposition of Z. Show that

$$\widehat{\gamma} = \sum_{j=1}^{q} \frac{\omega_j}{\omega_j^2 + \lambda}(u_j^ty)v_j.$$

(ii). Show that

$$\widehat{\gamma}^t\widehat{\gamma} = \sum_{j=1}^{q}\left(\frac{\omega_j}{\omega_j^2 + \lambda}\right)^2 (u_j^ty)^2.$$

Hint : use what you know on the v_j.

(iii). Conclude that $\lambda \mapsto \|\widehat{\beta}_{\text{ridge}}\|_2^2$ is a decreasing function of λ.

Assignment 3. Let $\lambda^* = 2\max_{1\le j\le q}|Z_j^Ty|$. We would like to show that

$$\begin{cases} \lambda > \lambda^* \implies & \widehat{\gamma}_{\text{lasso}} = 0 \\ \lambda < \lambda^* \implies & \widehat{\gamma}_{\text{lasso}} \neq 0. \end{cases}$$

Let $f(\gamma)$ be the lasso objective function, and let $g(\gamma) = f(\gamma) - \lambda\|\gamma\|_1$. The idea is to check how the objective value behaves around 0. We consider g by its derivative at 0, where as the nondifferentiable L_1 norm will require a direct inspection.

(a) Define the centred data $y^* = y - \overline{y}\mathbf{1}$. Show that

$$g(\gamma) = \sum_{i=1}^{n} \left(y_i^* - \sum_{j=1}^{q} Z_{ij}\gamma_j \right)^2 .$$

(b) Show that

$$\frac{\partial g}{\partial \gamma_j}(0) = -2Z_j^T y, \qquad j = 1, \ldots, q.$$

(c) Suppose that $\lambda < \lambda^*$. Then there exists j such that $2|Z_j^T y| > \lambda$. Show that zero is not a local minimum of f. *Hint : let* $e_j \in \mathbb{R}^q$ *be the* j*-th unit vector and consider* $f(te_j)$ *for* t *small.*

(d) Suppose that $\lambda > \lambda^*$. Show that 0 is the unique minimiser of f. *Hint : use the convexity* $g(v) \geq g(0) + [\nabla g(0)]^T v$ *and Hölder's inequality* $|u^T v| \leq \|u\|_\infty \|v\|_1$.

Assignment 4. Unlike ridge regression, the lasso solutions are not always unique. However, the fitted values are : let $\widehat{\beta}_1$ and $\widehat{\beta}_2$ be two solutions of the lasso (for the same λ).

(a) Show that $X\widehat{\beta}_1 = X\widehat{\beta}_2$. *Hint : it suffices to deal with the estimators of* γ *(why ?). Use strict convexity again.*

(b) Show that if $\lambda > 0$, then $\|\widehat{\beta}_1\|_1 = \|\widehat{\beta}_2\|_1$.

(c) Show that if

$$Z = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}, \qquad y^T = (1, -1), \qquad \lambda = 1$$

then solutions are not unique.

Assignment 5. Show that the gamma density

$$f(y; \mu, \nu) = \frac{1}{\Gamma(\nu)} y^{\nu-1} \left(\frac{\nu}{\mu}\right)^{\nu} \exp(-\nu y/\mu), \quad y > 0, \quad \nu, \mu > 0, \tag{2}$$

can be put in form

$$f(y; \theta, \phi) = \exp\left\{ \frac{y\theta - b(\theta)}{\phi} + c(y; \phi) \right\}. \tag{3}$$

with canonical parameter $\theta = -\mu^{-1}$, $b(\theta) = -\log(-\theta)$ and dispersion parameter $\phi = 1/\nu$. Give the canonical link function, and use $b(\theta)$ to find the mean and variance function.

Assignment 6. Verify that the inverse Gaussian density

$$f(y; \lambda, \mu) = \left(\frac{\lambda}{2\pi y^3}\right)^{1/2} \exp\left\{ -\frac{\lambda(y-\mu)^2}{2\mu^2 y} \right\}, \qquad y > 0, \quad \lambda > 0, \mu > 0,$$

can be written in form (3) by giving θ, $b(\theta)$, ϕ, and $c(y; \phi)$, and show that $Var(Y) = \mu^3/\lambda$.

Assignment 7.

(a) Show that the binomial density

$$f(r;\pi) = \binom{m}{r}\pi^r(1-\pi)^{m-r}, \quad 0 < \pi < 1, \quad r = 0,\ldots,m.$$

may be written as

$$exp\left[m\left\{\frac{r}{m}\log\left(\frac{\pi}{1-\pi}\right) + \log(1-\pi)\right\} + \log\binom{m}{r}\right]$$

(b) Hence show that a binomial density is a generalized linear model density (3) and give $\theta, b(\theta), \phi$, and $c(y;\phi)$ ($y = r/m$).

(c) Find the variance function for Y.

Assignment 8. If X is a Poisson variable with mean $\mu = \exp(x^t\beta)$ and Y is a binary variable indicating the event $X > 0$, find the link function between $\mathbb{E}(Y)$ and $x^t\beta$.

Assignment 9. `bliss` contains data on mortality of flour-beetles as a function of dose of a poison. To plot the death rates :

```
bliss
attach(bliss)
plot(log(dose),r/m,ylim=c(0,1),ylab="Proportion dead")
fit <- glm(cbind(r,m-r)~log(dose),binomial)
summary(fit)
points(log(dose),fitted(fit),pch="L")
```

Does the fit seem good to you ? Try again with the `probit` and `cloglog` link functions, for example :

```
fit <- glm(cbind(r,m-r)~log(dose),binomial(cloglog))
points(log(dose),fitted(fit),pch="C")
```

Which fits best ? Give an interpretation of the resulting model.