Assignment sheet 12

Spring 2025

Assignment 1 (Orthogonal variables)**.**
Consider the regression model

$$y = X\beta + \varepsilon = (X_1,\, X_2)\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \varepsilon,$$

where $X = (X_1, X_2)$, $\beta^t = (\beta_1^t, \beta_2^t)$, X_1 is $n \times p_1$, X_2 is $n \times p_2$ (both injective) such that

$$X_1^t X_2 = 0_{p_1 \times p_2}.$$

Let H_i the hat matrix associated with X_i.

(i). What is the geometrical interpretation of $X_1^t X_2 = 0$?

(ii). Compute H as a function of X_i and H_i, then compute the products

$$H_1H_2, H_2H_1, HH_1, H_1H.$$

Comment. What is their geometric interpretation ?

(iii). Show that each of the following quantities is equal to Hy :

(a) $H_1y + H_2y$;
(b) $H_1y + H_2e_1$, avec $e_1 = (I - H_1)y$;
(c) $H_1y + He_1$.

Finish by observing that the above equalities imply that for the model

$$y = X\beta + \varepsilon \qquad (M)$$

the fitted values under the full model M equal

(a) the sum of the fitted values under (M_1) and (M_2) (where the model M_i corresponds to the pair (y, X_i).
(b) The sum of the fitted values under (M_1) (with input data (y, X_1)) and of the residuals of (M_1) computed under (M_2) (with variables (e_1, X_2)).
(c) The sum of the fitted values under (M_1) (with variables (y, X_1)) and of the residuals of (M_1) computed under (M) (with variables (e_1, X)).

Assignment 2 (Orthogonal variables and ANOVA)**.**
Consider the regression model

$$y = X\beta + \varepsilon = (X_1, \ldots, X_k)\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \varepsilon$$

where X_i is $n \times p_i$, all the X_i are injectives and

$$i \neq j \implies X_i^t X_j = 0.$$

Let H be the hat matrix associated with X, H_i the hat matrix associated with X_i and $\hat\beta = (X^tX)^{-1}X^ty = (\hat\beta_1^t, \ldots, \hat\beta_k^t)^t$. Denote by δ_{ij} the Kronecker delta : $\delta_{ij} = 1$ if $i = j$, and 0 otherwise. For a set $L \subset \{1, \ldots, k\}$ we define $X_L = (X_i : i \in L)$ and $\hat{\beta}_L = (\hat{\beta}_i^t : i \in L)^t$. For example if $L = \{1, 2, 4\}$, $X_L = (X_1, X_2, X_4)$ et

$$\hat{\beta}_L = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_4 \end{pmatrix}.$$

Define $RSS_L = \|y - H_L y^2\|$, où $H_L = X_L(X_L^t X_L)^{-1} X_L^t$.

(i). Show that $H = H_1 + \cdots + H_k$ and that $H_L = \sum_{i \in L} H_i$.

(ii). Show that $H_i H_j = \delta_{ij} H_i$.

(iii). Show that $\hat{\beta}_j = (X_j^t X_j)^{-1} X_j^t y$.

(iv). For $j \notin L$, compute

$$RSS_L - RSS_{L \cup \{j\}},$$

and show that such expression doesn't depend on L.

(v). What is the interpretation of point 4 w.r.t. ANOVA ?

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

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

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

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

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

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

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

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

(ii). Show that

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

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 5. Let $\lambda^* = 2\max_{1\leq j\leq q}|Z_j^T y|$. 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

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

(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 6. 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.