1 TD : Linear Models

Exercise 1 OLS and re-centering

1. What are the vectors $\mathbf{y} \in \mathbb{R}^n$ such that $var_n(\mathbf{y}) = 0$ (var_n is the empirical variance) ?
2. For a matrix $X \in \mathbb{R}^{n\times p}$, what is Ker $\left(X^\top X\right)$ in terms of Ker (X) ?
3. For a matrix $X \in \mathbb{R}^{n\times(p+1)}, n \geqslant 1$ and $p \geqslant 1$, which a first column full of 1's, denote $\left(1, \tilde{x}_i^T\right)^T \in \mathbb{R}^{p+1}$, the rows of X.Show that $X^T X$ non-invertible is equivalent to

$$\operatorname{cov}_n(\tilde{X}) = n^{-1} \sum_{i=1}^n \left(\tilde{x}_i - \hat{\mu}_n\right)\left(\tilde{x}_i - \hat{\mu}_n\right)^T \qquad \text{non-invertible}$$

   where $\hat{\mu}_n = n^{-1} \sum_{i=1}^n \tilde{x}_i$
4. Assume that $cov_n(\tilde{X})$ is invertible. Give the OLS esimate on centered data.

Exercise 2 OLS and scaling/invariance

We assume X is of full rank and denote $\hat{\boldsymbol{\theta}}_n$ the OLS estimate. Denote $\tilde{X} = \left(\tilde{X}_1, \ldots, \tilde{X}_p\right)$ and $X = \left(1_n, \tilde{X}_1, \ldots, \tilde{X}_k, \ldots, \tilde{X}_p\right)$. We change the scale of one variable : $\tilde{X}_k$ is replaced by $\tilde{X}_k b$, with $b > 0$

1. Let $X_b = \left(1_n, \tilde{X}_1, \ldots, \tilde{X}_k b, \ldots, \tilde{X}_p\right)$. Show that $X_b = XD$ where D is a diagonal matrix to explicit.
2. Denote $\hat{\boldsymbol{\theta}}_{b,n}$ the OLS estimate associated to X_b. Write $\hat{\boldsymbol{\theta}}_{b,n}$ in terms of $\hat{\boldsymbol{\theta}}_n$ and D.
3. Give the covariance/variance of $\hat{\boldsymbol{\theta}}_{b,n}$.
4. We have seen that $\hat{\boldsymbol{\theta}}_n$ is affected by a rescaling. What can we say about the predicted value of the model ?
5. Does the last conclusion hold with the RIDGE estimate ? What does in mean in practice ?

Exercise 3 Prediction Intervals

For $i = 1, \ldots, n$, we consider $y_i \in \mathbb{R}$ and $x_i = (x_{i,0}, \ldots, x_{i,p})^T \in \mathbb{R}^{p+1}$ with $x_{i,0} = 1$. The OLS estimator is any coefficient vector $\hat{\boldsymbol{\theta}}_n = \left(\hat{\boldsymbol{\theta}}_{n,0}, \ldots, \hat{\boldsymbol{\theta}}_{n,p}\right)^T \in \mathbb{R}^{p+1}$ such that

$$\hat{\boldsymbol{\theta}}_n \in \underset{\boldsymbol{\theta}\in\mathbb{R}^{p+1}}{\arg\min} \sum_{i=1}^n \left(y_i - x_i^T \boldsymbol{\theta}\right)^2$$

With the notations

$$X = \begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix} = \begin{pmatrix} x_{1,0} & \ldots & x_{1,p} \\ \vdots & & \vdots \\ x_{n,0} & \ldots & x_{n,p} \end{pmatrix} \in \mathbb{R}^{n\times(p+1)}, \quad Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

We have

$$\hat{\boldsymbol{\theta}}_n \in \underset{\boldsymbol{\theta}\in\mathbb{R}^{p+1}}{\arg\min} \|Y - X\boldsymbol{\theta}\|_2^2$$

We assume the following Gaussian model, for all $i = 1, \ldots, n, y_i = x_i^T \boldsymbol{\theta}^\star + \epsilon_i$ with $(\epsilon_i) \sim_{iid} \mathcal{N}\left(0, \sigma^2\right)$ such that $\ker(X) = \{0\}$

1. Let $x = \left(1, \tilde{x}^T\right)^T$ with $\tilde{x} \in \mathbb{R}^p$. Give $\hat{p}(x)$ the predicted value at x by the OLS.
2. Give the distribution of $\hat{p}(x)$. The mean $p(x)$ and variance $u(x)$ should be made explicit.
3. Define
$$\hat{\sigma}_n^2 = \frac{1}{n-(p+1)} \sum_{i=1}^{n} \left(y_i - x_i^T \hat{\theta}_n\right)^2$$
and recall that $\left(\hat{\sigma}_n^2/\sigma^2\right)(n-(p+1))$ is independent from $\hat{\theta}_n$ and follows a chi-squared distribution with $n-(p+1)$ degrees of freedom. Show that
$$\frac{(\hat{p}(x) - p(x))}{\hat{\sigma}_n \sqrt{\left(x^T \left(X^T X\right)^{-1} x\right)}} \sim t(n-(p+1))$$
where $t(k)$ is the Student distribution with k degrees of freedom $\left(\text{hint } \mathcal{N}(0,1)/\sqrt{\chi_k^2/k} \simeq t(k) \text{ if } \mathcal{N}(0,1) \perp \chi_k^2\right)$
4. Let y be the output associated to the predictor x. The value y is supposed to be independent from the sample (y_i). Show that
$$\frac{y - \hat{p}(x)}{\hat{\sigma}_n \sqrt{1 + \left(x^T \left(X^T X\right)^{-1} x\right)}} \sim t(n-(p+1))$$
5. Build confidence intervals for $p(x)$ and Y. The last one is often called prediction interval.

Exercise 4 Under the same setting as the previous exercise

1. Show that the ridge is unique and that $\hat{\boldsymbol{\theta}}_n^{(rdg)} = \left(X^T X + \lambda I_p\right)^{-1} X^T Y$.
2. Give the bias and the variance of the Ridge.
3. In the Gaussian regression model, show that $\hat{\boldsymbol{\theta}}_n^{(rdg)}$ is distributed according to a normal distribution of which the mean and variance shall be specified.
4. Let $k \in \{1, \ldots, p\}$ and $\alpha \in (0, 1/2)$. Assuming that the variance of the noise is $\sigma^2 = 1$ and that the bias is negligible, give a confidence interval for $\hat{\boldsymbol{\theta}}_{n,k}^{(rdg)}$ with level α.