
Exercise class : Linear regression

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

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

With the notations

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

We have

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

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

Exercise 1 (prediction intervals)**.**

1) Let $x = (1, \tilde{x}^T)^T$ 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 $v(x)$ should be made explicit.

3) Show that

$$\frac{(\hat{p}(x) - p(x))}{\hat{\sigma}\sqrt{(x^T (X^T X)^{-1} x)}} \sim t(n - (p+1)).$$

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

$$\frac{y - \hat{p}(x)}{\hat{\sigma}\sqrt{1 + (x^T (X^T X)^{-1} x)}} \sim t(n - (p+1)).$$

5) Build confidence intervals for $p(x)$ and Y. The last one is often called prediction interval.

For $i = 1, \dots, n$, we consider $y_i \in \mathbb{R}$ and $x_i = (x_{i,1} \dots, x_{i,p})^T \in \mathbb{R}^p$. We assume that each x_i is deterministic. The Ridge estimator is any coefficient vector $\hat{\boldsymbol{\theta}}_n^{(rdg)} = (\hat{\boldsymbol{\theta}}_{n,1}^{(rdg)}, \dots, \hat{\boldsymbol{\theta}}_{n,p}^{(rdg)})^T \in \mathbb{R}^p$ such that

$$\hat{\boldsymbol{\theta}}_n^{(rdg)} \in \arg\min_{\boldsymbol{\theta} \in \mathbb{R}^p} \sum_{i=1}^{n} (y_i - x_i^T \boldsymbol{\theta})^2 + n\lambda \sum_{k=1}^{p} \boldsymbol{\theta}_k^2.$$

where $\lambda > 0$. With the notations

$$X = \begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix} = \begin{pmatrix} x_{1,1} & \dots & x_{1,p} \\ \vdots & & \vdots \\ x_{n,1} & \dots & x_{n,p} \end{pmatrix} \in \mathbb{R}^{n \times p}, \qquad Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

We have

$$\hat{\boldsymbol{\theta}}_n^{(rdg)} \in \arg\min_{\boldsymbol{\theta} \in \mathbb{R}^p} \|Y - X\boldsymbol{\theta}\|^2 + n\lambda \|\boldsymbol{\theta}\|_2^2.$$

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

Exercise 2 (ridge)**.**

1) *Show that the ridge is unique and that* $\hat{\boldsymbol{\theta}}_n^{(rdg)} = (X^T X + n\lambda I_p)^{-1} X^T Y$.
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* $k \in \{1, \dots, p\}$ *and* $\alpha \in (0, 1/2)$. *Assuming that the variance of the noise is* $\sigma^2 = 1$, *give a confidence interval for* $\hat{\boldsymbol{\theta}}_{n,k}^{(rdg)}$ *with level* α.