TELECOM Paris

Exercise class : Linear regression

Exercise 1 (MCO et re-centrage).

1) *Quels sont les vecteurs* $\mathbf{y} \in \mathbb{R}^n$ *tels que* $\mathrm{var}_n(\mathbf{y}) = 0$ *(*var_n *est la variance empirique) ?*
2) *Pour une matrice* $X \in \mathbb{R}^{n\times p}$*, que vaut* $\mathrm{Ker}(X^\top X)$ *?*
3) *Pour une matrice* $X \in \mathbb{R}^{n\times(p+1)}$*,* $n > 1$ *et* $p \geqslant 1$*, qui possède comme première colonne une colonne de 1, notons* $(1, \tilde{x}_i^\top) \in \mathbb{R}^{p+1}$*, les lignes de* X*. Montrer que* $X^\top X$ *non-inversible est équivalent à*

$$\mathrm{cov}_n(\tilde{X}) = n^{-1}\sum_{i=1}^{n}(\tilde{x}_i - \hat{\mu}_n)(\tilde{x}_i - \hat{\mu}_n)^T \quad \textit{non-inversible,}$$

   où $\hat{\mu}_n = n^{-1}\sum_{i=1}^n \tilde{x}_i$*.*
4) *Supposons que* $\mathrm{cov}_n(\tilde{X})$ *soit inversible. Exrpimer l'estimateur des moindres carrée calculé sur des variables centrées.*

Exercise 2 (MCO et invariance).

On suppose que X *est de rang plein et on note* $\hat{\boldsymbol{\theta}}_n$ *l'estimateur OLS. On note* $\tilde{X} = (\tilde{X}_1, \ldots, \tilde{X}_p)$*. On change l'échelle d'une des variables :* $\tilde{X}_k$ *est remplacé par* $\tilde{X}_k b$*, où* $b > 0$*.*

1) *Soit* $X_b = (1, \tilde{X}_1, \ldots, \tilde{X}_k b, \ldots, \tilde{X}_p)$*. Montrer que* $X_b = XD$ *où* D *est une matrice diagonale que l'on précisera.*
2) *Soit* $\hat{\boldsymbol{\theta}}_{b,n}$ *l'estimateur OLS associé à* X_b*. Exprimer* $\hat{\boldsymbol{\theta}}_{b,n}$ *en fonction de* $\hat{\boldsymbol{\theta}}_n$ *et* D*.*
3) *Donner la variance de* $\hat{\boldsymbol{\theta}}_{b,n}$*.*
4) *On a vu que l'estimateur* $\hat{\boldsymbol{\theta}}_n$ *était affecté par un changement d'échelle. Qu'en est-il de la valeur prédite par le modèle ?*
5) *Peut-on dire la même chose de l'estimateur Ridge ?*

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* $\hat{\boldsymbol{\theta}}_n = (\hat{\boldsymbol{\theta}}_{n,0}, \ldots, \hat{\boldsymbol{\theta}}_{n,p})^T \in \mathbb{R}^{p+1}$ *such that*

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

With the notations

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

We have

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

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

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* $u(x)$ *should be made explicit.*

3) *Define*

$$\hat{\sigma}_n^2 = \frac{1}{n-(p+1)} \sum_{i=1}^{n} (y_i - x_i^T \hat{\theta}_n)^2,$$

and recall that $\hat{\sigma}_n^2(n-(p+1))$ *is independent from* $\hat{\theta}_n$ *and follows a chi-squared distribution with* $n-(p+1)$ *degrees of freedom . Show that*

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

where $t(k)$ *is the Student distribution with* k *degrees of freedom (hint* $\mathcal{N}(0,1)/\sqrt{\chi_k^2/k} \simeq t(k)$ *if* $\mathcal{N}(0,1) \perp \chi_k^2$*).*

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*

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

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

Exercise 4 (ridge)**.** *Under the same setting as the previous exercise*

1) *Show that the ridge is unique and that* $\hat{\boldsymbol{\theta}}_n^{(rdg)} = (X^TX + \lambda I_p)^{-1}X^TY$*.*
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* $\sigma^2 = 1$*, give a confidence interval for* $\hat{\boldsymbol{\theta}}_{n,k}^{(rdg)}$ *with level* α*.*