TELECOM ParisTech **SD-TSIA204**

2017/2018 **Joseph Salmon, François Portier**

Exam : Linear Models

Reminder : no document authorized

Exercise 1. (Quiz (12 points))

1) What is the orthogonal projection of a vector $\mathbf{y} \in \mathbb{R}^n$ over $\mathrm{Vect}(1_n)$, where $1_n = (1, \ldots, 1)^\top \in \mathbb{R}^n$?

2) Express the expectation of $\frac{1}{n} \sum_{i=1}^{n} y_i (n \overline{y}_n - x_i)$ in terms of $\mathbb{E}[y_1^2]$ and $\mathbb{E}[x_1 y_1]$, ($\overline{y}_n$ is the empirical mean), (y_i, x_i) are i.i.d with finite variance.

3) Let $\hat{\boldsymbol{\theta}}_n$ be the OLS associated with $(\mathbf{y}, X)$ where $X = (1_n, \tilde{X})$ with $1_n = (1, \ldots, 1)^\top \in \mathbb{R}^n$ and $\tilde{X} \in \mathbb{R}^{n \times (p-1)}$. Show that $\sum_{i=1}^{n} (y_i - x_i^T \hat{\boldsymbol{\theta}}_n) = 0$ (one might express the normal equation).

- Consider the following problem :

$$\hat{\boldsymbol{\theta}}_n^{\Omega} \in \arg\min_{\boldsymbol{\theta} \in \mathbb{R}^p} \frac{1}{2} (\mathbf{y} - X\boldsymbol{\theta})^\top \Omega (\mathbf{y} - X\boldsymbol{\theta}) \ ,$$

where $\mathbf{y} = (y_1, \ldots, y_n)^\top \in \mathbb{R}^n$, $X \in \mathbb{R}^{n \times p}$ and Ω is a strictly positive symmetric.

4) Show that there exists $\Omega^{1/2}$ such that $\Omega^{1/2}\Omega^{1/2} = \Omega$ and that if $\ker(X) = 0$, then $\ker(\Omega^{1/2} X) = 0$.

5) Suppose $\ker(X) = 0$, give an expression for $\hat{\boldsymbol{\theta}}_n^{\Omega}$. Is it unique ? Express the prediction $\hat{\mathbf{y}} = X\hat{\boldsymbol{\theta}}_n^{\Omega}$ with respect to $X, \Omega, \mathbf{y}$.

- Ridge notation in what follows : for any $\lambda > 0$,

$$\hat{\boldsymbol{\theta}}^{\mathrm{Ridge},\lambda} = \arg\min_{\boldsymbol{\theta} \in \mathbb{R}^p} \frac{1}{2n} \|\mathbf{y} - X\boldsymbol{\theta}\|_2^2 + \frac{\lambda}{2} \|\boldsymbol{\theta}\|_2^2 \ .$$

6) Give a closed-form solution for the ridge estimator $\hat{\boldsymbol{\theta}}^{\mathrm{Ridge},\lambda}$ w.r.t. $\mathbf{y}$ and λ when $X = \mathrm{Id}_n$ (here $n = p$).

7) Using the SVD of X, show that $(X^\top X + \lambda \mathrm{Id}_p)^{-1} X^\top \mathbf{y} = X(XX^\top + \lambda \mathrm{Id}_n)^{-1}\mathbf{y}$ for any $X \in \mathbb{R}^{n\times p}$ and $\mathbf{y} \in \mathbb{R}^n$.

8) Assuming that the noise model is as follows : $\boldsymbol{\varepsilon} = \mathbf{y} - X\boldsymbol{\theta}^\star$ is a centered random vector with covariance matrix $\sigma^2 \mathrm{Id}_n$. What is the covariance matrix of the Ridge estimator $\hat{\boldsymbol{\theta}}^{\mathrm{Ridge},\lambda}$?

9) For any matrix $D \in \mathbb{R}^{q\times p}$ satisfying $\mathrm{Ker}(D) = \{0\}$, give a closed-form solution of the following problem

$$\hat{\boldsymbol{\theta}}^{D,\lambda} = \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \frac{1}{2}\|\mathbf{y} - X\boldsymbol{\theta}\|_2^2 + \lambda \|D\boldsymbol{\theta}\|_2^2 \ .$$

10) Give a closed-form solution of $\eta_\lambda(z) = \arg\min_{x\in\mathbb{R}} x \mapsto \dfrac{1}{2n}(z-x)^2 + \lambda|x|$ w.r.t. z, $\mathrm{sign}(z)$, λ and the positive part function $(\cdot)_+$.

11) Give a closed-form solution of $\eta_\lambda(z) = \arg\min_{x\geqslant 0} x \mapsto \dfrac{1}{2}(z-x)^2 + \lambda|x|$ w.r.t. z and λ.

12) Consider the regression model where X is deterministic and $\boldsymbol{\varepsilon} = \mathbf{y} - X\boldsymbol{\theta}^\star$ is a Gaussian, centered, with covariance matrix $\sigma^2 \mathrm{Id}_n$. Describe a test procedure (the test statistic and the reject region) to assess $H_0 : \theta_1 = \theta_2$. What is the distribution of your statistic under H_0 ?

EXERCISE 2. (On-line OLS and cross validation (8 points))

The goal of this exercise is to show that the OLS estimator $\hat{\boldsymbol{\theta}}_n$ associated with design matrix $X_{(n)} \in \mathbb{R}^{n\times p}$ and output $\mathbf{y}_{(n)} \in \mathbb{R}^n$ can be easily updated when a new pair of observation $(\mathbf{x}_{n+1}, y_{n+1}) \in \mathbb{R}^p \times \mathbb{R}$ is given. We apply the result to cross validation procedure in the end.

To clarify the notation :

$$X_{(n+1)} = \begin{pmatrix} X_{(n)} \\ \mathbf{x}_{n+1}^\top \end{pmatrix} \in \mathbb{R}^{(n+1)\times p}, \quad \text{and} \quad \mathbf{y}_{(n+1)} = \begin{pmatrix} \mathbf{y}_{(n)} \\ y_{n+1} \end{pmatrix} \in \mathbb{R}^{n+1}$$

We assume from now on that $X_{(n)}$ and $X_{(n+1)}$ are full column rank (*i.e.*, the columns of each matrix are independent vectors).

<u>NB</u> : Some questions require harder computation than others (in particular obtaining (1) and (3)). Even if you could not prove the results, they can be use later.

1) Let A, B, C, D be matrices with respective sizes (d, d), (d, k), (k, k), (k, d). Show that if A and C are invertible, then

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(DA^{-1}B + C^{-1})^{-1}DA^{-1}. \qquad (1)$$

NB : this result is called the “Sherman–Morrison–Woodbury formula”.

2) Obtain that

$$(X_{(n+1)}^\top X_{(n+1)})^{-1} = (X_{(n)}^\top X_{(n)})^{-1} - \frac{\boldsymbol{\zeta}_{n+1}\boldsymbol{\zeta}_{n+1}^\top}{1+b_{n+1}} \tag{2}$$

where $\boldsymbol{\zeta}_{n+1} = (X_{(n)}^\top X_{(n)})^{-1}\mathbf{x}_{n+1}$ and $b_{n+1} = \mathbf{x}_{n+1}^\top (X_{(n)}^\top X_{(n)})^{-1}\mathbf{x}_{n+1}$.

3) Express $X_{(n+1)}^\top \mathbf{y}_{(n+1)}$ with respect to $X_{(n)}^\top \mathbf{y}_{(n)}$ and $y_{n+1}\mathbf{x}_{n+1}$.

4) Show that the OLS estimator $\hat{\boldsymbol{\theta}}_{n+1}$ associated with design matrix $X_{(n+1)}$ and output $\mathbf{y}_{(n+1)}$ can be obtained as follows :

$$\hat{\boldsymbol{\theta}}_{n+1} = \hat{\boldsymbol{\theta}}_n + \frac{u_{n+1}}{1+b_{n+1}}\boldsymbol{\zeta}_{n+1}, \tag{3}$$

where $u_{n+1} = y_{n+1} - \mathbf{x}_{n+1}^\top \hat{\boldsymbol{\theta}}_n$.

5) Assuming $(X_{(n)}^\top X_{(n)})^{-1}$ and $\hat{\boldsymbol{\theta}}_n$ have been stored, explain how to update $\hat{\boldsymbol{\theta}}_{n+1}$ using a minimal number of operations of the kind : matrix-vector multiplications (for matrices of size $p \times p$ and vector of size p). How many such operations are needed ?

6) Using Equation (2) above, show that

$$1 + b_{n+1} = \frac{1}{1-h_{n+1}}$$

where $h_{n+1} = \mathbf{x}_{n+1}^\top (X_{(n+1)}^\top X_{(n+1)})^{-1}\mathbf{x}_{n+1}$.

7) The prediction of y_{n+1} given by the model is $\hat{y}_{n+1} := \mathbf{x}_{n+1}^\top \hat{\boldsymbol{\theta}}_{n+1}$. With the following formula

$$\hat{y}_{n+1} = \mathbf{x}_{n+1}^\top \hat{\boldsymbol{\theta}}_n + \frac{u_{n+1}b_{n+1}}{1+b_{n+1}}.$$

prove that

$$y_{n+1} - \hat{y}_{n+1} = u_{n+1}(1-h_{n+1}).$$

8) Given some data $(\mathbf{y}, X)$, leave-one-out cross-validation consists in computing the risk

$$R_{cv} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\theta}}_{(-i)})^2$$

where $\hat{\boldsymbol{\theta}}_{(-i)}$ is the OLS estimator based on $(\mathbf{y}_{(-i)}, X_{(-i)})$, *i.e.,* the data $(\mathbf{y}, X)$ without the i-th line. Applying what have been done so far, show that

$$R_{cv} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2/(1-\hat{h}_i)^2,$$

with $\hat{h}_i = \mathbf{x}_i^\top (X^\top X)^{-1}\mathbf{x}_i$ and $\hat{y}_i = \mathbf{x}_i^\top \hat{\boldsymbol{\theta}}_n$, $\hat{\boldsymbol{\theta}}_n$ being the OLS estimator of $(\mathbf{y}, X)$.