TELECOM ParisTech **SD-TSIA204**
2017/2018 **Joseph Salmon, François Portier**

Quiz : Linear Models

---

1) For any deterministic $\mu \in \mathbb{R}^p$ and any random variable $X \in \mathbb{R}^p$ express $Cov(X + \mu)$.

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

3) For any matrix $A \in \mathbb{R}^{m \times p}$ and any random vector $X \in \mathbb{R}^p$, express $Cov(AX)$.

4) What is the bias the of $\displaystyle\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y}_n)^2$ ($\overline{y}_n$ is the empirical mean) for i.i.d y_i Gaussian variables centered and with variance σ^2 ?

5) Let $y_1, \ldots, y_n$ be random Gaussian variables i.i.d., centered with variance σ^2. What is the quadratic risk of $\displaystyle\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y}_n)^2$ as an estimator of σ^2 ($\overline{y}_n$ is the empirical mean) ?

6) What are the vectors $\mathbf{y} \in \mathbb{R}^n$ such that $\mathrm{var}_n(\mathbf{y}) = 0$ (var_n(where $\mathbf{y}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y}_n)^2$ is the empirical variance) ?

- Least-squares notation : we write $\mathbf{y} = (y_1, \ldots, y_n)^\top \in \mathbb{R}^n$ and $X \in \mathbb{R}^{n \times p}$, and

$$\hat{\boldsymbol{\theta}}^{\mathrm{OLS}} \in \operatorname*{arg\,min}_{\boldsymbol{\theta} \in \mathbb{R}^p} \frac{1}{2}\|\mathbf{y} - X\boldsymbol{\theta}\|_2^2 \ . \tag{1}$$

7) Let $y_1, \ldots, y_n$ and $x_1, \ldots, x_n$ be real numbers. Is the following function convex or concave ?

$$\begin{aligned} \mathbb{R}^2 &\to \mathbb{R} \\ (\theta_0, \theta_1) &\mapsto \frac{1}{2}\sum_{i=1}^{n}(y_i + 3\theta_0 - \theta_1 x_i)^2 \ . \end{aligned}$$

8) Write a pseudo-code to perform the gradient descent algorithm for solving the least squares problem given in Eq. (1) (with input $X, \mathbf{y}$ and α being the step size).

9) For any $X \in \mathbb{R}^{n \times p}$ express $\mathrm{Ker}(X^\top X)$ in terms of $\mathrm{Ker}(X)$.

10) Let $X \in \mathbb{R}^{n \times n}$ satisfies $X^\top X = \mathrm{Id}_n$ (Id_n being the identity matrix). Can you provide a closed-form solution for the least squares and show it is unique ?

11) Let $X \in \mathbb{R}^{n \times p}$ be a full (column) rank matrix. What is the covariance (matrix) of the least squares estimator (assuming that the noise model is as follows : $\boldsymbol{\varepsilon} = \mathbf{y} - X\boldsymbol{\theta}^\star$ is a centered random vector with covariance matrix $\sigma^2 \mathrm{Id}_n$).

12) Express the pseudo inverse of X thanks to its SVD : $X = \sum_{i=1}^r s_i \mathbf{u}_i \mathbf{v}_i^\top$, with $r = \mathrm{rg}(X)$ and $s_1 \geqslant \cdots \geqslant s_r > 0$.

13) Give an explicit solution of the following problem :

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

for positive-definite matrix $\Omega = \mathrm{diag}(w_1, \dots, w_n)$, in the case where X is full (column) rank.

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

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

14) Give a closed-form solution for the ridge estimator $\hat{\boldsymbol{\theta}}^{\mathrm{Ridge},\lambda}$ w.r.t. $X, \mathbf{y}$ and λ.

15) 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 $n = p$).

16) 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 $\hat{\boldsymbol{\theta}}^{\mathrm{Ridge},\lambda}$?

17) Give a closed-form solution of the following problem w.r.t.

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

18) 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, $\mathrm{sign}(z)$ and the positive part function $(\cdot)_+$.

19) What is the sub-differential of the real function $x \mapsto \max(-2x, 0)$?

20) Provide the main step in the coordinate descent algorithm to solve the Elastic Net problem :

$$\hat{\boldsymbol{\theta}}^{\mathrm{ENET},\lambda,\alpha} = \arg\min_{\boldsymbol{\theta} \in \mathbb{R}^p} \left[\frac{1}{2} \|\mathbf{y} - X\boldsymbol{\theta}\|_2^2 + \lambda \left(\alpha \|\boldsymbol{\theta}\|_1 + (1 - \alpha) \frac{\|\boldsymbol{\theta}\|_2^2}{2} \right) \right] \ .$$

21) Provide the main step in the coordinate descent algorithm to solve the Positive Lasso :

$$\hat{\boldsymbol{\theta}}^{\text{Lasso}+,\lambda} = \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p_+} \frac{1}{2}\|\mathbf{y} - X\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \ .$$

22) Let us assume one has a Lasso solver Lasso$(X, \mathbf{y}, \lambda)$ that solve the following problem

$$\hat{\boldsymbol{\theta}}^{\text{Lasso},\lambda} = \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \frac{1}{2n}\|\mathbf{y} - X\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \ .$$

What transformation on X can you perform to solve the following problem with positive $w_1, \dots, w_p$.

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

23) What is the solution of $\begin{cases} \max\limits_{u\in\mathbb{R}^n, v\in\mathbb{R}^p} & u^\top X v \\ \text{s.c. } \|u\|_2^2 = 1 \text{ et } \|v\|_2^2 = 1 \end{cases}$?

24) For $X_1, \dots, X_n$ i.i.d. with values in $\{0, 1\}$, propose a procedure to test the hypothesis $p = P(X_1 = 1) = 1/2$.

25) In the regression model, assuming that X is deterministic and that $\boldsymbol{\varepsilon} = \mathbf{y} - X\boldsymbol{\theta}^\star$ is a Gaussian, centered, with covariance matrix $\sigma^2\mathrm{Id}_n$, what is the distribution of the least square $\hat{\boldsymbol{\theta}}^{\text{OLS}}$ (one could assume that X is full (column)rank here).

26) Let $X_1, \dots, X_n$ be i.i.d Gaussian variables with (unknown) mean μ and known variance σ^2, i.e., for all $i = 1, \dots, n$, $X_i \sim \mathcal{N}(\mu, \sigma^2)$. Propose a way to test the hypothesis $\mu = 1$.

27) Let $X_1, \dots, X_n$ be independent Gaussian variables with (unknown) mean μ and known variance σ_i^2, i.e., for all $i = 1, \dots, n$, $X_i \sim \mathcal{N}(\mu, \sigma_i^2)$. Propose a way to test the hypothesis $\mu = 1$.

28) Let $X_1, \dots, X_n$ be i.i.d. random variables with expectation μ and variance σ^2. Write a pseudo-code based on the bootstrap for testing if $\mu = 1$ or not.

29) Suppose that $Y \in \mathbb{R}^n$ and $X \in \mathbb{R}^{n\times p}$. Write a bootstrap procedure to estimate the variance of the ordinary least squares estimates $\hat{\beta}_k$ of the linear regression $Y \simeq X\beta$.