TELECOM ParisTech **SD-TSIA204**
2018/2019 **Pavlo Mozharovskyi and François Portier**

Quiz : Linear Models

General questions :

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 $\text{Vect}(1_n)$, where $1_n = (1, \dots, 1)^\top \in \mathbb{R}^n$?
3) For any matrix $A \in \mathbb{R}^{m \times p}$ and any random vector $X \in \mathbb{R}^p$, express $Cov(AX)$.
4) Let $V_n = \dfrac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y}_n)^2$ ($\overline{y}_n$ is the empirical mean). Compute $\mathbb{E}(V_n) - \sigma^2$ when the y_i's are i.i.d centered Gaussian variables with variance σ^2 ?
5) Let $y_1, \dots, y_n$ be random Gaussian variables i.i.d., centered with variance σ^2. What is the quadratic risk of $\dfrac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y}_n)^2$ 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 $\text{var}_n(\mathbf{y}) = 0$, where $\text{var}_n(\mathbf{y}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y}_n)^2$ is the empirical variance ?
7) What is the solution of $\begin{cases} \max\limits_{u \in \mathbb{R}^n, v \in \mathbb{R}^p} & u^\top X v \\ \text{s.c. } \|u\|_2^2 = 1 \text{ et } \|v\|_2^2 = 1 \end{cases}$?

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

$$\hat{\boldsymbol{\theta}}_n \in \arg\min_{\boldsymbol{\theta} \in \mathbb{R}^p} \frac{1}{2}\|\mathbf{y} - X\boldsymbol{\theta}\|_2^2 \ . \tag{1}$$

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

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

9) 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).

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

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

12) 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 $\sigma^2 \mathrm{Id}_n$).

13) 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 $s_1 \geqslant \cdots \geqslant s_r > 0$.

14) 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 $\Omega = \mathrm{diag}(w_1, \ldots, w_n)$, in the case where X is full (column) rank.

Tests and CI :

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

16) 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 $\sigma^2 \mathrm{Id}_n$ where σ^2 is known, what is the distribution of $\hat{\boldsymbol{\theta}}_n$ (one could assume that X is full column rank here). Based on this, provide a confidence interval for $(1, \ldots, 1)\hat{\boldsymbol{\theta}}_n$.

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

Ridge : For any $\lambda > 0$,

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

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

19) 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 $n = p$).

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

21) For any matrix $D \in \mathbb{R}^{q\times p}$ satisfying $\mathrm{Ker}(D) = \{0\}$, give a closed-form solution of the following problem
$$\hat{\boldsymbol{\theta}}^{D,\lambda} = \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \frac{1}{2}\|\mathbf{y} - X\boldsymbol{\theta}\|_2^2 + \lambda\|D\boldsymbol{\theta}\|_2^2 \ .$$

LASSO :

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

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

24) For a fixed $y \in \mathbb{R}$, and $\lambda, \alpha > 0$ provide a close form solution for solving the 1D Elastic Net problem :
$$\hat{\theta}^{\mathrm{ENET},\lambda,\alpha} = \arg\min_{\theta\in\mathbb{R}} \left[\frac{1}{2}(y-\theta)^2 + \lambda\left(\alpha|\theta| + (1-\alpha)\frac{\theta^2}{2}\right)\right] \ .$$

25) Let us assume one has a Lasso solver $\mathrm{Lasso}(X, \mathbf{y}, \lambda)$ that solves the following problem
$$\hat{\boldsymbol{\theta}}^{\mathrm{Lasso},\lambda} = \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \frac{1}{2n}\|\mathbf{y} - X\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1 \ .$$

What transformation on X can you perform to solve the following problem with positive $w_1, \ldots, w_p$:
$$\hat{\boldsymbol{\theta}}^{\mathrm{Lasson},\lambda} = \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \frac{1}{2}\|\mathbf{y} - X\boldsymbol{\theta}\|_2^2 + \lambda\sum_{j=1}^{p} w_j|\theta_j| \ .$$