TELECOM Paris
2022/2022

SD TSIA204

EXAM : Linear models

The duration of the exam is 3 hours.

General questions

1) Show that the hat matrix $H = X(X^TX)^{-1}X^T$ is an orthogonal projector onto the column space of X.
2) What is the orthogonal projection of $\mathbf{y} = (Y_1, \dots, Y_n) \in \mathbb{R}^n$ on $\text{Vect}(1_n)$, with $1_n = (1, \dots, 1)^\top \in \mathbb{R}^n$?
3) Express the pseudo inverse of X thanks to its SVD : $X = \sum_{i=1}^r s_i \mathbf{u}_i \mathbf{v}_i^\top$.
4) Show that the variance of the OLS estimator θ is $Var(\theta) = \sigma^2 \sum_{i=1}^r s_1^{-2} v_i v_i^T$.
5) Describe the "PCA before OLS" technique.
6) Let $X \in \mathbb{R}^n$ be normally distributed $X \sim \mathcal{N}(\mu_X, \Sigma_X)$ and Y an affine transformation of X, $Y = LX + u$ with $L \in \mathbb{R}^{m \times n}$, $u \in \mathbb{R}^m$ deterministic. Then Y is also normally distributed with mean μ_Y and covariance Σ_Y, $Y \sim \mathcal{N}(\mu_Y, \Sigma_Y)$. Show that $\mu_Y = u + L\mu_X$ and $\Sigma_Y = L\Sigma_X L^T$.

Ordinary Least Squares (OLS)

7) Let $Y = (Y_1, \dots, Y_n)^T \in \mathbb{R}^n$ and $X = (1_n, \tilde{X}) \in \mathbb{R}^{n \times (p+1)}$, $1_n = (1, \dots, 1)^T \in \mathbb{R}^n$. The identity matrix of dimension n is denoted as I_n. Denote the residuals as $\hat{\varepsilon} = Y - \hat{Y}$. Let $\hat{\theta} \in \arg\min_{\theta \in \mathbb{R}^{p+1}} \|Y - X\theta\|_2^2$ and $\hat{Y} = (\hat{Y}_1, \dots, \hat{Y}_n)^T = X\hat{\theta}$.
   (a) Show that $\min_{\theta \in \mathbb{R}^{p+1}} \|Y - X\theta\|_2^2 \leqslant \min_{\theta_0 \in \mathbb{R}} \|Y - 1_n\theta_0\|_2^2$
   (b) Show that $\arg\min_{\theta_0 \in \mathbb{R}} \|Y - 1_n\theta_0\|_2^2 = \overline{Y} = n^{-1}\sum_{i=1}^n Y_i$
   (c) Deduce the following inequality $\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \overline{Y})^2} \leqslant 1$
   (d) State the normal equations for the OLS and use it to show that

$$\begin{aligned} &< \hat{\varepsilon}, X > = 0 \\ &< \hat{\varepsilon}, \hat{Y} > = 0 \\ &< \hat{\varepsilon}, \overline{Y}\mathbf{1}_n > = 0. \end{aligned} \tag{1}$$

   (e) Show that we can write the R^2 as

$$R^2 = 1 - \frac{\|Y - \hat{Y}\|^2}{\|Y - \overline{Y}\mathbf{1}_n\|^2} = \frac{\|\hat{Y} - \overline{Y}\mathbf{1}_n\|^2}{\|Y - \overline{Y}\mathbf{1}_n\|^2}.$$

   (f) Elaborate on when $R = 0$ and $R = 1$.
8) Give the coordinate descent algorithm for the OLS specifying the gradients.
9) Show that if X is full column rank the OLS estimator is unique.

Ridge

10) We consider the Ridge problem in the following points,

$$\hat{\boldsymbol{\theta}}_\lambda^{\text{rdg}} = \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left(\|\mathbf{y} - X\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_2^2\right)$$

Give a closed-for expression for $\hat{\boldsymbol{\theta}}_\lambda^{\text{rdg}}$. When is it unique?

11) Give an expression for the prediction for a new point x_0 and the residuals ϵ_δ. Show that $\mathbb{E}[\epsilon_\delta] = [I_n - X(X^TX + \lambda I_n)^{-1}X^T]X\hat{\boldsymbol{\theta}}_\lambda^{\text{rdg}}$.

12) Show that the ridge regression estimator for dataset (X, Y) can be obtained by ordinary least squares regression on an augmented data set $(\tilde{X}, \tilde{Y})$. Both X and Y are augmented by adding p rows. Specify the values of those rows. X Y

LASSO. Here again $X \in \mathbb{R}^{n\times p}$.

13) Let $\Omega = \text{diag}(w_1, \ldots, w_n)$ with $w_i > 0$ for all i. Express (justifying) the coordinate descent algorithm for the following problem

$$\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} (Y - X\boldsymbol{\theta})^T\Omega(Y - X\boldsymbol{\theta}) + 2\lambda\|\boldsymbol{\theta}\|_1$$

(hint (indication) : when solving the previous problem over one single direction it can be expressed into a simpler problem where the function $\eta_\lambda(z) = \arg\min_{x\in\mathbb{R}}(z - x)^2 + 2\lambda|x|$ is useful)

14) (a) Show that 0 is a solution of the Lasso $\min_\theta \|Y - X\theta\|_2^2 + 2\lambda\|\theta\|_1$ if and only if

$$Y^TX\boldsymbol{\theta} \leqslant \boldsymbol{\theta}^TX^TX\boldsymbol{\theta}/2 + \lambda\|\boldsymbol{\theta}\|_1, \qquad \forall \boldsymbol{\theta} \in \mathbb{R}^p$$

(b) Show that for all $u = (u_1, \ldots, u_K)^T$, $v = (v_1, \ldots v_K)^T$, it holds that $|u^Tv| \leqslant \|u\|_1 \max_{1\leqslant k\leqslant K} |v_k|$

(c) Show that $Y^TX\boldsymbol{\theta} \leqslant \lambda_{max}\|\boldsymbol{\theta}\|_1$ with $\lambda_{max} = \max_{k=1,\ldots,p} |X_k^TY|$ and $X = (X_1, \ldots, X_p)$.

(d) Deduce that if $\lambda \geqslant \lambda_{max}$ then 0 is one solution of the Lasso.

(e) Deduce that if $\lambda \geqslant \lambda_{max}$, then 0 is the unique solution of the Lasso. One can start by considering $\boldsymbol{\theta} \neq 0$ such that

$$Y^TX\boldsymbol{\theta} = \boldsymbol{\theta}^TX^TX\boldsymbol{\theta}/2 + \lambda\|\boldsymbol{\theta}\|_1,$$

and then the 2 cases : $\boldsymbol{\theta} \in \ker(X)$ and its contrary.

Tests and CI

15) Chi-square test for the variance : In the framework of the linear model of Question 7, we suppose in addition that the vector of noises ϵ is Gaussian of covariance matrix $\sigma^2\mathbb{I}_n$.

(a) We are interested in the variance of the noises σ^2. Recall the expression for the unbiased estimator $\hat{\sigma}^2$ seen in class for this quantity. What is the distribution of $(n - p - 1)\hat{\sigma}^2/\sigma^2$?

(b) Deduce a confidence interval for σ^2 constructed from $\hat{\sigma}^2$ and the quantiles 0.025 and 0.975 of the probability distribution. What is the confidence level of this interval?

(c) Denote q_a, q_b the quantiles 0.025 and 0.975 respectively. Is $q_a = -q_b$?

16) We want to design a test to see if the coefficient θ_j is equal to 1.

(a) Detail the test, including : the null and alternative hypotheses, the statistic, the probability distribution of the statistic, the p-value, the rejection region for a α-level test, and the first order risk.

(b) For a given database and a given coefficient j, we obtain in this test a p-value p_1. What is the decision of the test of acceptance or rejection according to α?

(c) What is the type-I error?

17) Let $X_1, \ldots, X_n$ be Gaussian variables of known variance σ^2 and unknown mean μ. Detail a hypothesis test for $\mu > 1$.

10 $\frac{17}{10}$