Apprentissage statistique et grande dimension

TD 1 : grande dimension

Dans toute la fiche de TD $\|\cdot\|$ désigne la norme usuelle de $\mathbb{R}^p$,

$$\|\boldsymbol{x}\|^2 = \sum_{j=1}^{p} x_j^2, \text{ pour tout } \boldsymbol{x} = (x_1, ..., x_p) \in \mathbb{R}^p$$

et, pour tout $r > 0$, $\boldsymbol{x} \in \mathbb{R}^p$

$$B_p(\boldsymbol{x}, r) := \{\boldsymbol{y} \in \mathbb{R}^p, \|\boldsymbol{x} - \boldsymbol{y}\| \leq r\},$$

la boule de centre x et de rayon r. On notera également $V_p(r)$ le volume de $B_p(r)$. Pour simplifier, $B_p(r) = B_p(0, r)$.

Exercice 1 (Estimateur des moindres carrés et grande dimension)

On suppose que l'on observe des couples $\{(Y_i, \mathbf{x}_i), i = 1, \dots, n\}$ tels que Y_i est la valeur de la variable réponse et $\mathbf{x}_i = (x_i^1, \dots, x_i^d)$ d variables explicatives observées sur le i-ème individu. L'objectif de cet exercice est d'étudier les propriétés de l'estimateur des moindres carrés en grande dimension dans le modèle linéaire gaussien :

$$\mathbf{Y} = \mathbf{X}\beta^* + \boldsymbol{\varepsilon}$$

où l'on rappelle que :

— $\mathbf{Y} = (Y_1, \dots, Y_n)^t$ est un vecteur colonne de taille n,

— $\mathbf{X} = \begin{pmatrix} x_1^1 & \dots & x_1^d \\ \vdots & \vdots & \vdots \\ x_n^1 & \dots & x_n^d \end{pmatrix}$ est une matrice à n lignes et d colonnes,

— $\beta^* = (\beta_1^*, \dots, \beta_d^*)^t$ est un vecteur colonne de taille d, qui est le *paramètre* du modèle et que nous allons chercher à estimer,

— $\boldsymbol{\varepsilon}$ est un vecteur gaussien de moyenne nulle et de matrice de covariance $\sigma^2 I = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sigma^2 \end{pmatrix}$.

On supposera dans cet exercice que la matrice $\mathbf{X}$ est non aléatoire (on parle alors de régression linéaire multivariée *à design fixe*).

1. Montrer que toute solution $\overline{\beta}$ du problème d'optimisation
$$\min_{\beta \in \mathbb{R}^d} \left\{ \frac{1}{n} \sum_{i=1}^{n} (Y_i - \mathbf{x}_i \beta)^2 \right\}. \tag{1}$$
est solution de l'équation $X^t Y = X^t X \overline{\beta}$.
2. La matrice $X^t X$ est-elle toujours inversible ? Que se passe-t'il si elle ne l'est pas ? Dans la suite de l'exercice, nous supposerons que $X^t X$ est inversible et nous noterons $\widehat{\beta}^{(MCO)}$ l'unique solution du problème (1).
3. Montrer que $\widehat{\beta}^{(MCO)}$ est un vecteur gaussien dont on précisera la moyenne et la matrice de covariance.
4. Quel est le biais de $\widehat{\beta}^{(MCO)}$?
5. Calculer l'erreur quadratique moyenne de $\widehat{\beta}^{(MCO)}$
$$\mathbb{E}\left[\left\|\widehat{\beta}^{(MCO)} - \beta^*\right\|^2\right].$$
Que se passe-t'il en grande dimension (on pourra regarder le cas $\frac{1}{n} X^t X = I$ par exemple) ?

Exercice 2 (Loi uniforme en grande dimension)

On considère des vecteurs aléatoires tirés uniformément sur deux ensembles très différents : la boule unité $B_p(1)$ et le pavé $[-1,1]^p$.

1. Montrer que $V_p(r) = r^p V_p(1)$ (sans calculer $V_p(1)$).
2. Soit $X \sim \mathcal{U}(B_p(1))$ de loi uniforme sur la boule unité, calculer, pour tout $r > 0$, $\mathbb{P}(\|X\| \leq r)$.
3. Soient $X_1, X_2 \sim \mathcal{U}(B_p(1))$ indépendantes.
   (a) Calculer $\mathbb{E}[\|X_1\|^2]$.
   (b) Vérifier que X_i est un vecteur aléatoire centré et en déduire la valeur de $\mathbb{E}[\|X_1 - X_2\|^2]$.
4. Mêmes questions pour $X_1, X_2 \sim \mathcal{U}([-1,1]^p)$.
5. Que pouvez-vous dire sur la loi uniforme en grande dimension pour les deux situations ? Étudier la loi limite et/ou l'espérance des normes des vecteurs.

Exercice 3 (Volume de la boule unité en dimension p)

On étudie en détail le volume de la boule unité en grande dimension.

1. Montrer que $V_1(1) = 2$ et $V_2(1) = \pi$.
2. Prouver que, pour $p \geq 3$, $V_p(1) = \frac{2\pi}{p} V_{p-2}(1)$.
3. En déduire que, pour tout $k \in \mathbb{N}\backslash\{0\}$,
$$V_{2k}(1) = \frac{\pi^k}{k!}$$
   et, pour tout $k \in \mathbb{N}$,
$$V_{2k+1}(1) = \frac{2^{2k+1} k! \pi^k}{(2k+1)!}.$$
4. Donner un équivalent de $V_p(r)$ quand $p \to \infty$. On rappelle la formule de Stirling
$$n! \sim_{n\to\infty} \sqrt{2\pi n} n^n e^{-n}.$$
5. Nous souhaitons couvrir l'hypercube $[0,1]^p$ par des boules de rayon 1 c'est-à-dire trouver $x_1, ..., x_n \in \mathbb{R}^p$ tels que
$$[0,1]^p \subset \bigcup_{i=1}^{n} B_p(x_i, 1). \tag{2}$$
   Notons n_p le nombre minimal de points nécéssaire pour couvrir $[0,1]^p$.
   (a) Montrer que $n_p = 1$ lorsque $p \leq 4$. Que se passe-t'il pour $p = 5$?
   (b) Montrer que si (2) est vérifiée alors $n_p V_p(1) \geq 1$.

Exercice 4 (Loi normale en grande dimension)

On s'intéresse à la loi normale centrée réduite en dimension p dont on rappelle la densité.

$$g_p(x) = \frac{1}{\sqrt{2\pi}^p} e^{-\|x\|^2/2}.$$

1. Quel est le maximum de g_p (le mode de la distribution) ? que se passe-t-il quand $p \to +\infty$?
2. Nous souhaitons déterminer la masse de la "cloche" de la gaussienne c'est-à-dire la partie de plus grande densité. Soit $\delta > 0$ proche de 0, nous notons
$$B_{p,\delta} := \{x \in \mathbb{R}^p, g_p(x) \geq \delta g_p(0)\}.$$
   (a) Faire un dessin en dimension $p = 1$.
   (b) Faire un dessin en dimension $p = 2$.
   (c) À l'aide de l'inégalité de Markov, montrer que
$$\mathbb{P}(X \in B_{p,\delta}) \leq \delta^{-1} 2^{-p/2}.$$
   (d) Conclure sur l'effet de la dimensions sur la loi normale centrée réduite.