Apprentissage Statistique

Exercice du diapo

S1-2023

1 Supervised classification

1.1 Excercice

Exercice 1 :

$$\begin{aligned}L^* &= 1 - P(g^*(X) = Y)\\&= 1 - \mathbb{E}[g^*(X) = 1]P(Y = 1|X)\\&= 1 - \mathbb{E}[\mathbb{1}_{P(Y=1|X)} + \mathbb{1}_{g^*(X)=0}P(Y = 0|X)]\\&= 1 - \mathbb{E}[\mathbb{1}_{\eta(X)>1/2}\eta(X) + \mathbb{1}_{\eta(X)\leq 1/2}(1 - \eta(X))]\end{aligned}$$

Exercice 2

$$\begin{aligned}L^* &= 1 - P(g^*(X) = Y)\\&= 1 - \mathbb{E}[P(g^*(X) = Y|X)]\\&= 1 - \mathbb{E}[\max(1 - \eta(X), \eta(X))]\\&= 1 + \mathbb{E}[\min(\eta(X) - 1, -\eta(X))]\\&= \mathbb{E}[\min(\eta(X), 1 - \eta(X))]\end{aligned}$$

Exercice 3: Si $L * (X) = 0$ ça veut dire que c'est un processus déterministe. Que Y a un lien déterministe avec X

$$P(g^*(X) \neq Y) = 0 \Rightarrow Y = g^*(X)as..$$

$$Y = \phi(X) \Rightarrow P(Y \neq \phi(X)) = 0 \Rightarrow L* = 0.$$

1.2 Statistical Learning

Exercice diapo 18 : $consistency \Leftrightarrow L(g_n) \to^{L^1} L* \Leftrightarrow L(g_n) \to^{\mathbb{P}} L*$

- Pour la convergence L1 (je crois)

$$\begin{aligned}&P(g_n(X) \neq Y|\mathcal{D}_n) \to^{L1} L*\\&\mathbb{E}[P(g_n(X) \neq Y|\mathcal{D}_n) - L*]\\&= \mathbb{E}[P(g_n(X) \neq Y[\mathcal{D}_n) - L*])]\\&= P(g_n(X) \neq Y) - L*\\&\to 0(?)\end{aligned}$$

- Pour la convergence en proba dans le sens non instinctif, on veut montrer que

$$\begin{aligned}&Z_n \to^{\mathbb{P}} 0\\&|Z_n| \leq 1 \text{ car proba}\\&\text{alors } Z_n \to^{L1} 0\end{aligned}$$

Preuve :

$$\begin{aligned}&\mathbb{E}[|Z_n|] = \mathbb{E}[Z_n|\mathbb{1}_{|Z_n|>\epsilon}] + \mathbb{E}[Z_n|\mathbb{1}_{|Z_n|\leq\epsilon}]\\&\leq P(|Z_n| > \epsilon)\end{aligned}$$

1.3 Exercice TD1

1.3.1 Exercice 1

$(X,Y) \in \mathbb{R} \times [0,1], X \sim \mathcal{U}([-2,2]), ect$

$$\begin{aligned} Y &= \mathbb{1}_{X<0,U\leq 2} + \mathbb{1}_{X>0,U>1} \\ \eta &= P(Y=1|X) \\ &= P(X<0, u\leq 2|X) + P(X>0, U>1|X) \\ &= \mathbb{1}_{X<0}P(U\leq 2) + \mathbb{1}_{X>0}P(U>1) \\ &= \mathbb{1}_{X<0}\frac{2}{10} + \mathbb{1}_{X>0}\frac{9}{10} \end{aligned}$$

$$g*(X) = \mathbb{1}_{\eta(X)>\frac{1}{2}} = \mathbb{1}_{X>0}$$

Bayes error

$$\begin{aligned} L* &= P(g*(X) \neq Y) \\ &= \frac{1}{2} - \frac{1}{2}\mathbb{E}[|2\eta(X)-1|] \\ &= \frac{1}{2} - \frac{1}{2}\mathbb{E}[\left|\frac{4}{10}\mathbb{1}_{X<0} + \frac{18}{10}\mathbb{1}_{X>0} - 1\right|] \\ &= \frac{1}{2} - \frac{1}{2}\mathbb{E}[\left|\frac{6}{10}\mathbb{1}_{X<0} + \frac{8}{10}\mathbb{1}_{X>0}\right|] \\ &= \frac{1}{2} - \frac{1}{2} * \frac{6}{10} * \frac{1}{2} - \frac{1}{2} * \frac{8}{10} * \frac{1}{2} \\ &= \frac{3}{20} \end{aligned}$$

1.3.2 Exercice 2

1.

$$\begin{aligned} L* &= \mathbb{E}[\min(\eta(X), 1-\eta(X))] \\ &= \mathbb{E}[\min(\frac{x}{c+x}, \frac{c}{c+x})] \\ &= \mathbb{E}[\frac{\min(x,c)}{c+x}] \end{aligned}$$

2. $f_x(x) = \frac{\mathbb{1}_{x\in[0,\alpha c]}}{\alpha c}$

$$\begin{aligned} L* &= \int_0^{\alpha c} \frac{\min(x,c)}{x+c}\frac{dx}{\alpha c} \\ &= \int_0^{c} \frac{x}{x+c}\frac{dx}{\alpha c} \\ &= \frac{1}{\alpha}\log(\frac{(\alpha+1)e}{4}) \end{aligned}$$

3. exercice, étudier la fonction ou
   - f continue sur $[1, +\infty]$
   - $\lim_{\alpha\to\infty} f(\alpha) = 0$

   Docn f admet un maximum

1.3.3 Exerice 3

$X \sim (T, B, E) \sim \mathcal{E}(1)$
densité d'une loi exp $Z \sim \mathcal{E}(1)$

$$\begin{aligned} f_Z(z) &= e^{-z\mathbb{1}_{\mathbb{R}_+}(z)} \\ F_Z(t) &= (1-e^{-t})\mathbb{1}_{\mathbb{R}_+}(t) \\ G_z(t) &= \mathbb{1}_{\mathbb{R}_-} + e^{-t}\mathbb{1}_{\mathbb{R}_+}(t) \end{aligned}$$

1. Y est une fonction de X donc $L* = 0$. C'est déterministe
2. $$\begin{aligned} P(Y=1|T,B) &= P(T+B+E<7|T,B) \\ &= F_Z(7-T-B) \text{ car } E\perp(T,B) \qquad = (1-e^{-(7-T-B)})\mathbb{1}_{7-T-B>0} \end{aligned}$$
3. $$\begin{aligned} &g^*(T,B) = \mathbb{1}_{\eta(T,B)>1/2} \\ &\eta(T,B) > 1/2 \\ \Leftrightarrow &1 - e^{T+B-7} > \frac{1}{2} \\ \Leftrightarrow &-\ln 2 > T+B-7 \\ \Leftrightarrow &T+B < 7-\ln 2 \end{aligned}$$
   Donc $g*(T,B) = \mathbb{1}_{T+B<7-\ln 2}$
4. $T, B \sim \mathcal{E}(1)$ et $T\perp B$ donc $T+B \sim \gamma(2,1)$, $f_{T+B}(u) = ue^{-u}\mathbb{1}_{\mathbb{R}_+(u)}$
5. $$\begin{aligned} L^* = P(g^*(T,B) \neq Y) &= P(g^*(T,B)=1, Y=0) + P(g^*(T,B)=1, Y=0) \\ &= P(T+B<7-\ln 2, T+B+E\geq 7) + P(T+B\geq 7-\ln 2, T+B+E<7) \\ &= a+b \\ a &= \mathbb{E}[P(T+B<7-\ln 2, T+B+3\geq 7|T,B)] \\ &= \mathbb{E}[\mathbb{1}_{T+B<7-\ln 2}G(7-T-B)] \\ &= \mathbb{E}[\mathbb{1}_{T+B<7-\ln 2}e^{T+B-7}] \\ &= \int_0^{7-\ln 2} e^{u-7}ue^{-u}du \\ &= e^{-7}[\frac{u^2}{2}]_0^{7-\ln 2} \\ &= e^{-7}\frac{(7-\ln 2)^2}{2} \\ b &= \text{ same} \\ a+b &= e^{-7}(\frac{(7-\ln 2)^2}{2} + 2(8-\ln 2) - 8 - \frac{7^2}{2} + \frac{(7-\ln 2)^2}{2}) \end{aligned}$$
6. $$\begin{aligned} P(Y=0) &= P(T+B+E\geq 7) \\ &= P(\gamma(3)\geq 7) \\ &= \int_7^{+\infty}\frac{1}{2}u^2e^{-u}du \\ &= 0.029 \end{aligned}$$

2 Exercice TD2

2.1 Exercice 1

1.
$$g^*(x) = \begin{cases} 1 & \text{si } \eta(x) > 1/2 \\ 0 & \text{sinon} \end{cases}.$$

En posant $G^* = \{x \in \mathbb{R}^d, \eta(x) > 1/2\}$, on a bien $g^*(x) = \mathbb{1}_{x \in G^*}$

2.
$$\begin{aligned} P(g(x) \neq Y|X) &= 1 - P(g(x) = Y|X) \\ &= 1 - P(g(X) = 1, Y = 1|X) - P(g(X) = 0, Y = 0|X) \\ &= 1 - \mathbb{1}_{g(X)=1} P(Y = 1|X) - \mathbb{1}_{g(X)=0} P(Y = 0|X) \\ &= 1 - \mathbb{1}_{g(X)=1} \eta(X) - \mathbb{1}_{g(X)=0}(1 - \eta(X)) \\ P(g^*(X) \neq Y|X) &= 1 - \mathbb{1}_{g^*(X)=1} \eta(X) - \mathbb{1}_{g^*(X)=0}(1 - \eta(X)) \end{aligned}$$

Assemblons les deux termes

$$\begin{aligned} &P(g(x) \neq Y|X) - P(g^*(X) \neq Y|X) \\ &= \mathbb{1}_{g(X)=1}\eta(X) - \mathbb{1}_{g(X)=0}(1 - \eta(X)) - \mathbb{1}_{g^*(X)=1}\eta(X) - \mathbb{1}_{g^*(X)=0}(1 - \eta(X)) \\ &= \mathbb{1}_{g^*(X)=1}\eta(X) + (1 - \mathbb{1}_{g^*(X)=1})(1 - \eta(X)) - \mathbb{1}_{g(X)=1}\eta(X) - (1\mathbb{1}_{g(X)=1})(1 - \eta(X)) \\ &= \eta(X)(\mathbb{1}_{g^*(X)} - \mathbb{1}_{g(X)=1}) + (1 - \eta(X))[1 - \mathbb{1}_{g^*(X)=1} - 1 + \mathbb{1}_{g(X)=1}] \\ &= [\mathbb{1}_{g^*=1} - \mathbb{1}_{g(X)=1}][\eta(X) - 1 + \eta(X)] \\ &= (2\eta(X) - 1)(\mathbb{1}_{g^*(X)=1} - \mathbb{1}_{g(X)=1}) \\ &= |2\eta(X) - 1| \, \mathbb{1}_{g(X) \neq g^*(X)} \text{ par définition de } g^* \end{aligned}$$

CCL : $P(g(X \neq Y) - P(g^*(X) \neq Y) = \mathbb{E}[|2\eta(X) - 1| \, \mathbb{1}_{g(x) \neq g^*(X)}]$

3. On pose $G = [g = 1], G^* = [n \geq 1/2] = [g^* = 1]$; $G \Delta G^* = ([g = Z] \cap [g^* =]) \cup ([g = 0] \cap [g^* = 1]) = [g \neq g^*]$

$$\begin{aligned} d(G, G^*) &= P(g(X) \neq Y) - L^* \\ &= \int_{G\Delta G^*} |2\eta(X) - 1| \, d\mu(x) \end{aligned}$$

4. Comme $0 \leq \eta(X) \leq 1$ on obtient que $0 \leq |2\eta(x) - 1| \leq 1$, donc

$$d(G, G^*) = \int_{G\Delta G^*} |2\eta(x) - 1| \, d\mu(x) \leq \mu(G\Delta G^*) = d_\delta(G, G^*) \leq 1.$$

5. $\forall t \in (0, t^*], (0 < t^* \leq 1/2)$

$$P(\left|\eta(X) - \frac{1}{2}\right| \leq t) \leq C_\eta t^\alpha, C_\eta > 0, \alpha > 0.$$

When $\alpha \to \infty, P(\left|\eta(X) - \frac{1}{2}\right| \leq t) \to 0$

6. $P(\left|\eta(X) - \frac{1}{2}\right| \leq t) \leq C_\eta t^\alpha$ pour tout $t \in]0, t^*]$ et $0 < t^* < \frac{1}{2}$.

$$\forall x \in V, \eta(x) = \begin{cases} \frac{1}{2} + x^{1/\alpha} & \text{si } x > 0 \\ \frac{1}{2} - x^{1/\alpha} & \end{cases}.$$

X a une densité

$$\begin{aligned} P(\left|\eta(X) - \frac{1}{2}\right| \leq t) &= P(|X|^{1/\alpha} \leq t) \\ &= \int_{[-t^\alpha, t^\alpha]} f_x(dx) \\ &\leq M2t^\alpha \end{aligned}$$

t dans le voisinage, suffisament petit.

7. Montrer que $(1) \Rightarrow d(G, G^*) \geq 2t[d_\Delta(G, G^*) - C_\eta t^\alpha]$

$$\begin{aligned}
d(G, G^*) &= \int_{G\Delta g^*} |2\eta(x) - 1| \, \mu(dx) \\
&\geq 2 \int_{G\Delta G^*} \left|\eta(x) - \frac{1}{2}\right| \mathbb{1}_{|\eta(x)-1/2|>t} \mu(dx) \\
&\geq 2t\mu(G\Delta G^* \cap \{\left|\eta(x) - \frac{1}{2}\right| > t\}) \\
&\geq 2t[\mu(G\Delta G^*) - \mu(\left|\eta(x) - \frac{1}{2}\right| \leq t)] \\
&\geq 2t[d_\Delta(G, G^*) - C_\eta t^\alpha]
\end{aligned}$$

On note : $\mu(A \cap B) \geq \mu(A) - \mu(B)$

8. On cherche $kappa, c_0, \epsilon_0$ tel que $d_\Delta(G, G^*) \leq \epsilon_0 \Rightarrow d(G, G^*) \geq c_0 d_\Delta^\kappa(G, G^*)$
Notons $\phi : 2td_\Delta - 2C_\eta t^{\alpha+1}$

$$\begin{aligned}
\phi(t) &= 2d_\delta - 2C_\eta(\alpha + 1)t^\alpha \\
&= 0 \\
&\Rightarrow t = \frac{d_\Delta}{C_\eta(\alpha + 1)} \\
&\Rightarrow t = (\frac{d_\Delta}{C_\eta(\alpha + 1)})^{1/\alpha}
\end{aligned}$$

$$\begin{aligned}
2(\frac{d_\Delta}{C_\eta(\alpha + 1)})^{1/\alpha}[d_\Delta - C_\eta \frac{d_\Delta}{C_\eta(\alpha + 1)}] &= 2(\frac{d_\Delta}{C_\eta(\alpha + 1)})^{1/\alpha}(1 - \frac{1}{\alpha + 1})d_\Delta \\
&= 2\frac{\alpha}{\alpha + 1} \frac{1}{(C_\eta(\alpha + 1))^{1/\alpha}} d_\Delta^{\alpha+1/\alpha} \\
&= \frac{2\alpha}{(\alpha + 1)^{\alpha+1/\alpha} C_\eta^{1/\alpha}}
\end{aligned}$$

donc

$$\begin{aligned}
\epsilon_0 &= t^* \alpha C_\eta(\alpha + 1) \\
\kappa &= \frac{\alpha + 1}{\alpha} \\
C_0 &= \frac{\alpha}{(\alpha + 1)^{\alpha+1/\alpha}} \frac{2}{C_\eta^{1/\alpha}}
\end{aligned}$$

9.

$$\begin{aligned}
d(G, G^*) &= \int_{G\Delta G^*} |2\eta(x) - 1| \, \mu(dx) \\
&= \int_{G\Delta G^*} |2\eta(x) - 1| \, (\mathbb{1}_{|\eta(x)-1/2|\leq t} + \mathbb{1}_{|\eta(x)-1/2|>}) \mu(dx) \\
&\leq 2tP(|\eta(X) - 1/2| \leq t) + \mathbb{E}[|2\eta(X) - 1| \, \mathbb{1}_{g(X)\neq g^*(X)} \mathbb{1}_{|\eta(X)-1/2|>t}] \\
&\leq 2C_\eta t^{\alpha+1} + \mathbb{E}[|2\eta(X) - 1| \, \mathbb{1}_{g(X)\neq g^*(X)} \mathbb{1}_{|\eta(X)-1/2|>t}]
\end{aligned}$$