Correction des travaux dirigés de Machine learning

Cycle pluridisciplinaire d'études supérieures
Université Paris sciences et lettres

Joon Kwon

vendredi 20 mars 2020

Exercice 1. — On rappelle que pour $f \in \mathscr{F}(\mathscr{X}, \mathscr{Y})$:

$$R(f) = \mathbb{E}\left[\ell(Y, f(X))\right] = \mathbb{E}\left[\mathbb{1}_{Y \neq f(X)}\right],$$

où la second égalité est valable ici car on considère la fonction de perte 0–1.

1) On pose $f_*(x) = \begin{cases} 1 & \text{si } x \geqslant 1/2 \\ 0 & \text{sinon.} \end{cases}$. Alors étant donné la distribution P, on a toujours $f_*(X) = Y$. Montrons que f_* minimise bien le risque. On calcule son risque :

$$R(f_*) = \mathbb{E}[\underbrace{\mathbb{1}_{\{Y \neq f(X)\}}}_{=0}] = 0.$$

Pour tout prédicteur $f \in \mathscr{F}(\mathscr{X}, \mathscr{Y})$,

$$R(f) = \mathbb{E}[\underbrace{\mathbb{1}_{\{Y \neq f(X)\}}}_{=0 \text{ ou } 1}] \geqslant 0 = R(f_*).$$

Donc, f_* minimise bien le risque, autrement dit :

$$R(f_*) = \min_{f \in \mathscr{F}(\mathscr{X}, \mathscr{Y})} R(f).$$

2) On pose $f_*(x) = 1$ pour tout $x \in \mathscr{X}$ et montrons que f_* minimise le risque. On a,

$$\begin{aligned} R(f_*) &= \mathbb{E}\left[\mathbb{1}_{\{Y \neq f(X)\}}\right] = \mathbb{E}\left[\mathbb{1}_{\{Y \neq 1\}}\right] \\ &= \mathbb{P}[Y \neq 1] = 1 - \mathbb{P}[Y = 0] \\ &= \frac{1}{3} \end{aligned}$$

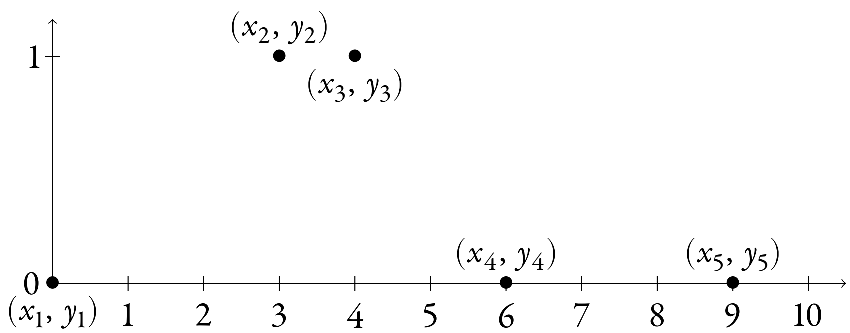
Pour $f \in \mathscr{F}(\mathscr{X}, \mathscr{Y})$,

$$\begin{aligned} R(f) &= \mathbb{E}\left[\mathbb{1}_{\{Y \neq f(X)\}}\right] = \mathbb{P}[Y \neq f(X)] \\ &= \mathbb{P}[Y \neq f(X) \text{ et } f(X) = 1] + \mathbb{P}[Y \neq f(X) \text{ et } f(X) = 0] \\ &= \mathbb{P}[Y = 0 \text{ et } f(X) = 1] + \mathbb{P}[Y = 1 \text{ et } f(X) = 0] \\ &= \mathbb{P}[Y = 0]\,\mathbb{P}[f(X) = 1] + \mathbb{P}[Y = 1]\,\mathbb{P}[f(X) = 0] \\ &\geqslant \frac{1}{3}\left(\mathbb{P}[f(X) = 1] + \mathbb{P}[f(X) = 0]\right) \\ &= \frac{1}{3}. \end{aligned}$$

où on a pu passer à la quatrième ligne car X et Y sont indépendants. On a donc bien

$$R(f_*) = \min_{f \in \mathscr{F}(\mathscr{X}, \mathscr{Y})} R(f).$$

Exercice 2. — On représente graphiquement l'échantillon d'apprentissage.



Pour $k = 1$, le découpage de l'ensemble d'entrées est donné par :

$$[0, 10] = \underbrace{A^{(1)}_{\{1\}}}_{=[0,3/2]} \sqcup \underbrace{A^{(1)}_{\{2\}}}_{=]3/2,7/2]} \sqcup \underbrace{A^{(1)}_{\{3\}}}_{=]7/2,5]} \sqcup \underbrace{A^{(1)}_{\{4\}}}_{=]5,15/2]} \sqcup \underbrace{A^{(1)}_{\{5\}}}_{=]15/2,10]},$$

où on rappelle que chaque $A^{(1)}_{\{m\}}$ pour $m \in [5]$ correspond à l'ensemble des $x \in [0, 10]$ pour lesquels x_m est le plus proche voisin (au sens de la première statistique de rang). Pour tout $m \in [5]$ et $x \in A^{(1)}_{\{m\}}$, on a $\hat{f}^{(1)}(x) = y_m$, ce qui donne après simplification :

$$\hat{f}^{(1)}(x) = \begin{cases} 0 & \text{si } x \leqslant 3/2 \\ 1 & \text{si } 3/2 < x \leqslant 5 \\ 0 & \text{si } x > 5, \end{cases}$$

On procède de même pour les autres prédicteurs.

$$\hat{f}^{(2)}(x) = \begin{cases} 0 & \text{si } x \leqslant 2 \\ 1 & \text{si } 2 < x \leqslant 9/2 \\ 0 & \text{si } x > 9/2, \end{cases}$$

$$\hat{f}^{(5)}(x) = 0, \quad \forall x \in \mathscr{X}.$$

Le calcul des erreurs d'apprentissage donne :

$$\varepsilon_{\text{train}}(\hat{f}^{(1)}) = \varepsilon_{\text{train}}(\hat{f}^{(2)}) = 0 \qquad \varepsilon_{\text{train}}(\hat{f}^{(5)}) = \frac{2}{5}.$$