FTML Exercices 3 solutions

Pour le 21 mars 2025

TABLE DES MATIÈRES

1 CONVEXITÉS

2)

— $x \mapsto x$ est convexe car

$$\forall x, y \in \mathbb{R}, \alpha \in [0, 1], \alpha x + (1-\alpha)y \leqslant \alpha x + (1-\alpha)y \tag{1}$$

— $x \mapsto 7x + 2$ est convexe car $\forall x, y \in \mathbb{R}, \alpha \in [0, 1]$,

$$\begin{aligned} 7(\alpha x + (1-\alpha)y) + 2 &= 7(\alpha x + (1-\alpha)y) + (\alpha + (1-\alpha))2 \\ &= \alpha(7x+2) + (1-\alpha)(7y+2) \end{aligned} \tag{2}$$

Plus généralement, toute application linéaire à valeurs dans $\mathbb{R}$ (forme linéaire) est convexe.

— $x \mapsto \sin(x)$ n'est pas convexe. En effet, prenons $x = 0$, $y = \pi$, $\alpha = 1/2$. Alors

$$\alpha x + (1-\alpha)y = \frac{\pi}{2} \tag{3}$$

Donc

$$\sin(\alpha x + (1-\alpha)y) = 1 \tag{4}$$

Or

$$\alpha \sin(x) + (1-\alpha)\sin(y) = 0 \tag{5}$$

Ainsi,

$$\sin(\alpha x + (1-\alpha)y) > \alpha \sin(x) + (1-\alpha)\sin(y) \tag{6}$$

— $x \mapsto x^2$ est convexe car

$$\begin{aligned} (\alpha x + (1-\alpha)y)^2 &= \alpha^2x^2 + 2\alpha(1-\alpha)xy + (1-\alpha)^2y^2 \\ &= \alpha^2x^2 + 2\alpha(1-\alpha)xy + y^2 - 2\alpha y^2 + \alpha^2y^2 \end{aligned} \tag{7}$$

On veut montrer que 7 est inférieur à $\alpha x^2 + (1-\alpha)y^2$.
C'est-à-dire :

$$\alpha^2x^2 + 2\alpha(1-\alpha)xy + y^2 - 2\alpha y^2 + \alpha^2y^2 \leqslant \alpha x^2 + (1-\alpha)y^2 \tag{8}$$

Ce qui équivaut à

$$\alpha x^2 + (1-\alpha)y^2 - (\alpha^2x^2 + 2\alpha(1-\alpha)xy + y^2 - 2\alpha y^2 + \alpha^2y^2) \geqslant 0 \tag{9}$$

Or

$$\begin{aligned}
\alpha x^2 + (1-\alpha)y^2 - (\alpha^2x^2 + 2\alpha(1-\alpha)xy + y^2 - 2\alpha y^2 + \alpha^2y^2) \\
= (\alpha-\alpha^2)x^2 - \alpha y^2 - 2\alpha(1-\alpha)xy + 2\alpha y^2 - \alpha^2y^2 \\
= (\alpha-\alpha^2)x^2 + \alpha y^2 - 2\alpha(1-\alpha)xy - \alpha^2y^2 \\
= (\alpha-\alpha^2)x^2 + (\alpha-\alpha^2)y^2 - 2\alpha(1-\alpha)xy \\
= (\alpha-\alpha^2)(x^2 - 2xy + y^2) \\
= (\alpha-\alpha^2)(x-y)^2 \\
\geqslant 0
\end{aligned} \tag{10}$$

On remarque que l'inégalité est stricte dès que $x \neq y$. On dit que $x \mapsto x^2$ est strictement convexe.

— $x \mapsto |x|$ par inégalité triangulaire. Plus généralement, tout norme est convexe pour la même raison.

3) Supposons qu'une fonction f soit convexe et concave. Par définition on aurait donc

$$\forall x, y \in \mathbb{R}, \forall \alpha \in [0,1], f(\alpha x + (1-\alpha)y) = \alpha f(x) + (1-\alpha)f(y) \tag{11}$$

On peut montrer que cela implique que f est affine.
Réciproquement, une application affine est bien convexe et concave.
Remarque : Si on suppose que f est dérivable, la démonstration est rapide : en effet on sait que la dérivée est croissante et décroissante, elle est donc constante, égale par exemple à $a \in \mathbb{R}$. Par intégration, on sait que

$$\forall x \in \mathbb{R}, f(x) = f(0) + \int_0^x f'(t)dt = f(0) + ax \tag{12}$$

Ce qui montre que f est affine.

4) Oui, on peut le voir en appliquant la convexité aux 2 fonctions dont on fait la somme.

2 RISQUES DE BAYES

2.1 Exemple 1 (penalty)

On utilise la 01 loss, donc on sait que le prédicteur de Bayes prédit, pour une valeur x de X, la sortie la plus probable. Il prédit donc 1 si $X = 1$, et 0 si $X = 0$.

Pour calculer le risque de Bayes, c'est le même calcul que l'exercice 1 du cours magistral 3 avec l'estimateur f_1 et

— $p = 0.6$
— $q = 1 - p$

Ainsi :

$$\begin{aligned} R(f_1) &= E[l(Y, f(X))] \\ &= 1 \times P(Y \neq f(X)) + 0 \times P(Y = f(X)) \\ &= P(Y \neq f(X)) \\ &= P((Y \neq f(X)) \cap (X = 1)) + P((Y \neq f(X)) \cap (X = 0)) \\ &= P((Y \neq f(X))|X = 1)P(X = 1) \\ &+ P((Y \neq f(X))|X = 0)P(X = 0) \\ &= \frac{1}{2}P((Y = 0)|X = 1) + \frac{1}{2}P((Y = 1)|X = 0) \\ &= \frac{1}{2}(1 - p) + \frac{1}{2}q \\ &= 0.4 \end{aligned} \tag{13}$$

2.2 Exempler 2 : Spotify streams

On utilise la squared loss, donc on sait que le prédicteur de Bayes prédit, pour une valeur x de X, l'espérance conditionnelle de Y sachant que $X = x$. Pour une valeur x donnée, Y suit une loi binomiale de paramètres $n_Y(x) = 3^x$ et $p_Y(x) = 0.5$. L'espérance de cette variable conditionnelle est $n_Y(x)p_Y(x) = 0.5n_Y(x)$, qui est donc la prédiction faite par le prédicteur de Bayes. Calculons maintenant le risque de Bayes. Pour cela, on utilise à nouveau la loi de l'espérance totale pour calculer dans un premier temps le "risque conditionnel" : l'espérance de la loss sachant X. Etant donnée une valeur x de X, on remarque que l'espérance de l'erreur est

$$\begin{aligned} E[l(Y, f^*(X))|X = x] &= E[(Y - f^*(X))^2|X = x] \\ &= E[(Y - E[Y|X])^2|X = x] \\ &= Var[Y|X = x] \\ &= n_Y(x)p_Y(x)(1 - p_Y(x)) \\ &= 3^x/4 \end{aligned} \tag{14}$$

Pour calculer le risque de Bayes, il nous suffit, avec le théorème de l'espérance totale, de calculer l'espérance de $E[l(Y, f^*(X))|X = x]$. Mais on connaît la loi de X qui est discrète et suit une loi binomiale de paramètres $n_X = 20$ et $p_X = 0.2$. Ainsi, ce calcul d'espérance est une simple somme.

$$\begin{aligned} R^* &= \sum_{x=0}^{n_X} P(X = x)E[l(Y, f^*(X))|X = x] \\ &= \sum_{x=0}^{n_X} \binom{n_X}{x} p_X^x (1 - p_X)^{n_X - x} n_Y(x)/4 \\ &= \sum_{x=0}^{20} \binom{20}{x} (0.2)^x (0.8)^{20-x} 3^x/4 \end{aligned} \tag{15}$$

Et on calcule cette somme numériquement pour arriver au résultat (c'est ce que fait la fonction **compute_bayes_risk()** dans la solution du tp).

Remarque : le fait que le risque de Bayes soit l'espérance de la variance conditionnelle n'est pas spécifique à ce problème. Ce sera vrai dès qu'on utilise la squared loss.