

Master de Mathématiques et Applications- UL FDS
Rattrapage : Méthodes Mathématiques pour Machine Learning 2024. Durée 3h

NB les notations sont celles du cours. Le prêt de tout objet est strictement interdit.

Problème 1

A) Régression linéaire simple

On a relevé sur un échantillon d'individus les observations d'un couple de caractères (X, Y) données ci-dessous.

x_i	249	255	260	267	288	301	320	334	350	364	382	399	413	432	450
y_i	160	165	170	174	182	193	207	220	240	243	253	262	273	277	290

- 1) Coefficient de corrélation linéaire entre X et Y ?
- 2) Droite de régression de Y en X ? et la valeur ajustée y_a si $x = 300$?
- 3) Soit M_i le point du plan $\mathbb{R}^2$ de coordonnées (x_i, y_i) et $\Delta : y = \alpha + \beta x$ une droite du plan. Soit O l'origine du repère de $\mathbb{R}^2$ et P_i la projection orthogonale de M_i sur Δ . On note par $d(M, P)$ la distance euclidienne entre deux points M et P et on souhaite déterminer la droite Δ qui rend minimum la somme

$$S = \sum_{i=1}^n d^2(M_i, P_i).$$

a) Démontrer que

$$\min \sum_{i=1}^n \|M_i P_i\|^2 \iff \max_{u'u=1} u'Vu$$

où u est un vecteur directeur de la droite Δ et V est la matrice var-covar de X et Y .

- b) En déduire que u est le vecteur propre associé à la plus grande valeur de V .
- c) A.N. :

c1) Donner la matrice V .

c2) Montrer que

$$\lambda_{\max} = s_x^2 + s_y^2 + \sqrt{(s_x^2 - s_y^2)^2 + 4s_{xy}^2}$$

où s_x^2 , s_y^2 et s_{xy} sont les variances et covariance de X et Y et trouver u . En déduire l'équation de Δ dans le cas des observations ci-dessus.

B) Régression linéaire multiple

On considère 5 observations de X_1 , X_2 , Y données par

X_1	1	2	3	1	2	3	1	2	3
X_2	6	5	4	4	3	2	2	1	0
Y	1	1	1	2	2	2	3	3	3

- a) Rappeler la formule donnant l'estimateur $\hat{\beta}$ de β du modèle linéaire multiple

$$Y = X\beta + \varepsilon$$

où X , Y , β et ε sont à préciser.

b) Rappeler la formule de l'erreur empirique MSE du modèle ajusté $\hat{Y} = X\hat{\beta}$ avec la fonction de coût $l(y, y') = (y - y')^2$.

c) Rappeler la formule de l'estimateur sans biais S^2 de la variance σ^2 de ε .

d) A.N. : En utilisant les observations ci-dessus, donner l'expression du modèle d'apprentissage à l'aide de l'estimateur $\hat{\beta}$ et les valeurs de MSE et de S^2 .

C) Régression linéaire régularisée

1) Rappeler la formule de la régularisation avec la fonction de coût précédente. On pourra utiliser la fonction Argmin ou Argmax.

2) En déduire celles de la régression Ridge, Lasso et Elastic Net.

Problème 2

A) Réseaux de neurones

1) On donne $x_1 = (0, 0, 1, 1)$ et $x_2 = (0, 1, 0, 1)$.

1) Calculer $\hat{y} = f(x_1, x_2)$ et le comparer à $y = (0, 0, 0, 1)$ à l'aide d'un perceptron avec les poids $w_0 = -1.5$, $w_1 = w_2 = 1$ et la fonction d'activation seuil définie par $\sigma(x) = 1$ si $x > 0$, $\sigma(x) = 0$ si $x < 0$.

2) Calculer $\hat{y} = f(x_1, x_2)$ et le comparer à $y = (0, 1, 1, 0)$ à l'aide d'un perceptron avec une couche cachée à deux neurones avec les poids $w_{01}^1 = -0.5$, $w_{02}^1 = -0.5$, $w_{11}^1 = 1$, $w_{12}^1 = -1$, $w_{21}^1 = -1$, $w_{22}^1 = 1$, $w_0^2 = -0.5$, $w_1^2 = 1$, $w_2^2 = 1$ et la fonction d'activation Seuil précédemment définie.

3) On considère un réseau de neurones décrit par le graphe ci-dessous.

a) Donner les expressions de a_1^1 , a_2^1 , a_2 et $\hat{y}$.

b) On prend pour fonction d'activation la sigmoïde. Donner la valeur de $\hat{y}$ pour $x = 2$, $w_1 = 0.5$, $w_2 = -1$, $w_3 = -2$, $w_4 = 1.5$, $w_5 = 1$, $w_6 = 1$, $w_7 = -1$. L'entrée de w_5 est 1.

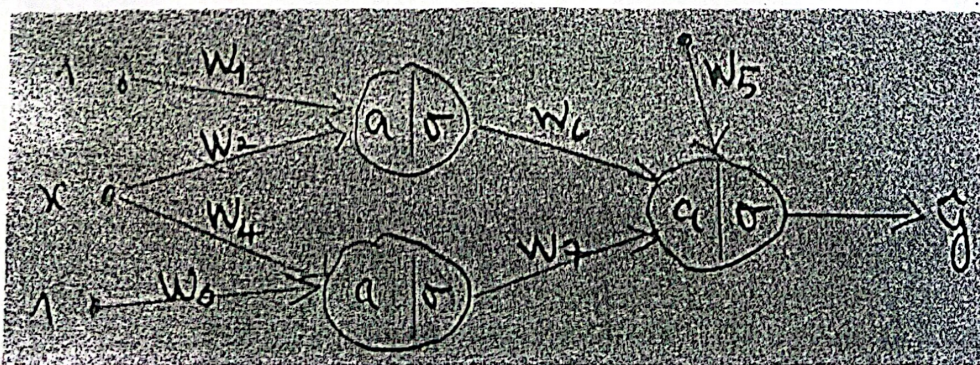


FIGURE 1 – Graphe du réseau

B) Support vector machine

Soit $\mathcal{D} = \{(x_i, y_i), i = 1 \dots n\}$ un jeu de données avec $x_i \in \mathbb{R}^p$, $y_i \in \{-1, 1\}$. Soit

$$H : \langle w, x \rangle + b = 0$$

l'hyperplan séparateur optimal de $\mathcal{D}$.

- 1) Rappeler les formulations primale et duale de la SVM à marge rigide de $\mathcal{D}$ et leurs fonctions de décision.
- 2) Rappeler le problème d'optimisation de la SVM à noyau de $\mathcal{D}$. Donner un exemple simple de noyau.