

Correction : Série de Révision et Exercices de Travaux Dirigés (TD)

Correction : Série de QCM

1. Réponse correcte : A) Apprentissage supervisé
2. Réponse correcte : B) K-means
3. Réponse correcte : C) Maximiser une récompense cumulative
4. Réponse correcte : B) SVM (Machines à vecteurs de support)
5. Réponse correcte : C) Identifier des schémas cachés ou des regroupements
6. Réponse correcte : C) La valeur de qualité d'une action
7. Réponse correcte : A) La règle de probabilité bayésienne
8. Réponse correcte : B) Régression linéaire
9. Réponse correcte : B) À regrouper des données similaires
10. Réponse correcte : C) La conduite autonome

Correction : Exercice 1 - Naive Bayes

Type d'apprentissage : Apprentissage supervisé, car les données sont étiquetées (Spam ou Non-Spam).

Type de classification : Classification binaire (Spam/Non-Spam).

Calcul avec Naive Bayes :

$$- P(\text{Spam}) = 3/6, P(\text{Non-Spam}) = 3/6$$

$$- P(\text{'Offre'}|\text{Spam}) = 3/3, P(\text{'Gratuit'}|\text{Spam}) = 2/3$$

$$- P(\text{'Offre'}|\text{Non-Spam}) = 0, P(\text{'Gratuit'}|\text{Non-Spam}) = 2/3$$

Probabilité que l'email soit Spam : $P(\text{Spam}|\text{'Offre'}, \text{'Gratuit'}}) = (P(\text{'Offre'}|\text{Spam}) * P(\text{'Gratuit'}|\text{Spam}) * P(\text{Spam})) / P(\text{'Offre'}, \text{'Gratuit'})$

$$\Rightarrow (2/3 * 2/3 * 1/2) / P(\text{'Offre'}, \text{'Gratuit'})$$

Effectuez les calculs pour finaliser.

Correction : Exercice 2 - k-Nearest Neighbors (kNN)

Type de problème : Classification supervisée.

Distances euclidiennes calculées :

- Avec (65, 5.5) et (60, 5) : $\sqrt{((65-60)^2 + (5.5-5)^2)} = \sqrt{(25 + 0.25)} = \sqrt{25.25}$

- Avec (65, 5.5) et (70, 6) : $\sqrt{((65-70)^2 + (5.5-6)^2)} = \sqrt{(25 + 0.25)} = \sqrt{25.25}$

- Autres calculs à effectuer.

Prédiction : Les 3 voisins les plus proches sont... (complétez avec les distances exactes).

Correction : Exercice 3 - Régression linéaire

Équation de la régression linéaire : $y = ax + b$

- Calcul des coefficients :

$$a = \text{Cov}(X, Y) / \text{Var}(X)$$

$$b = \text{Moyenne}(Y) - a * \text{Moyenne}(X)$$

Pour les données : Tasses de café (X) et Température (Y), effectuez les calculs pour trouver a et b.

Prédiction pour $X = 7$: $y = a*7 + b$.

Correction : Exercice 4 – K-means

Client	Montant Total Dépensé (x)	Fréquence d'Achats (y)
Ali	500	2
Badr	100	1
Chiheb	800	4
Dhaker	200	1
Emir	1200	5
Fedi	350	2
Ghalia	600	3
Helmi	900	4

2) Centres initiaux et K=3

Pour K=3, nous allons choisir les trois premiers clients comme centroïdes initiaux :

- $C1=(500,2)$ (Ali)
 - $C2=(100,1)$ (Badr)
 - $C3=(800,4)$ (Chiheb)
-

3) Calcul des distances et attribution des clusters

Pour chaque client, nous allons calculer la distance euclidienne à chaque centroïde et attribuer chaque client au cluster le plus proche.

Formule de la distance euclidienne :

$$d = \sqrt{(x_{\text{client}} - x_{\text{centre}})^2 + (y_{\text{client}} - y_{\text{centre}})^2}$$

$$d = \sqrt{(x_{\text{client}} - x_{\text{centre}})^2 + (y_{\text{client}} - y_{\text{centre}})^2}$$

$$d = \sqrt{(x_{\text{client}} - x_{\text{centre}})^2 + (y_{\text{client}} - y_{\text{centre}})^2}$$

Calculons les distances pour chaque client :

Client	d(C1)	d(C2)	d(C3)	Cluster assigné
Ali	0	403.12	316.23	C1
Badr	403.12	0	728.01	C2
Chiheb	316.23	728.01	0	C3
Dhaker	316.23	100.00	640.31	C2
Emir	707.11	1105.45	412.31	C3
Fedi	158.11	250.00	452.77	C1
Ghalia	100.00	509.90	223.61	C1
Helmi	412.31	806.23	100.00	C3

4) Mise à jour des centres des clusters

Recalculons les centres des clusters en prenant la moyenne des coordonnées (x, y) des clients de chaque cluster.

- **Cluster C1** (Ali, Fedi, Ghalia) :

$$C1 = ((500+350+600)/3, (2+2+3)/3)$$

$$= (483.33, 2.33)$$

- **Cluster C2** (Badr, Dhaker) :

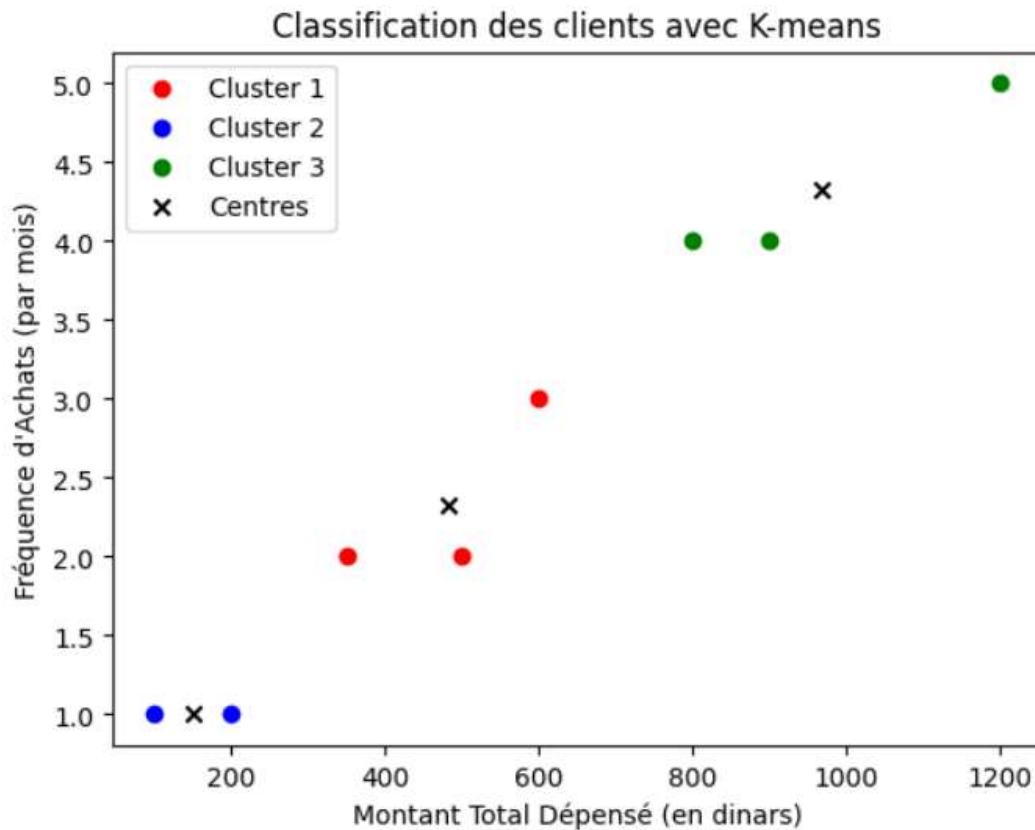
$$C2 = ((100+200)/2, (1+1)/2)$$

$$= (150, 1)$$

- **Cluster C3** (Chiheb, Emir, Helmi) :

$$C3 = ((800+1200+900)/3, (4+5+4)/3) \\ = (966.67, 4.33)$$

5) Représentation graphique des clusters



6) Nouveau client : (1200,6)

Calculons la distance du nouveau client aux centres finaux :

- Distance à C1=(483.33,2.33) :
 $d = \text{Racine}((1200 - 483.33)^2 + (6 - 2.33)^2) = 717.74$
- Distance à C2=(150,1) :
 $d = \text{Racine}((1200 - 150)^2 + (6 - 1)^2) = 1050.01$
- Distance à C3=(966.67,4.33)
 $d = \text{Racine}((1200 - 966.67)^2 + (6 - 4.33)^2) = 233.39$

Le nouveau client appartient au **Cluster C3C_3C3 : Clients avec des dépenses élevées et une fréquence d'achat élevée.**