

## Série de Révision et Exercices de Travaux Dirigés (TD)

### QCM

1. Quel type d'apprentissage utilise des données étiquetées pour entraîner un modèle ?
  - A) Apprentissage supervisé
  - B) Apprentissage non supervisé
  - C) Apprentissage par renforcement
  - D) Aucun des trois
2. Quel algorithme appartient à l'apprentissage non supervisé ?
  - A) Arbres de décision
  - B) K-means
  - C) Régression linéaire
  - D) Q-learning
3. Dans l'apprentissage par renforcement, l'objectif est de :
  - A) Classer des données en catégories
  - B) Trouver des structures cachées dans les données
  - C) Maximiser une récompense cumulative
  - D) Minimiser une erreur quadratique
4. Lequel des concepts suivants est un exemple de modèle d'apprentissage supervisé ?
  - A) Algorithme de clustering
  - B) SVM (Machines à vecteurs de support)
  - C) PCA (Analyse en composantes principales)
  - D) Réseau auto-encodeur
5. Dans l'apprentissage non supervisé, l'objectif est de :
  - A) Minimiser la perte sur des données étiquetées
  - B) Maximiser la récompense en explorant
  - C) Identifier des schémas cachés ou des regroupements
  - D) Générer de nouvelles données
6. Dans l'algorithme Q-learning, Q représente :
  - A) Une mesure de similarité
  - B) La fonction de récompense
  - C) La valeur de qualité d'une action
  - D) La probabilité conditionnelle

7. L'algorithme Naive Bayes est basé sur :
- A) La règle de probabilité bayésienne
  - B) Les distances euclidiennes
  - C) La descente de gradient
  - D) La propagation en arrière
8. Quel type d'algorithme utilise une fonction de perte pour ajuster ses paramètres ?
- A) K-means
  - B) Régression linéaire
  - C) Apprentissage par renforcement
  - D) Aucune des réponses ci-dessus
9. Dans un problème de clustering, nous cherchons principalement :
- A) À prédire une variable cible
  - B) À regrouper des données similaires
  - C) À classer des données étiquetées
  - D) À minimiser une récompense
10. Un exemple d'application de l'apprentissage par renforcement est :
- A) La reconnaissance faciale
  - B) Le clustering d'images
  - C) La conduite autonome
  - D) La régression polynomiale

### Exercice 1 : Naive Bayes

Une entreprise souhaite prédire si un email est spam ou non-spam. Vous avez les données suivantes :

Email	Contient "Offre"	Contient "Gratuit"	Spam/Non-Spam
1	Oui	Non	Spam
2	Oui	Oui	Spam
3	Non	Oui	Non-Spam
4	Non	Non	Non-Spam
5	Oui	Oui	Spam
6	Non	Oui	Non-Spam

1. Dans ce problème quel est le type d'apprentissage.
2. Expliquer le type de classification.
3. En utilisant l'algorithme Naive Bayes pour calculer la probabilité qu'un email contenant les mots "Offre" et "Gratuit" soit spam.

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

$$P(x_1, \dots, x_k | C) = P(x_1 | C) \cdot \dots \cdot P(x_k | C)$$

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

Une base de données contient les informations suivantes sur les plantes :

Hauteur (cm)	Largeur des feuilles (cm)	Type de plante
60	5	Plante A
70	6	Plante A
40	3	Plante B
50	4	Plante B
65	5.5	Inconnu (à prédire)

1. Expliquer le type de problème.
2. En utilisant la distance euclidienne ci-dessous, nous voulons prévoir le type de plante pour une nouvelle donnée (Hauteur = 65, Largeur des feuilles = 5.5) en utilisant **k = 3**.

$$d_{(x,y)} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

## Exercice 3 : Régression linéaire

Un scientifique collecte les données suivantes pour prédire la température en fonction du nombre de tasses de café consommées :

Tasses de café	Température (°C)
1	20
2	22
3	23
4	26
5	28
6	30

1. Ajustez un modèle de régression linéaire à ces données et prédisez la température **pour 7 tasses de café**.

$$b_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$Y = b_0 + b_1 X + e$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

### Exercice 3 : (K-moyenne)

Dans une société de sondage qui fait une enquête sur les clients des produits informatiques, un agent responsable a utilisé l'ensemble de données illustré dans le tableau ci-dessous représentant le montant total dépensé (en dinar) et la fréquence des achats mensuels pour chaque client.

A l'aide de l'algorithme K-moyenne, cet agent veut classer ces clients.

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

- 1) K-moyenne est un algorithme permettant un apprentissage supervisé, non supervisé ou renforcé ?
- 2) Expliquer brièvement le principe de l'algorithme K-moyenne.
- 3) En fixant  $K = 3$  et en choisissant les trois centres initiaux de nos clusters les trois premiers clients : (Ali (500,2), Badr (100,1) et Chiheb (800,4)) et en utilisant la formule de distance euclidienne calculée comme suit :

$$d(x,y)=\sqrt{\sum_{i=1}^n (x_i-y_i)^2}$$

Faire les calculs nécessaires pour obtenir les centres finaux des clusters.  
(Détaillez les étapes de calcul).

- 4) Représenter graphiquement les données tout en plaçant le montant total dépensé (sur l'axe des abscisses) en fonction des fréquences d'achat (sur l'axe des ordonnées) et en spécifiant les clusters.

- 5) Ajouter ces différents titres à vos clusters résultants :

- Clients avec des dépenses élevées et une fréquence d'achat élevée
- Clients avec des dépenses moyennes et une fréquence d'achat modérée.
- Clients avec des dépenses faibles et une fréquence d'achat faible.

- 6) Supposons que vous ajoutez un nouveau client avec un total de dépense est 1200 dinars et une fréquence d'achat est de 6. D'après votre représentation graphique, à quel cluster va-t-il appartenir ?