

M2_MIAGE_2

1. [Clustering dans un graphe] On peut construire un graphe de voisinage en...

- A** Fixant $\epsilon > 0$ puis en cherchant tous les points à distance inférieure à ϵ de chacun des x_i .
- B** Fixant $\epsilon > 0$ puis en cherchant tous les points à distance supérieure à ϵ de chacun des x_i .
- C** Fixant K entier > 0 puis en cherchant les K points les plus proches de chacun des x_i .
- D** Fixant K entier > 0 puis en cherchant les K points les plus éloignés de chacun des x_i .

2. [Clustering dans un graphe] De quelles distances avons-nous parlé dans le cours ?

- A** Distance euclidienne.
- B** Distance de graphe (plus court chemin dans le graphe de voisinage).
- C** Distance "temps moyen d'aller-retour".
- D** Distance d'édition entre graphes.

3. [Clustering dans un graphe] Si l'on retire k arêtes à un arbre couvrant de poids minimal, il reste...

- A** $k+1$ composantes connexes.
- B** k composantes connexes.
- C** $k-1$ composantes connexes.

4. [Chaînes de Markov] On considère un pigeon "markovien" n'évoluant que dans 3 états : Dormir, Voler, Chanter.

Quand il dort, il se réveille puis vole avec probabilité 0.1, chante avec probabilité 0.3.
Quand il vole, il se se pose puis s'endort avec probabilité 0.1, chante avec probabilité 0.4.
Quand il chante, il s'endort avec probabilité 0.3, et s'envole avec probabilité 0.2.

Quelle matrice correspond aux valeurs ci-dessus ? (Ordre : D, V, C).

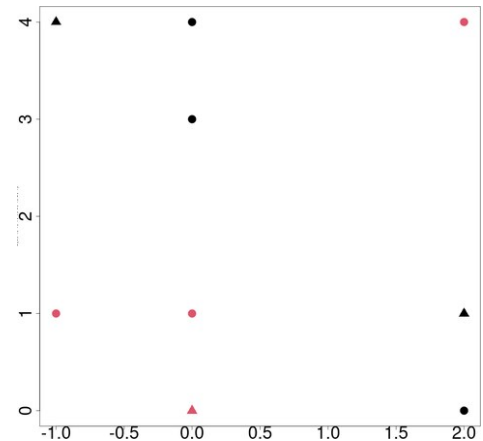
- (A) (0.5 0.1 0.3)
(0.1 0.5 0.4)
(0.3 0.2 0.6)
- (B) (0.1 0.1 0.3)
(0.1 0.1 0.4)
(0.3 0.2 0.1)
- (C) (0.6 0.1 0.3)
(0.1 0.5 0.4)
(0.3 0.2 0.5)

5. [Chaînes de Markov] Une matrice et un graphe de transition sont deux manières équivalentes de définir une chaîne de Markov.

- (T) True
(F) False

6. [k plus proches voisins] Combien d'erreurs réalisera-t-on en classant les triangles (données test) par un vote majoritaire parmi les 3 plus proches voisins ? (Cercles = données d'entraînement).

- (A) 0
(B) 1
(C) 2
(D) 3



7. [k plus proches voisins] Cochez les bonnes stratégies d'optimisation d'un modèle (= essayant de minimiser l'erreur future sur de nouvelles données).

- (A) Entraîner le modèle sur l'ensemble des données, puis évaluer l'erreur sur ce même ensemble.
- (B) Entraîner le modèle sur un sous-ensemble fixé (choisi aléatoirement), puis évaluer l'erreur sur son complémentaire.
- (C) Partitionner le jeu de données D en sous-ensembles D_1, \dots, D_k , entraîner le modèle successivement sur $D \setminus D_i$ et tester sur D_i : effectuer alors la moyenne des erreurs.
- (D) Répéter une dizaine de fois ou plus : choisir un sous-ensemble D' au hasard, entraîner le modèle sur $D \setminus D'$ puis calculer l'erreur sur D' . Retourner la moyenne des erreurs.

- 8.** [k plus proches voisins] Considérant un modèle de régression des 2 plus proches voisins (simple : pondération uniforme) avec distance euclidienne, quelle valeur est prédite pour la nouvelle ligne $(x_1, x_2) = (1, 2)$ sur le jeu de données ci-dessous ?

$(x_1, x_2, y) =$

(0, 0, 0)

(0, 1, 1)

(1, 1, 2)

(1, 0, 3)

A 1.5

B 2

C 2.5

D 3

- 9.** [Régression linéaire] Le R^2 est d'autant plus grand que la variance des données est élevée.

T True

F False

- 10.** [Régression linéaire] L'ajout d'un terme de pénalité dans l'expression à minimiser permet de...

A Éviter d'avoir de trop grands coefficients (notamment quand les données sont hautement corrélées).

B Annuler certains coefficients (considérés comme non significatifs) dans le cas du LASSO.

C S'assurer que tous les coefficients aient à peu près la même valeur absolue.