



Exemple d'examen

SCIPER: XXXXXX

Attendez le début de l'épreuve avant de tourner la page. Ce document est imprimé recto-verso, il contient X pages, les dernières pouvant être vides. Ne pas dégrafer.

Cette collection de questions est représentative de la couverture des thèmes et du type de questions incluses dans l'examen final. Cependant, l'examen final comptera environ le double de questions.

- Posez votre carte d'étudiant sur la table.
- Les notes de cours ne sont pas autorisées.
- Seul est autorisé un “cheat-sheet” sous forme d'une feuille papier au format A4 avec des inscriptions (formulaire, résumés, etc.) au recto et au verso. Ce dernier doit impérativement avoir été préparé de manière personnelle. La technique d'élaboration est libre (manuscrit, dactylographié, collage, etc.) mais il ne pourra en aucun cas prendre la forme d'un assemblage d'artéfacts copié-collés directement du cours ou des labos (pas de micro-fiche, photocopies directes, etc.).
- L'utilisation d'une **calculatrice** et de tout outil électronique est interdite pendant l'épreuve.
- Les questions à choix multiples peuvent n'avoir que des déclarations incorrectes; dans ce cas, il faut impérativement remplir le carré “Aucune de ces réponses n'est correcte” ou la question sera considérée comme non répondue et pas de points ne seront ajoutés au score.
- La notation est telle que des réponses entrées aléatoirement résultent en un score nul:
 - Chaque question avec un simple choix parmi N réponses possibles donne +3pts lorsque la réponse est correcte, $-\frac{3}{N-1}$ pts si la réponse est incorrecte et 0pt s'il n'y a pas de réponse.
 - À l'intérieur de chaque question à choix multiples, la réponse à chacune des N cases donne +4/Npt si l'état (coché ou vide) est correct et -4/Npt s'il est incorrect. Si aucune case n'a été cochée la question est considérée comme non-répondue et donne 0pt.
- Une question pourra être annulée s'il s'avère par la suite qu'elle est incorrecte ou ambiguë.
- Utilisez un **stylo** à encre **noire ou bleu foncé** et effacez proprement avec du **correcteur blanc** si nécessaire.
- Si une question est erronée, l'enseignant se réserve le droit de l'annuler.

Respectez les consignes suivantes Observe this guidelines Beachten Sie bitte die unten stehenden Richtlinien		
--	--	--

choisir une réponse select an answer Antwort auswählen	ne PAS choisir une réponse NOT select an answer NICHT Antwort auswählen	Corriger une réponse Correct an answer Antwort korrigieren
---	--	---



ne PAS choisir une réponse NOT select an answer NICHT Antwort auswählen		
--	--	--



Corriger une réponse Correct an answer Antwort korrigieren		
---	--	--



ce qu'il ne faut **PAS** faire | what should **NOT** be done | was man **NICHT** tun sollte



CORRECTION

Première partie, questions à choix multiple

Pour chaque question marquer la case correspondante à la réponse correcte sans faire de ratures. Il n'y a qu'**une seule** réponse correcte par question.

Question 1 Après avoir entraîné une régression linéaire sur ses données Martine se rend compte que le modèle sous-apprend. Laquelle des actions suivantes serait adaptée?

- Rendre le modèle plus complexe en ajoutant des variables et/ou utiliser un modèle polynomial avec un degré plus élevé
- Simplifier le modèle en enlevant des variables et/ou utiliser un modèle polynomial avec un degré inférieur
- Garder le même modèle mais réduire la taille du jeu d'entraînement
- Aucune des actions proposées ci-dessus n'est adaptée au problème

Question 2 Laquelle des affirmations suivante qui compare les avantages et inconvénients du choix d'une fonction de modélisation simple ou complexe est, en général, correcte?

- un modèle simple permet d'éviter le sur-apprentissage, requiert moins de temps de calcul, mais sa performance et sa capacité de généralisation seront moindres
- un modèle complexe permet d'éviter le sous-apprentissage, requiert plus de temps de calcul, peut permettre une meilleure performance, mais sa capacité de généralisation sera moindre
- un modèle complexe permet d'éviter le sur-apprentissage, requiert plus de temps de calcul mais permet une meilleure généralisation et performance
- un modèle simple permet d'éviter le sous-apprentissage, requiert moins de temps de calcul mais sa capacité de généralisation et sa performance seront moindres

Question 3 L'extraction de variables pour des données à p variables consiste à

- créer m nouvelles variables à partir des p variables dont on dispose initialement
- éliminer un nombre $p - m$ de variables des données

Question 4 Pour chaque vente d'un accessoires de mode sur son site de vente en ligne, une compagnie stocke le nom de l'article, le numéro d'article (qui identifie le type d'article), la date de l'achat, le prix de l'article, le rabais accordé (pourcentage), et les coordonnées géographiques (longitude et latitude) de l'adresse d'envoi. La compagnie désire visualiser toutes les ventes sur un graphique en 2 dimensions avec les paramètres les plus pertinents sur les deux axes. A quel type d'apprentissage ce problème correspond-il?

- apprentissage non-supervisé (clustering)
- apprentissage non-supervisé (réduction de dimension)
- apprentissage supervisé (régression linéaire)
- apprentissage supervisé (classification)

Question 5 La matrice de confusion d'un problème de classification binaire

- est un tableau avec 2 lignes et deux colonnes qui décompte les résultats selon qu'ils ont été classifié en TN, TP, FP, FN
- une matrice pour caractériser la moralité et la justice d'un modèle
- un tableau avec autant de lignes et colonnes que de jeux d'entraînement distincts qui permet de comparer la performance du modèle évalué sur les données d'entraînement et de test.

CORRECTION

Question 6 Pour évaluer la qualité d'un modèle appris, on sépare communément les données en jeux de données tel qu'on ait:

- pas de séparation du jeu de données (intégrité des données)
- un jeu d'entraînement, un jeu de validation, un jeu de test
- un jeu d'évaluation, un jeu de validation
- un jeu d'entraînement par classe

Question 7 Un test de dépistage qui a une haute précision indique:

- parmi les résultats identifiés comme positifs, une haute proportion est effectivement malade (vrais positifs)
- le test identifie correctement comme positif tous les cas de personnes vraiment malades (pas ou peu de faux négatifs)
- les résultats identifiés comme positifs ne sont probablement pas fiables (beaucoup de faux positifs)

Question 8 La fonction $k : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$, $x, x' \mapsto k(x, x') = (x - x')^2$ est-elle un noyau?

- Non, car pour les observations $x^1 = -1$, $x^2 = 1$, et les valeurs $a_1 = -3$ et $a_2 = 4$, on a: $\sum_{i=1}^2 \sum_{\ell=1}^2 a_i a_{\ell} k(x^i, x^{\ell}) < 0$.
- Oui, car quel que soit le choix de $x^1 \in \mathbb{R}$, $x^2 \in \mathbb{R}$, et $a_1 \in \mathbb{R}$, $a_2 \in \mathbb{R}$, on a: $\sum_{i=1}^2 \sum_{\ell=1}^2 a_i a_{\ell} k(x^i, x^{\ell}) \geq 0$.

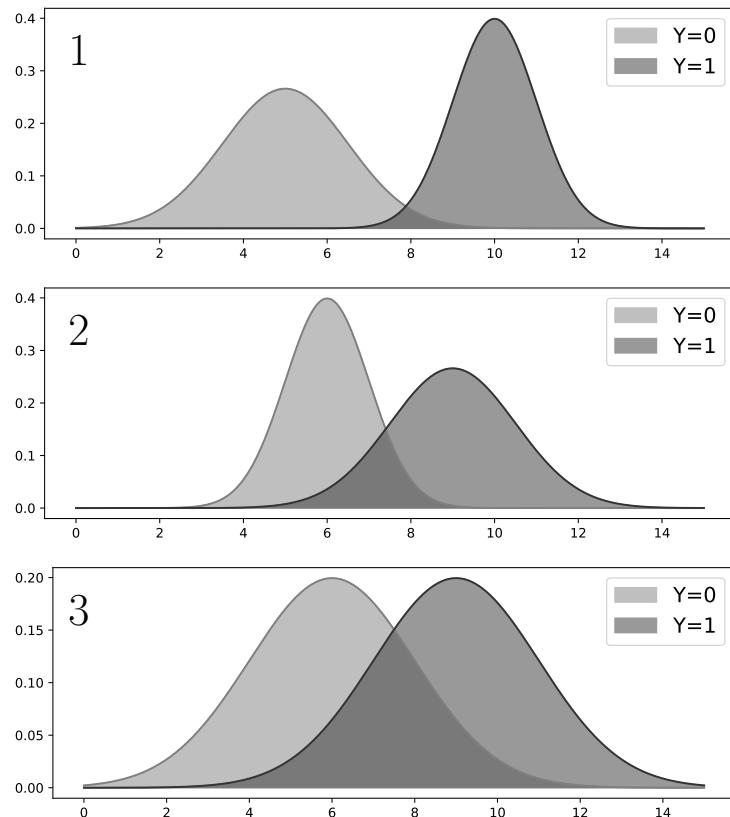
Question 9 Quelle est l'erreur empirique d'un modèle de classification ayant la matrice de confusion suivante?

16	3	2
9	21	4
3	4	38

- 25
- 1/4
- 0.33
- 75%

CORRECTION

Question 10 Soit un jeu de données à 3 dimensions. Les densités de population des observations dans chacune des dimensions sont représentées ci-dessous. En appliquant une sélection de variables par filtrage avec le critère de corrélation, quel est l'ordre de pertinence des dimensions du jeu de données, par importance décroissante?



2, 3, 1

1, 2, 3

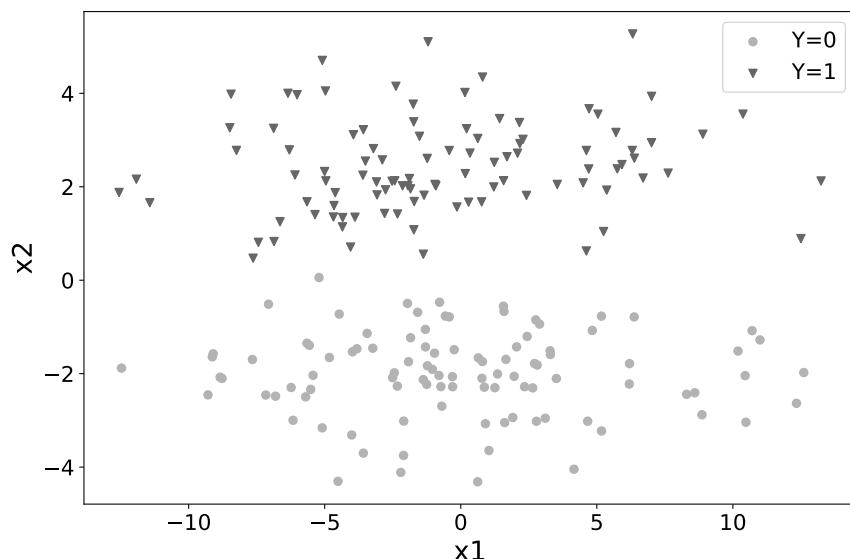
3, 2, 1

3, 1, 2

1, 3, 2

2, 1, 3

Question 11 Soit le jeu de données représenté ci-dessous. Quelle composante sera conservée si on le réduit à une dimension par une analyse en composantes principales (PCA)?



x1 x2

CORRECTION

Question 12 En considérant la propagation vers l'avant du réseau de neurones artificiel suivant:

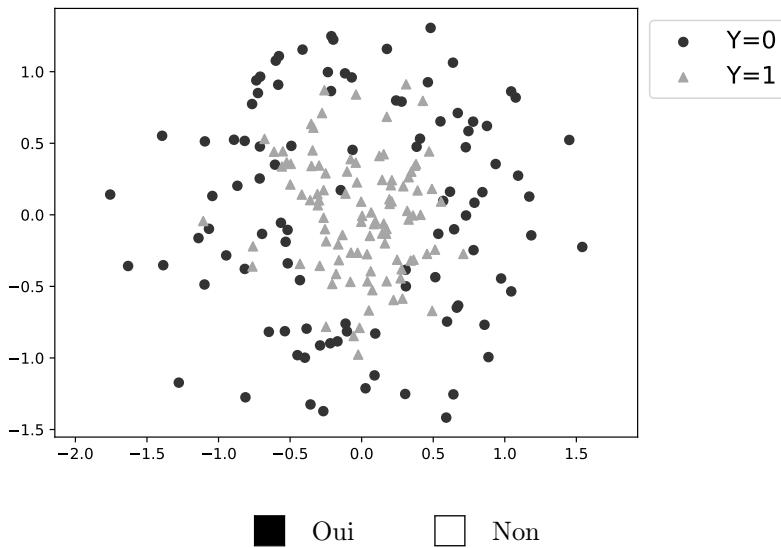
$$x^{(0)} \xrightarrow{W^{(1)}, b^{(1)}} s^{(1)} \xrightarrow{\sigma} x^{(1)} \xrightarrow{W^{(2)}, b^{(2)}} s^{(2)} \xrightarrow{\sigma} x^{(2)}$$

Avec $x^{(0)} \in \mathbb{R}^3$, $x^{(1)} \in \mathbb{R}^{100}$ et $x^{(2)} \in \mathbb{R}^2$; quelle est la taille des matrices W^1 et W^2 ?

- $W^1 \in \mathbb{R}^{100 \times 2}$ et $W^2 \in \mathbb{R}^{3 \times 100}$
 $W^1 \in \mathbb{R}^{2 \times 100}$ et $W^2 \in \mathbb{R}^{3 \times 100}$

- $W^1 \in \mathbb{R}^{3 \times 100}$ et $W^2 \in \mathbb{R}^{2 \times 100}$
 $W^1 \in \mathbb{R}^{100 \times 3}$ et $W^2 \in \mathbb{R}^{2 \times 100}$

Question 13 Un réseau de neurones artificiel contenant suffisamment de paramètres peut-il obtenir une erreur d'entraînement nulle sur les données suivantes?



- Oui Non

Question 14 En utilisant les conventions habituelles des réseaux de neurones convolutifs (CNN), quel est le résultat de la convolution suivante?

Entrée				
0	0	1	0	0
0	0	1	0	0
1	1	1	1	1
0	0	1	0	0
0	0	1	0	0

Filtre		
-1	0	1
-2	0	2
-1	0	1

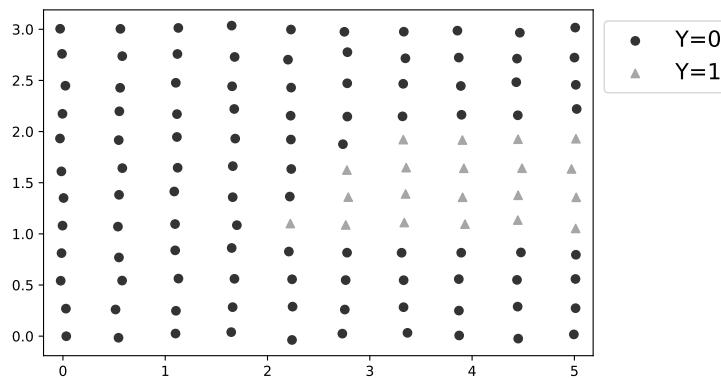
- $\begin{bmatrix} 3 & 2 & 3 \\ 0 & 0 & 0 \\ -3 & -2 & -3 \end{bmatrix}$ $\begin{bmatrix} 3 & 0 & -3 \\ 2 & 0 & -2 \\ 3 & 0 & -3 \end{bmatrix}$ $\begin{bmatrix} -3 & 0 & 3 \\ -2 & 0 & 2 \\ -3 & 0 & 3 \end{bmatrix}$ $\begin{bmatrix} -3 & -2 & -3 \\ 0 & 0 & 0 \\ 3 & 2 & 3 \end{bmatrix}$

Question 15 Dans un réseau de neurones convolutif, on applique une convolution avec un filtre de taille 20×15 à une image de taille 580×580 . Quelle est la taille de l'image résultante?

- 566×561 20×15 560×565
 580×580 565×560 561×566

CORRECTION

Question 16 Un arbre de décision est entraîné sur le jeu de données suivant, en utilisant l'algorithme CART. Combien de plans séparateurs possède le classificateur au minimum s'il obtient une erreur d'entraînement nulle?



- 2 3 4 5 6 7 8

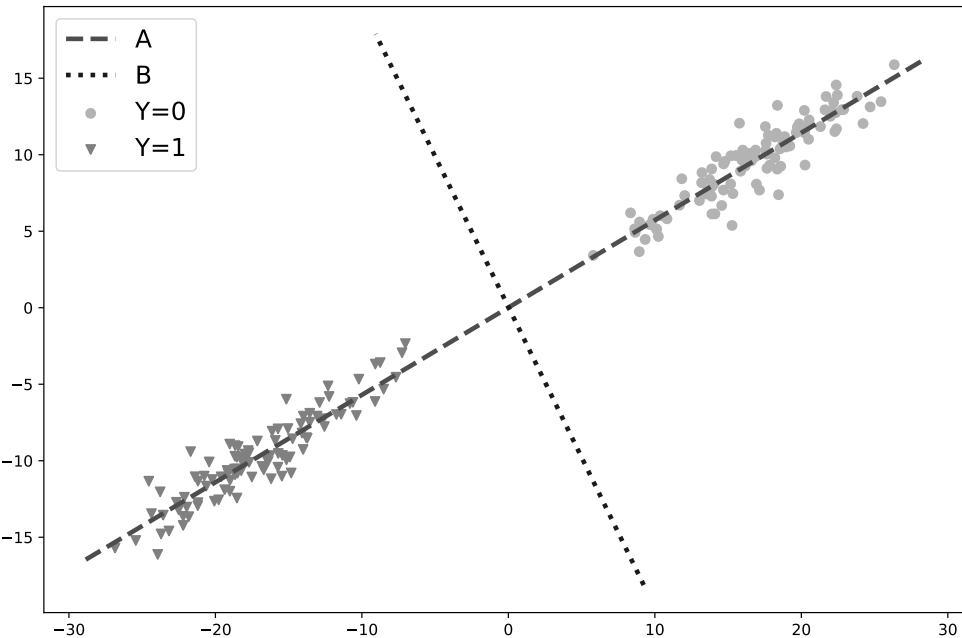
Question 17 Xavier, Yvette et Zahra observent les nuages depuis la place Cosandey, quand soudain ils aperçoivent un objet volant mais ils n'arrivent pas à se mettre d'accord sur ce que c'est. Le tableau suivant répertorie leurs impressions. En utilisant le bagging sur leurs prédictions, quel objet doit-on conclure qu'ils ont observé?

Objet	Xavier	Yvette	Zahra
Mouette	3%	0%	4%
Soucoupe volante	96%	22 %	44%
Hélicoptère	0%	3 %	0%
Drône	1%	75 %	52%

- Drône Mouette Soucoupe volante Hélicoptère

CORRECTION

Question 18 Dans la figure suivante, que représente la droite B?



- La prédiction d'un modèle de régression linéaire
- La prédiction d'un modèle de régression logistique
- La frontière de décision d'un modèle de régression logistique

Question 19 Le gradient de la fonction de coût logistique est donné par

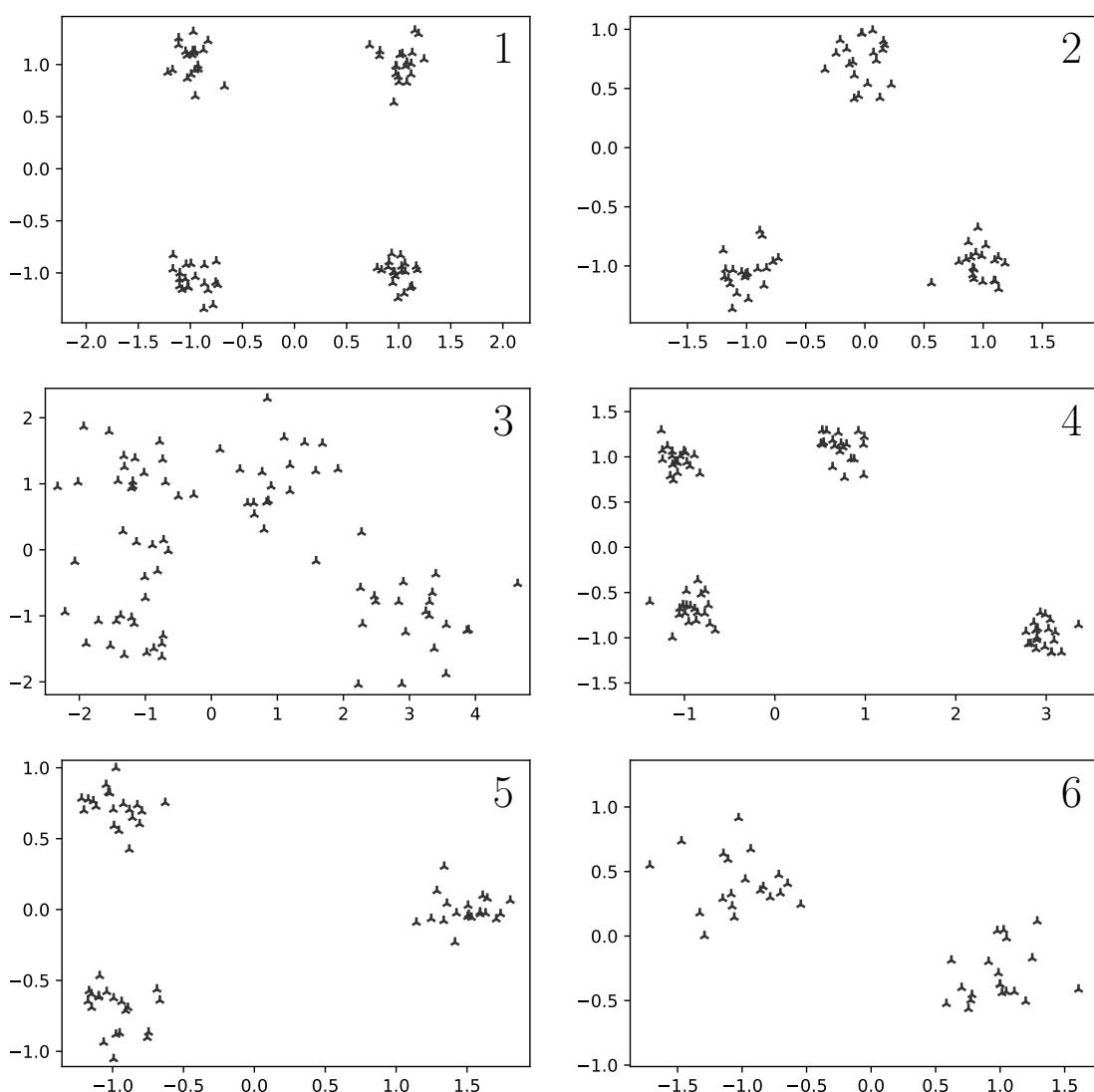
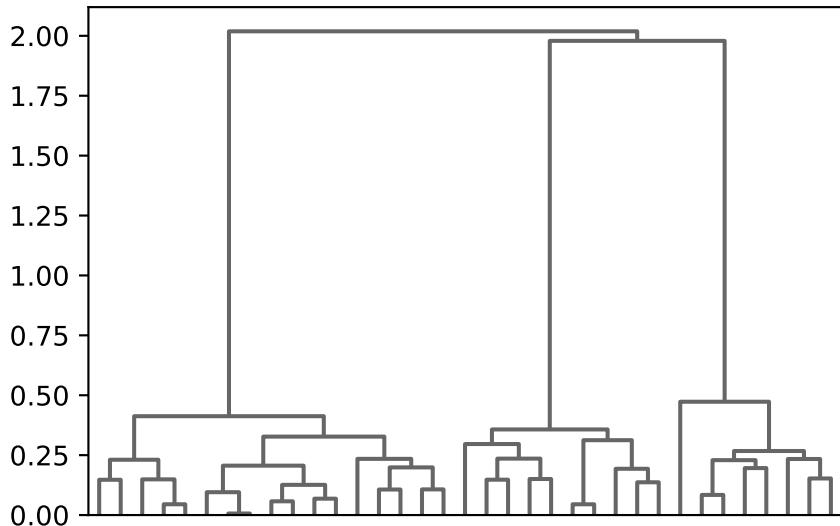
$$\frac{dL}{d\vec{\beta}} = \left(\sigma(\vec{\beta}^T \vec{x}) - y \right) \vec{x}$$

avec σ la fonction logistique. Quelle ligne de code calcule correctement ce gradient?

- $-(y - 1 / (1 + np.exp(-x.dot(beta)))) * x$
- $(1 / (1 + np.exp(-x.dot(beta))) - y) * x$
- $(1 / (1 + np.exp(x.dot(beta))) - y) * x$
- $(y - 1 / (1 - np.exp(x.dot(beta)))) * x$
- $(1 / (1 - np.exp(x.dot(beta))) - y) * x$

CORRECTION

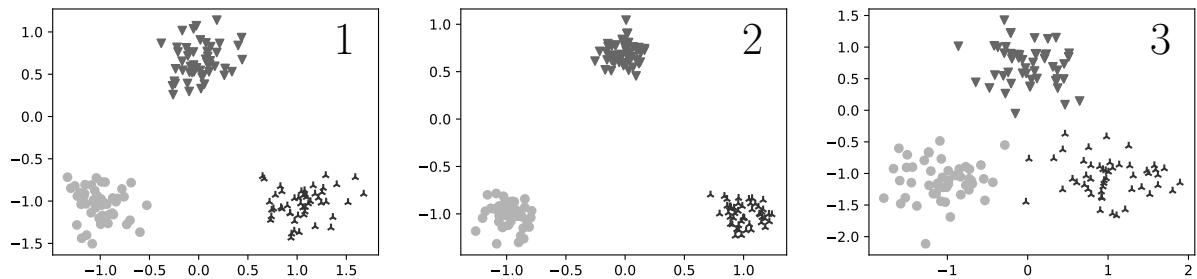
Question 20 Quel jeu de données correspond au dendrogramme suivant, construit en utilisant un clustering agglomératif à lien moyen?



<input type="checkbox"/>	1	<input checked="" type="checkbox"/>	2	<input type="checkbox"/>	3	<input type="checkbox"/>	4	<input type="checkbox"/>	5	<input type="checkbox"/>	6
--------------------------	---	-------------------------------------	---	--------------------------	---	--------------------------	---	--------------------------	---	--------------------------	---

CORRECTION

Question 21 Soient trois jeux de données séparés en trois clusters chacun, affichés ci-dessous. Trier les trois cas par indice de Davies-Bouldin croissant.



3, 1, 2
 2, 1, 3

3, 2, 1
 2, 3, 1

1, 2, 3
 1, 3, 2

Question 22 Soient Y le vecteur contenant les étiquettes et F le vecteur contenant les prédictions, deux tableaux numpy. Quelle ligne de code permet de calculer la précision des prédictions?

- $(Y==F).mean()$
 $np.mean(Y=F)$
 $np.sum(Y==F) / Y.shape$

CORRECTION

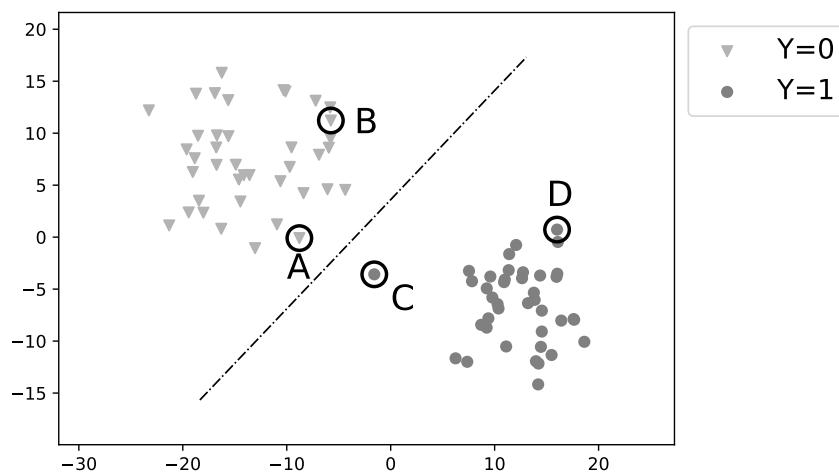
Deuxième partie, questions à choix multiples

Pour chaque question, marquer la case ou les cases qui correspondent à la réponse ou aux réponses correctes.

Question 23 ♣ Quelle(s) ligne(s) de code python calcule(nt) la solution optimale du problème de régression linéaire $\vec{y} = X\vec{b}$?

- `b = X.T @ y @ np.linalg.inv(X.T @ X)`
- `b = np.linalg.inv(X.T @ X) @ X.T @ y`
- `b = np.linalg.inv(X.T.dot(X)).dot(X.T).dot(y)`
- `b = X.Ty np.linalg.inv(X.TX)`
- Aucune de ces réponses n'est correcte.*

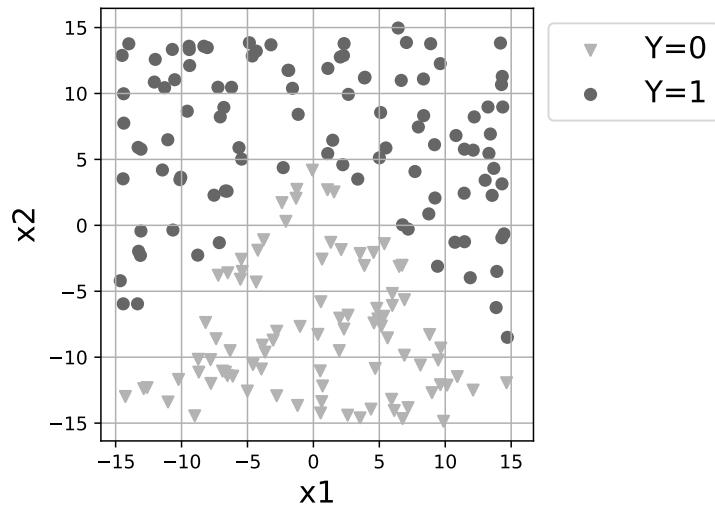
Question 24 ♣ Soit un modèle SVM entraîné sur un problème de classification binaire. Le jeu de données et l'hyperplan séparateur du modèle sont représentés ci-dessous. Parmi les points proposés, lesquels sont des vecteurs de support du modèle?



- A
- B
- C
- D
- Aucune de ces réponses n'est correcte.*

CORRECTION

Question 25 ♣ Soit le jeu de données représenté ci-dessous. Quelle(s) transformation(s) de l'espace permet(tent) de le rendre linéairement séparable pour la classification binaire?



- $x_1, x_2 \rightarrow x_1, x_2^2$
- $x_1, x_2 \rightarrow x_1^2, |x_2|$
- $x_1, x_2 \rightarrow x_1^2, x_2$

- $x_1, x_2 \rightarrow x_1^2, x_2^2$
- $x_1, x_2 \rightarrow x_1, |x_2|$
- $x_1, x_2 \rightarrow |x_1|, x_2$

- $x_1, x_2 \rightarrow |x_1|, |x_2|$
- $x_1, x_2 \rightarrow |x_1|, x_2^2$
- Aucune de ces réponses n'est correcte.