

**Examen Final en Apprentissage Automatique Semestre 2 - 2019/2020 Durée : 01h00**

Les questions contenant le symbole ♣ peuvent présenter zéro, une ou plusieurs bonnes réponses. Les autres ont une unique bonne réponse.

Question 1 Un système d'apprentissage automatique est dit non-supervisé lorsque la cible est explicitement connue.

- A VRAI B FAUX

Question 2 Le nombre des représente les exemples de classe négative dont la classe est prédite comme positive.

- A faux négatifs (FN) B faux positifs (FP) C vrais négatifs (VN)
 D aucunes de ces réponses E vrais positifs (VP)

Question 3 Les domaines d'application de l'apprentissage non-supervisé n'incluent pas

- A Le Web et les réseaux sociaux B La médecine et la bio-informatique
 C La vision artificielle D Les services financiers E Aucune de ces réponses
 F Les domaines de l'énergie

Question 4 ♣ Les paramètres de la régression logistique sont estimés en utilisant

- A aucunes de ces réponses B la montée du gradient C une approche exacte
 D le maximum de vraisemblance E la fonction sigmoïde

Question 5 Le nombre des représente les exemples de classe négative dont la classe est prédite comme négative.

- A faux positifs (FP) B faux négatifs (FN) C vrais négatifs (VN)
 D vrais positifs (VP)

Question 6 La commande appropriée pour l'utilisation de la technique de validation croisée est :

- A $vc = cvpartition(y, 'Kfold', n);$ B $vc = cvpartition(y, 'Holdout', n);$
 C $vc = cvpartition(y, 'Leaveout', n);$ D $vc = cvpartition(y, 'VCroised', n);$

Question 7 Quand l'hypothèse de Bayes naïve n'est pas satisfaite, le classificateur de Bayes est biaisé et obtient plus de performance que la régression logistique.

- A FAUX B VRAI

Question 8 Pour trouver l'hypothèse f la plus optimale pour la prédication,

- A on prend w qui maximise la fonction d'erreur.
 B on prend w qui minimise la fonction d'erreur.
 C on prend w qui maximise la fonction $f(x)$.
 D on prend w qui minimise la fonction $f(x)$.

Question 9 $P(y = C_k | X = x, D)$ donne la probabilité d'observer la classe C_k si on observe la donnée $X=x$ sachant que l'on dispose de l'ensemble des exemples D .

- A FAUX B VRAI



Question 10 La commande suivante calcule les paramètres de la régression linéaire :

- A aucunes de ces réponses B $w = \text{inv}(x' * x) * x' * y;$ C $w = x' * x * \text{inv}(y * x');$
 D $w = \text{inv}(x * x) * x * y;$ E $w = x' * x * \text{inv}(x' * y);$

Question 11 Dans le contexte SVM, le produit scalaire est remplacé par un noyau afin de

- A aucunes de ces réponses
 B éviter de faire explicitement la transformation de l'espace R^D vers l'espace R^M .
 C faire la classification linéaire des données et tolérer les erreurs de classification.
 D régler le problème des classes de données qui ne sont pas linéairement séparables.

Question 12 Le maximum de vraisemblance vise à maximiser

- A les probabilités des classes B le taux de succès du classificateur
 C l'erreur de validation du classificateur D le produit des probabilités des classes
 E aucunes de ces réponses

Question 13 La commande `predict(A, B)` permet de

- A Générer le modèle A en utilisant les données B .
 B Estimer les classes des données A en utilisant le modèle B .
 C Estimer les classes des données B en utilisant le modèle A .
 D Générer le modèle B en utilisant les données A .

Question 14 ♣ Google translate et une application de quels domaines ?

- A Systèmes de recommandation des spams
 B Traduction automatique
 C Vision artificielle : reconnaissance d'objets, segmentation
 D Traitement automatique de la langue
 E Analyse financière

Question 15 On suppose le cas simple de la classification binaire avec la donnée x ayant D dimensions. Quel est le nombre de paramètres de la régression logistique ?

- A $D+1$ paramètres B $3D+1$ paramètres C D paramètres
 D $2D+1$ paramètres

Question 16 A quoi sert une fonction d'erreur ?

- A évaluer un modèle de prédiction.
 B comparer deux ou plusieurs modèles de prédiction.
 C aucunes de ces réponses
 D mesurer combien une prédiction est loin des "vraies" valeurs.

Question 17 Un algorithme est qualifié de bon algorithme d'apprentissage si :

- A aucunes de ces réponses
 B il est capable de généraliser ce qu'il a appris à de nouvelles données.
 C il est basé sur le modèle de régression logistique.
 D il fournit à l'algorithme des données d'entraînement.