

Ecole Polytechnique de Thies
Département Génie Informatique et Télécommunications
TD1 AD2-ML 2018-2019

Exercice 1 *Rappels*

1. *Quels sont les principaux avantages à réduire la dimension d'un jeu de donnée ? Quels en sont les principaux inconvénients ?*
2. *Qu'appelle t'on fléau de la dimension ?*
3. *Après avoir effectué une réduction de dimension sur un jeu de données, est-il possible d'inverser l'opération ? Si oui, comment ? Sinon, pourquoi ?*
4. *Peut on utiliser une PCA pour effectuer une réduction de dimension sur un jeu de données hautement linéaire ?*
5. *Supposons que vous effectuez une PCA sur un jeu de données de dimension 1000, en fixant la contribution à la variance à 95%. Combien le jeu de données résultant aura t-il de dimensions ?*
6. *Dans quel cas utiliseriez-vous une PCA ordinaire, une PCA incrémentale, une PCA randomisée ou une PCA à noyau ?*
7. *Comment pouvez-vous évaluer les résultats d'un algorithme de réduction de dimension sur votre jeu de données ?*
8. *Peut-il y avoir un intérêt à enchaîner deux algorithmes de réduction de dimension différents ?*
9. *Chargez le jeu de données MNIST et partagez-le entre un jeu d'entraînement et un jeu de test (prenez les 60000 premières observations pour l'entraînement et les 10000 restantes pour le test). Entraînez un classificateur de type forêt aléatoire sur le premier jeu et mesurez le temps de calcul, puis évaluez le modèle obtenu sur le jeu de données, en conservant 95% de la variance. Entraînez un nouveau classificateur de type forêt aléatoire sur le jeu de donnée réduit et mesurez le temps de calcul. Quel a été l'entraînement le plus rapide ? Évaluez ensuite le classificateur sur le jeu de test : est-il meilleur ou moins bon que le précédent ?*
10. *Utiliser t-SNE pour réduire le jeu de données MNIST à deux dimensions puis représentez graphiquement le résultat en utilisant Matplotlib.*

Exercice 2 *Généralités sur l'apprentissage*

1. Comment définirez vous le Machine Learning et par quelle autre appellation le désigne t'on en Français ?
2. Pouvez vous nommez quatre types de problèmes où il se distingue ?
3. Qu'est ce qu'un jeu d'entraînement étiqueté ?
4. Quelles sont les deux tâches supervisées les plus communes ?
5. Pouvez vous nommer quatre tâches non supervisées courantes ?
6. Quel type d'algorithme d'apprentissage automatique utiliserez vous pour permettre à un robot de s'aventurer sur divers terrains inconnus ?
7. Quel type d'algorithme utiliserez vous pour segmenter vos clients en plusieurs groupes ?
8. Classez-vous le problème de la détection de spam parmi les problèmes d'apprentissage supervisé ou parmi les problèmes d'apprentissage non supervisé ?
9. Qu'est ce qu'un système d'apprentissage en ligne ?
10. Qu'est ce que l'apprentissage hors-mémoire ?
11. Quel type d'algorithme d'apprentissage s'appuie sur une mesure de similarité pour effectuer des prédictions ?
12. Que recherche un algorithme d'apprentissage basé sur un modèle ? Quelle stratégie utilise-t'il le plus couramment pour réussir ? Comment effectue-t-il des prédictions ?
13. Pouvez vous citer quatre des principales difficultés de l'apprentissage automatique ?
14. Si votre modèle donne de bons résultats sur les données d'entraînement mais généralise mal aux nouvelles observations, que se passe t'il ? Pouvez vous indiquer trois solutions possibles ? à Qu'est-ce qu'un jeu de test et à quoi sert il ?
15. Quel est le but d'un jeu de validation ?
16. Que peut il se produire de facheux si vous ajustez vos hyperparamètres à l'aide du jeu test ?
17. Qu'est ce que la validation croisée et our qu'elle raison pourriez vous la préférez à un jeu de validation ?