

TP noté

Introduction à l'intelligence artificielle

Durée : 1h30

Documents et supports autorisés : supports de cours et de TD, sujets et notes personnelles de TD, documentations officielles des packages et réponses de forums en ligne. Tout élément non listé est interdit.

Exercice 0 – 3 points

L'exercice 0 concerne le nommage et la présentation des éléments que vous allez remettre. Si l'ensemble de vos remises respecte les critères énoncés ci-dessous, vous bénéficierez des points alloués à l'exercice 0, selon le barème. Sinon, vous recevrez 0 pour l'exercice.

Critères à respecter

- Les fichiers doivent être nommés selon le format **nom_prenom_l3maths_exoN.py**, où **nom**, **prenom** et **N** sont remplacés respectivement par votre nom, votre prénom et le numéro de l'exercice.
- En cas de réponse rédigée, cette dernière doit être placée dans le fichier python correspondant à l'exercice, à la fin, comme commentaire. Le numéro de la question doit être précisé.
- Votre mail doit avoir pour objet « TP noté L3 Maths 2024 » et contenir, dans le corps, votre nom et votre prénom. L'adresse à laquelle envoyer vos travaux sera précisée en séance.
- Si vous commentez votre code, les commentaires doivent être en français.

Exercice 1 – 6 points

Le fichier **wine.csv** contient diverses mesures issues d'analyses réalisées sur des vins. On cherche à se doter d'un outil permettant à un expert humain d'annoter les observations une par une.

1. Etudiez la structure du fichier **wine.csv**.
 - a. Où trouver le nom des variables ?
 - b. Combien d'observations contient le jeu de données ?
 - c. Quelle est la dimension du jeu de données ?
2. Créez une fonction **load_wine()** qui charge le fichier **wine.csv**, et qui renvoie deux éléments : une liste, contenant le nom des variables, et un tableau numpy, contenant les observations. La fonction devra prendre en paramètre le chemin d'accès vers le fichier.
3. Créez une fonction **annotate()** qui prend en paramètre une liste de noms de variables et un tableau numpy, et renvoie une liste d'entiers. La fonction doit permettre à un utilisateur de saisir une valeur numérique entre 0 et 10 pour chaque ligne du fichier. Cette liste est renvoyée à la fin de la fonction.
4. Créez une fonction **save_to_disk()** qui prend en paramètre un tableau numpy, un nom de fichier, et un séparateur de champ. La fonction sauvegarde sur le disque le tableau passé en paramètre, en le nommant et en le formattant selon les autres paramètres indiqués.

La qualité (ergonomie, praticité) de l'affichage, pour la question 3., sera pris en compte pour la notation.

Exercice 2 – 8 points

On s'intéresse à un jeu de données concernant les usagers victimes d'accidents de la route. Deux fichiers sont à votre disposition : **usagers-2022.csv** et **description-des-bases-de-donnees-annuelles-2022.pdf**. Pour chaque question, vous devrez produire du code source à remettre.

1. Le fichier contient-il des données manquantes ? Lesquelles, et combien ?
2. Combien d'observations contiennent les données ?
3. Combien de véhicules différents sont concernés par au moins un accident en 2022 ?
4. Quel âge avait l'utilisateur accidenté le plus jeune ? Le plus vieux ?
5. Quelle est la moyenne d'âge des accidentés conducteurs d'un véhicule ?
6. On se demande s'il est possible de prédire, au moyen d'un classifieur, la gravité d'un accident. Répondez en essayant de classifier les données au moyen d'au moins un classifieur. Les étapes de préparation de données, d'apprentissage et de généralisation doivent être suivies.

Attention : vous devez faire, au choix, l'exercice 3 ou l'exercice 4

Exercice 3 – 3 points

On s'intéresse au jeu de données iris.csv. Pour répondre à l'exercice, vous devez produire du code source, qui est à remettre.

Après avoir normalisé le jeu de données, comparez les performances en *hold-out 60/40* avec celles obtenues en *10-cross-validation*, pour un k-NN où k varie de 3 à 10. Commentez les résultats obtenus.

Exercice 4 – 3 points

On souhaite mettre au point une mesure de distance entre deux mots. Pour répondre à l'exercice, vous devez produire du code source, qui est à remettre.

On compare les positions une à une. Pour chaque position, si les caractères trouvés à la position courante sont identiques, on compte 0, sinon, on compte la distance qui sépare les deux lettres dans l'alphabet. Cette technique ne permet pas de comparer deux mots de longueur différentes.

1. Quelle sera la distance entre deux mots identiques ?
2. Quelle est la distance entre les mots *avenants* et *attendant* ?