

Apprentissage Artificiel

TP 1 - Introduction

Issam Falih
issam.falih@uca.fr

Instructions : Préparez un rapport incluant le code source et vos résultats, et déposez-le sur Moodle. Il est recommandé d'utiliser streamlit pour afficher vos résultats .

A. Manipulation d'un jeu de données

1. Chargez les données **Iris** avec la commande : `iris = datasets.load_iris()`

La variable `iris` est un objet qui contient :

- la matrice des données (`iris.data`),
 - un vecteur de numéro de classe (`target`),
 - les noms des variables (`feature_names`),
 - le nom des classes (`target_names`).
2. Affichez les données, les noms des variables et le nom des classes (utilisez `print`).
 3. Affichez le nom des classes pour chaque donnée.
 4. Affichez la moyenne (*mean*), l'écart-type (*std*), le min et le max pour chaque variable.
 5. En utilisant les attributs `size` et `shape`, affichez le nombre de données, le nombre de variables et le nombre de classes.

B. Téléchargement et importation de données

Il est possible de charger un jeu de données directement depuis le package `dataset` de `scikit-learn` qui contient de nombreux jeux de données gratuits.

1. Importez les données '**MNIST original**'. On va charger 1500 images d'une manière aléatoire.

```

from sklearn.datasets import fetch_openml
mnist = fetch_openml('mnist_784', version=1, as_frame=False)

shuffled_indices = np.random.permutation(len(mnist.data))

subset_indices = shuffled_indices[:1500]

X = mnist.data[subset_indices]
y = mnist.target[subset_indices]

```

2. Affichez la matrice des données, le nombre de données et de variables, les numéros de classes pour chaque donnée, ainsi que la moyenne, l'écart-type, les valeurs min et max pour chaque variable ; enfin donnez le nombre de classes avec la fonction `unique`.

C. Génération de données et affichage

1. Utilisez l'aide (`help`) pour voir comment utiliser la fonction `datasets.make_blobs`.
2. Générez 1000 données de deux variables réparties en 4 groupes.
3. Utilisez les fonctions `figure`, `scatter`, `title`, `xlim`, `ylim`, `xlabel`, `ylabel` et `show` pour afficher les données avec des couleurs correspondant aux classes. Les axes `x` et `y` seront dans l'intervalle `[-15, 15]` et devront avoir un titre. La figure doit aussi avoir un titre.
4. Générez 100 données de deux variables réparties en 2 groupes, puis 500 données de deux variables réparties en 3 groupes. Concaténez (`vstack` et `hstack`) les deux jeux de données et les numéros de classe pour fusionner les deux jeux de données. Affichez les trois ensembles avec `scatter` comme précédemment.