

# Analyse, classification et indexation des données

## (Mini) projet

### Présentation

Le but de ce devoir est de tester les différents algorithmes de machine learning vus en cours et en td. Nous allons pour cela travailler avec un jeu de donnée simple : le corpus `fashion_mnist`. Ce corpus fait partie des datasets disponibles sur `keras` et peut donc être chargé grâce aux instructions :

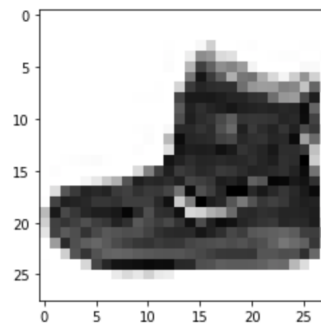
```
from keras.datasets import fashion_mnist
(X_train, y_train), (X_test, y_test) = fashion_mnist.load_data()
```

Le corpus contient 60000 images à  $28 \times 28$  pixels au niveau de gris. Chaque image représente un habit. Le nombre de catégories d'habit est de 10, chacune est codée par un chiffre :

- 0 : T-shirt/top
- 1 : Trouser
- 2 : Pullover
- 3 : Dress
- 4 : Coat
- 5 : Sandal
- 6 : Shirt
- 7 : Sneaker
- 8 : Bag
- 9 : Ankle boot

La figure suivante donne le code et un exemple d'une image du corpus.

```
import matplotlib.pyplot as plt
import matplotlib.cm as cm
%matplotlib inline
plt.imshow(X_train[0], cmap=cm.Greys)
plt.show()
print(y_train[0])
```



9

FIGURE 1 –

## Travail demandé

Le but de ce travail est de tester les différents algorithmes d'apprentissage vus en cours/tds sur le corpus présenté ci-dessus. Concrètement, l'équipe pédagogique attend de vous de :

1. tester les algorithmes et identifier ceux efficaces pour le type de données représenté par ce corpus,
2. étudier l'impact des différentes techniques de réduction de dimension sur la faisabilité, l'efficacité et la complexité de la classification.

## Modalités pratiques

1. Le projet est à réaliser en binômes
2. Il doit être déposé par un membre du groupe sur moodle. Le nom du fichier doit indiquer les noms du binôme et le groupe. Le format attendu des noms de fichiers est le suivant `GrX-Nom1_Prenom1-Nom2_Prenom2.format`. Attention : des pénalités seront appliquées en cas de non respect de cette convention.
3. La date limite pour déposer vos fichiers est fixée au 02 janvier 2022 à 23h59.