

Cette séance est consacrée à la classification non supervisée, aussi appelée clustering.

1 Application de l'algorithme KMeans

1.1 Extraction de couleurs dans une image

1. Charger l'image et importer les bibliothèques nécessaires avec les commandes suivantes:

```
import matplotlib.pyplot as plt
import matplotlib.image as img
import numpy as np

from sklearn.cluster import KMeans
from sklearn.metrics import pairwise_distances_argmin
from sklearn.utils import shuffle

# Importation de l'image
from sklearn.datasets import load_sample_image
image = load_sample_image("china.jpg")
image = np.array(image, dtype=np.float64) / 255

# Affichage
plt.figure()
plt.clf()
plt.axis("off")
plt.title("Image originale")
plt.imshow(image)
```

2. Créer un tableau `numpy` avec l'image et la redimensionner avec :

```
w, h, d = original_shape = tuple(image.shape)
im_size = w*h
assert d == 3
image_array = np.reshape(image, (w * h, d))
```

3. Créer et ajuster un classifieur de type KMeans avec la commande `KMeans` de `scikit-learn`.
4. Afficher les centres des clusters et leur label. Puis reconstruire l'image avec un nombre réduit de couleurs à l'aide de la fonction `predict` de votre classifieur.
5. Recréer l'image à partir des labels prédits (il faut associer chaque label au centre du cluster).