

## TP2 – Analyse en composantes principales, classification et reconstruction

### I. Chargement et mise en forme des données

On utilisera les mêmes données que lors du TP1. Charger les données et réaliser les prétraitements définis dans les parties I et II du TP1.

### II. Analyse en composantes principales et classification

Définissez la décomposition en composantes principales en utilisant la fonction `PCA()` en gardant le maximum de composantes puis tracer les variances en utilisant la méthode `pca.explained_variance_ratio_()`. Définissez la décomposition en utilisant la fonction `PCA()` en conservant 100 composantes et l'appliquer sur les données en utilisant la méthode `fit()`. Réaliser la classification sur les données de départ puis sur les nouvelles données avec la méthode du 5PPV et la distance de Manhattan. Conclure.

#### Questions

Que représentent les valeurs renvoyées par `pca.explained_variance_ratio_` ?

Comment varient les temps de calcul ?

### III. Analyse en composantes principales et reconstruction

Définissez la décomposition en utilisant la fonction `PCA()` en conservant 50 composantes et l'appliquer sur les données en utilisant la méthode `fit()`.

Récupérer les vecteurs propres en utilisant une méthode de `PCA()`. Redimensionner les vecteurs propres en images propres (`np.reshape()`) de manière à pouvoir les visualiser sous forme d'images (array de taille 50x62x47). On utilisera la fonction `plot_gallery()` pour la visualisation.

#### Questions

Que représentent les vecteurs propres ? Quelle est leur taille ?

On souhaite transmettre les images de `X_test` en utilisant le moins de bande passante possible. Pour cela, les 50 images propres sont transmises une fois. Pour chaque nouvelle image, on transmet uniquement ses composantes dans le nouveau système d'axe de dimension 50. L'image est ensuite reconstruite à l'arrivée.

Calculer l'ACP des images de `X_test` puis reconstruisez les images à partir d'une des méthodes de `PCA()`. Remettre les données dans leur forme initiale en utilisant une des méthodes de `StandardScaler`. Afficher les images reconstruites et les comparer aux images initiales avec une distance euclidienne.

#### Questions

Lorsque l'on conserve 50 composantes, quel est le taux de compression, pourquoi ?

Quel est le principe de la reconstruction des images ?

Faire varier le nombre de composantes conservées et calculer l'erreur de reconstruction (norme L2). Afficher l'erreur de reconstruction en fonction du nombre de composantes.

#### Questions

- Que représente l'erreur de reconstruction ?
- Comment varie-t-elle en fonction du nombre de composantes ?