

TP4 : Réduction de la dimension

- 1) Récupérer la base de données relative à l'activité humaine sur le site suivant <https://archive.ics.uci.edu/ml/machine-learning-databases/00240/>.
- 2) Décrire la base de données téléchargée.
- 3) Evaluer les performances, en termes de taux de bonne classification, rappel, précision et f-mesure, d'une approche de classification supervisée de votre choix pour la reconnaissance des activités humaines de la base de données téléchargée (utiliser comme base d'apprentissage X_train.txt et y_train.txt). Pour cela utiliser la méthode de validation croisée (**10-fold cross validation**).
- 4) Projetez les données dans l'espace sur les composantes principales de l'ACP.
- 5) Etudier successivement l'inertie expliquée par chaque axe (vecteur propre) puis l'inertie expliquée cumulée.
- 6) Appliquer l'approche de classification utilisée précédemment et calculer taux de bonne classification en fonction du nombre d'axes principaux (e.g de 1 à 400).
- 7) Dresser la courbe du taux de bonne classification moyen pour chaque nombre d'axes principaux.
- 8) Interpréter le résultat obtenu.
- 9) Choisir et justifier le nombre d'axes principaux. Donner le taux de bonne classification correspondant.
- 10) Conclure