

Cours PCD – Labo 4 : ACP sur des images

Objectifs

- Explorer et documenter avec précision un ensemble d'images de visages.
- Mettre en place des expériences de reconnaissance de visages.
- Comparer la qualité de la reconnaissance avant et après l'Analyse en Composantes Principales.

Tâches et questions

0. Examinez la [démonstration fournie par « Scipy Lectures »](#) sur la combinaison de l'ACP et d'un classifieur SVM, en exécutant le *notebook* fourni avec la démo (`plot_eigenfaces.ipynb`). Le but de ce labo est d'expliquer et d'étendre ce code, dans un nouveau *notebook*.
1. Veuillez décrire avec précision le jeu de données « Olivetti Faces » en écrivant des commandes utiles pour obtenir ses divers paramètres. Répondez p.ex. à ces questions :
 - a. Que contient la structure de données retournée par `datasets.fetch_olivetti_faces()` ?
 - b. Combien de personnes sont représentées ? Combien d'images par personne ?
 - c. Quelle est la taille des images ? Comment sont-elles codées ? Y a-t-il une duplication ?
 - d. Veuillez afficher une image extraite des données.
2. Préparation : veuillez constituer un jeu d'entraînement avec les 80% premières images de chaque personne, et un jeu de test avec les 20% dernières. N'oubliez pas d'inclure aussi les identifiants des personnes. Quels sont les dimensions des jeux, à savoir le nombre d'items et d'attributs ?
3. Tentez une expérience de classification avec chacun des classifieurs suivants : KNN, régression logistique, SVM. Quels scores de « [accuracy](#) » obtenez-vous sur le jeu de test ?
4. Veuillez appliquer la méthode ACP avec 150 composantes principales : construisez d'abord les composantes (appelées « [eigenfaces](#) »), puis appliquez la transformation aux données d'entraînement, puis de test. Devez-vous centrer les données avant d'appliquer l'ACP ?
5. Refaites une expérience du point (3) avec les données transformées : quels sont les changements ?
6. Pour chacun des trois classifieurs (SVM, régression logistique, KNN), veuillez chercher le nombre de composantes principales (à 10 près) qui donne le meilleur score sur le jeu de test.
7. Pour le meilleur système obtenu, affichez les 5 premières erreurs de classification, avec : (a) l'image mal classée, (b) une image de la personne supposée par le système, et (c) une image de la personne correcte. Veuillez commenter les similarités et différences.
8. Pour le meilleur système, veuillez tester en utilisant [un test de Student apparié](#) si l'utilisation de l'ACP est significativement supérieure (à un niveau de 5%) que l'absence d'ACP. Pour appliquer le test, veuillez utiliser la validation croisée, en construisant 5 « folds » 80%/20%.