

Date limite de rendu : 27/04/2019

Travail à faire par groupe de 1,2 ou 3.

Format de fichier accepté : pynb, doc, pdf

Que ce soit en notebook ou doc/pdf, joindre le code dans une rubrique Annexe à la fin. Dans vos réponses, vous ferez référence au code utilisé à l'aide d'une numérotation. Des points sont attribués pour la lisibilité du code, l'orthographe et la présentation générale.

Exercice 1 (60%)

C'est un exercice de classification d'images. L'objectif est d'abord de réduire la dimension avec l'ACP puis de classer les images. Vous pouvez utiliser l'ACP de scikit-learn. Le K plus proche voisin est à implémenter sans l'utilisation de packages (excepté numpy,pandas).

1. Décrivez les 4 datasets. Représentez graphiquement le visage moyen (moyenne de toutes les lignes) de l'échantillon train.
2. Appliquez une ACP pour réduire la dimension du dataset pour $k=5, 10$ et 50 .
3. Choisissez un visage et représentez le graphiquement pour $k=5,10,50$.
4. L'ACP est plutôt longue lorsque la quantité de colonnes est très très grande, proposez une solution.
5. En utilisant le dataset réduit par ACP, implémentez un K plus proche voisin, pour pouvoir classer les images du test.
6. Faites une analyse d'erreur sur vos labels prédits et les vrais labels du test.
7. Offrez des recommandations pour améliorer le modèle et appliquez les.

Exercice 2 (40%)

En prenant un exemple de jeu de données d'au moins 1000 observations, proposez une étude de votre choix incorporant des notions de cours et des notions extérieures si vous le souhaitez. Les travaux originaux, non plagés, seront valorisés.