

Analyse, classification et indexation des données

CC1 - Rapport

Exercice 1 :

1) C1 est une classe contenant 6000 échantillons à 3 dimensions. Les moyennes de chacune de ses dimensions sont : 7.0431, 7.0143 et 7.0667. La moyenne de ces trois moyennes est : 7.0414.

C2 est une classe contenant 3000 échantillons à 3 dimensions. Les moyennes de chacune de ses dimensions sont : 7.9412, 4.0167 et -0.0712. La moyenne de ces trois moyennes est : 3.9622.

Nous disposons d'un grand nombre d'échantillons pour nos classes ce qui va aider à une bonne répartition de la classification mais certaines des moyennes de ces deux classes sont très proches l'une de l'autre ce qui va faire augmenter le taux d'erreur de la classification.

2) Avec une classification de max à posteriori, on obtient une moyenne d'erreur de 0 sur la classe C1 et de presque 100 % sur la classe C2, ce qui semble fortement incorrect.

3) Avec une analyse en composantes principales à 2 dimensions, on obtient entre 150 et 600 erreurs en moyennes sur C1 et de 1000 à 2500 erreurs sur la classe C2. Cela se rapproche beaucoup plus des résultats attendus d'une classification standard mais souffre tout de moins d'un fort taux d'erreurs sur C2. Cette classification est préférable à la précédente grâce à la réduction en dimensions.

Exercice 2 :

Exercice 3 :

Les échantillons réduit par ACI sont à 1 dimension. On obtient en moyenne entre 200 et 250 erreurs sur C1 et près de 1250 erreurs sur C2. Cette classification est beaucoup plus avantageuse que celle par ACP et celle sans réduction en dimensions.