

Manifold Learning

Lab x003 : Estimation de la dimension intrinsèque

Jairo Cugliari

S1 2018–2019

1 3-sensor dataset

Ce jeu de données est proposé dans la thèse de P. Desmartines (1994). Le but c'est de construire un jeu de données plongé sur un espace de dimension (extrinsèque) 10 mais de dimension intrinsèque égal à 3.

Pour ce faire nous disposons de 10 capteurs dans le cube unité $[-1,1]^3$. Alors, chaque point du cube unité peut être exprimé à partir des distances aux 10 capteurs, élevant la dimension du jeu de données à 10. Les variables ainsi créées ne sont pas indépendantes et intègrent des non linéarités.

1. A partir du code fourni, compléter la fonction `generateData` afin de créer le jeu de données décrit ci dessus.
2. Simulez 3 jeux de données de 100, 1000 et 10000 observations.

2 Estimateur global: ACP

A partir des données simulées, estimez la dimensionalité à l'aide de l'estimateur par ACP. Obtenir un graphique qui compare les variances normalisées en fonction de la dimension selon la taille des données.

3 Estimateur basé sur la corrélation

1. Écrire une fonction R qui calcule l'estimateur de la dimension intrinsèque basé sur la corrélation.
2. Utiliser les données 3-sensor set pour estimer la dimension