Manifold Proejct :

Introduction / Objectifs

La réduciton de dimensions est un enjeu majeur dans l’apprentissage automatique. Plusieurs étude cas l’ont montré : lorsque l’on souhaite inféré un model à partir d’un jeu de données, bien souvent la structure de ces dernières est noyé dans une dimensions plus grande. Nous parlons ainsi de dimension intrinsèque qui est la dimension réelle des jeux de données et de dimension extrinsèque : la dimension du jeu de donnée. L’étude des données en grande dimensions est un problème dans beaucoup d’études car la plupart des algorithme d’apprentissage perde en efficacité du fait des propriété particulières des données en frande dimensions. Pour toutes ces raisons des algorithme sont nés afin de réduire la dimension des jeux de données. Leur tâche consiste à trouver les dimensions intrinsèques en espérant que cette espace soit de suffisement faible dimmension afin que cela pallie le fléaut des dimensions.

L’objectif de notre travail est d’étudier ces différents algorithmes de réduction de dimensions, nous les mettrons donc en place et les testerons sur des jeux de données artificielles et réel. Les jeux de données artificielles seront simuler suivant différentes techniques. Pour chaque méthode l’objectif sera de les calibrer au mieux afin d’avoir la meilleur représentation possible de nos données. Enfin nous comparerons les performances des différents algorithmes sur ces jeux de données en proposant un critère de comparaison.

Matériel et méthode.

Dans cette partie nous présenterons les différentes méthodes utilisées ainsi que le moyen de les implémenter. Toutes les méthodes implémentées ci-dessous ont été programmer dans le Language R (version 4.1.2) en utisant l’IDE RStudio (version 2022.02.0).

La première méthode que nous implémenterons et une ACP à noyau. Proposé par … en … cette méthode est une extension de l’algorithme Analyse en Composante Principale

Explication de l’algortihme

Pour appliquer cette méthode nous avons décidé de d’implémenter l’agortihme directement en R et de faire notre propre fonction :

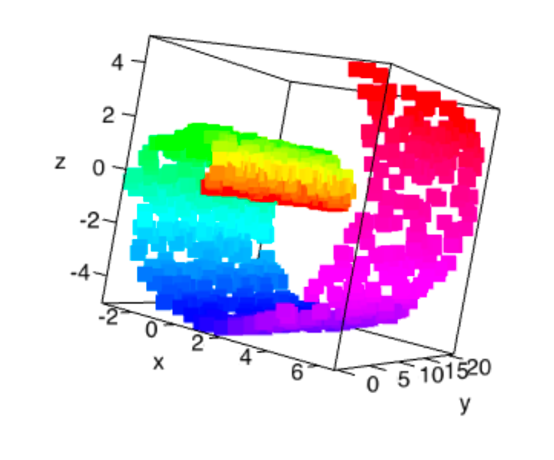
Pour calculer le noyeau Gaussien des données nous avons utiliser la méthode « gausskernel » proposée par le Package KRLS.

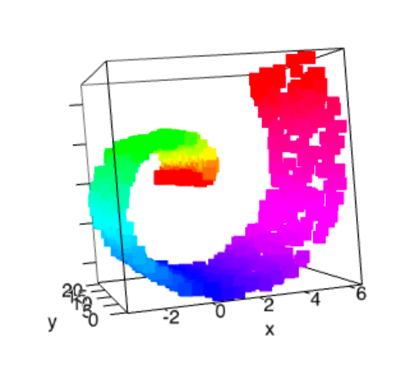
La seconde méthode mise en place est la Local

Jeu de données Simulée :

Pour comparer les différentes méthodes nous avons simulée 4 jeux de données ayant chacun ces spécificités. De manière générale nous les avons choisis dans le but de mettre en évidences les différents points fort et points faibles des méthodes présentées précédemment.

Le premier jeux de données est assez répendu lorsqu’il sagiit de l’étude des manifold. Il s’agit du swisrool. Représenté comme une feuille de papier que nous aurions enroulé sur elle-même Voir Image X :

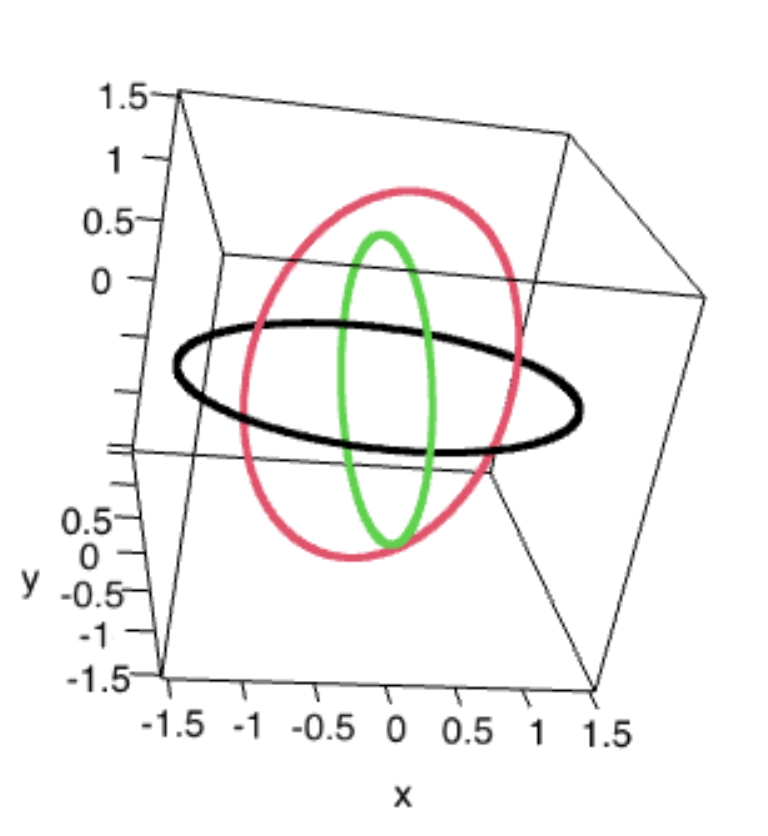
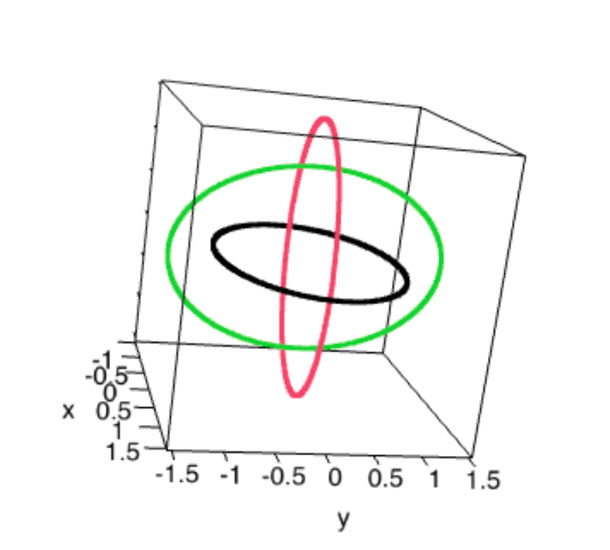




Cette structure enroulée pose de réel problème au méthodes linéaires. L’objectif ici est que l’algorithme « déroule » la structure.

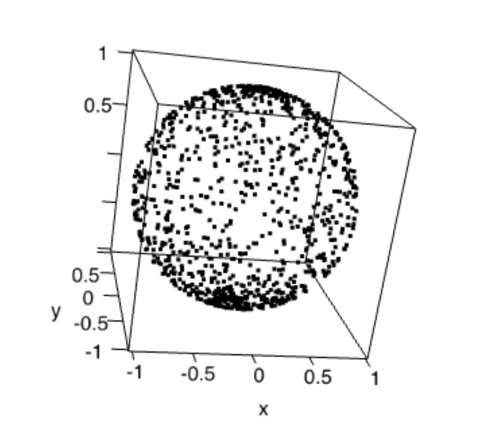
Pour simuler ces données nous avons implémenter une fonction qui reprenait l’équation implémentée dans la librairie Scikit Learn pour sa fonction sklearn.datasets.make\_swiss\_roll.

Le second jeux de données est une structure nommé les anneaux de Borommée, utilisé par différents artistes au cours du temps ils sont souvent utilisé en temps que symbole. Ici nous générons 3 cercles qui s’entrecroisent. Nous espérons que la méthode de dimensions consiste qu’ils sagissent de trois sous structures. Ici l’anneau rouge entoure le anneau vert qui lui-même entoure le anneau noir.

Nous avons directemps implémenter l’équations paramétriques de ces anneaux.

Le troisième jeu de données est un jeu de donnée simuler selon la proposition de … dans sa thèse … . L’objectif de cette méthode est construire une variété dans un espace de dimension faible, dans notre étude nous avons choisi une sphère unitaire de dimension 3. Cette structure va être plongé dans un espace de domension supérieur



Nous espérons pour ce jeu de donnée que la méthode de réductiond e dimensions parviendra à retrouver la dimension original de la variété dans notre cas cette sphère de dimension 3.