

Projet : Étude comparative des méthodes de réduction de la dimension Manifold Learning - M2 MALIA - UL2

Jairo Cugliari

S1 2024–2025

Dans ce projet, vous allez mener une étude numérique approfondie sur la performance des techniques de réduction de dimension vues en cours. Vous appliquerez ces méthodes à la fois sur des jeux de données artificiels et sur un jeu de données réel. Le travail s'articulera autour de 3 axes principaux : simulation, estimation et comparaison.

Simulation. Vous proposerez trois ou quatre jeux de données artificiels que vous obtiendrez par simulation. Ces jeux de données devront être conçus de manière à bien illustrer les différences entre les méthodes que vous allez comparer. Précisez les paramètres utilisés pour générer les données.

Apprentissage (a.k.a. estimation). Vous sélectionnerez trois ou quatre techniques de réduction de la dimension parmi celles vues en cours (par exemple : ACP, MDS, ISOMAP, LLE, t-SNE). Votre choix doit être justifié en fonction de la nature des données. Vous pouvez aussi explorer d'autres techniques pour peu que vous soyez certains de comprendre les techniques utilisées (autrement dit, vous devrez être capable de l'expliquer à un collègue, utiliser du code existant ne veut pas dire comprendre la technique). Cette partie comportera deux étapes :

- La calibration des méthodes, c'est-à-dire le réglage des paramètres comme la dimension intrinsèque, le nombre de voisins, etc.
- L'application des méthodes calibrées sur les jeux de données simulés afin de projeter ces données dans un espace de dimension réduite, de préférence euclidien.

Il sera particulièrement apprécié le soin que vous prêterez à la méthode de calibration. Vous pouvez proposer une nouvelle approche et la mettre en pratique.

Comparaison. Vous devrez définir un ou plusieurs critères de comparaison (par exemple, la préservation des distances, le coût de calcul, etc.) pour évaluer la performance des techniques choisies sur chaque jeu de données. Vous illustrerez vos résultats à l'aide de visualisations appropriées (graphiques, cartes de chaleur, etc.). Pour terminer, vous appliquerez une technique de réduction de dimension sur un jeu de données réel, que vous choisirez en fonction de son intérêt pour la problématique étudiée (par exemple, un jeu de données issu de l'apprentissage automatique, de la biologie, ou du traitement d'images).

Livrables

Le rapport devra être présenté sous la forme d'un article scientifique de 8 pages à deux colonnes. L'utilisation de Latex est recommandé mais pas obligatoire. Vous devrez y détailler chaque étape de votre travail de manière à ce qu'il soit reproductible. Vous devrez déposer sur Moodle le code et les données utilisées pour réaliser vos expériences numériques.

A titre indicatif, vous trouverez ci-dessous un plan de rédaction de l'article. Vous pouvez le modifier *a piacere*.

Plan proposé pour l'article :

1. **Introduction et objectifs** : Introduction du contexte, présentation des méthodes choisies et des objectifs du projet.
2. **Matériels et méthodes** : Description succincte des techniques de réduction de dimension et des critères de comparaison utilisés.
3. **Données** : Justification du choix des jeux de données artificiels et réels, avec description précise de ces derniers.
4. **Expériences numériques** : Explication du processus de simulation, et détails des expériences menées sur les données artificielles, avec un accent sur la reproductibilité (inclus les paramètres et les algorithmes utilisés).
5. **Comparaison des méthodes** : Analyse et comparaison des résultats obtenus avec les différentes techniques, avec visualisation des résultats et interprétation.
6. **Application sur données réelles** : Application d'une méthode de réduction de dimension sur les données réelles et discussion des résultats.
7. **Conclusion** : Synthèse des résultats, perspectives et limitations de l'étude.

Aspects pratiques

- Travail en équipes par binômes.
- Date de rendu : cf Moodle. Tout retard dans le rendu sera pénalisé.
- Le rendu inclura un rapport de 8 pages au format article (deux colonnes) ainsi que le code nécessaire pour reproduire toutes les expériences numériques, accompagné des graphiques et visualisations. Le code devra être propre, bien commenté et structuré.
- Dépôt des fichiers sur la plateforme Moodle, sous la forme d'une archive compressée (.zip ou .tar.gz) contenant le rapport au format PDF, le code source et les données.

Conseils supplémentaires

- Soignez vos visualisations : les graphiques doivent être lisibles et pertinents pour l'analyse (choisissez des couleurs et échelles adéquates, ajoutez des légendes et des titres clairs).
- Préparez-vous à justifier vos choix méthodologiques, tant pour les simulations que pour la sélection des techniques de réduction de dimension.
- Prenez soin de bien rédiger votre rapport. Un bon rapport doit être clair, concis, structuré et sans fautes de grammaire ou d'orthographe.