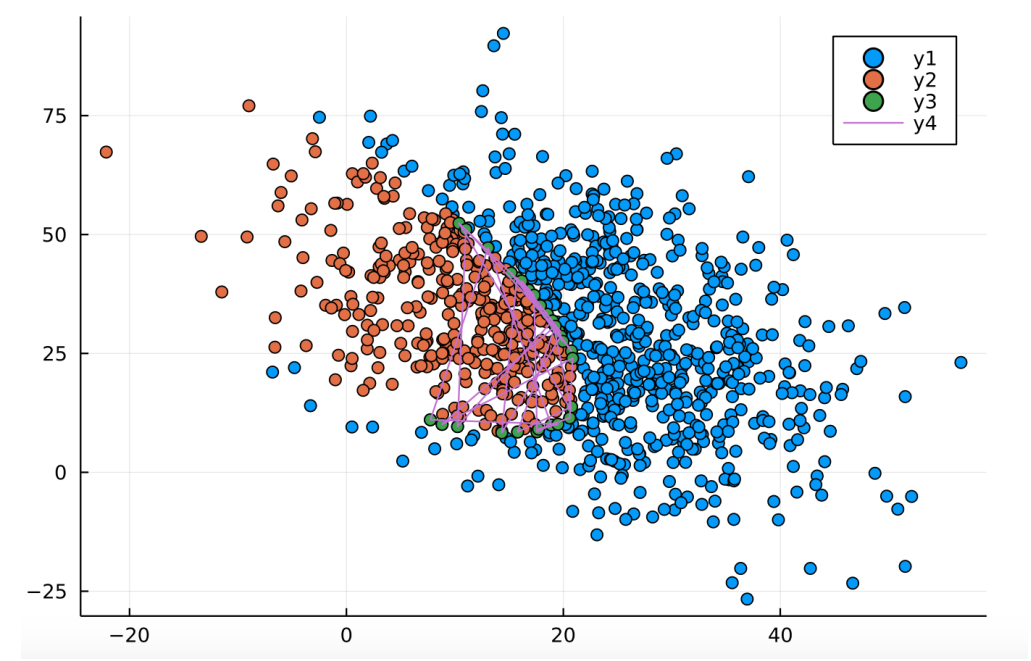
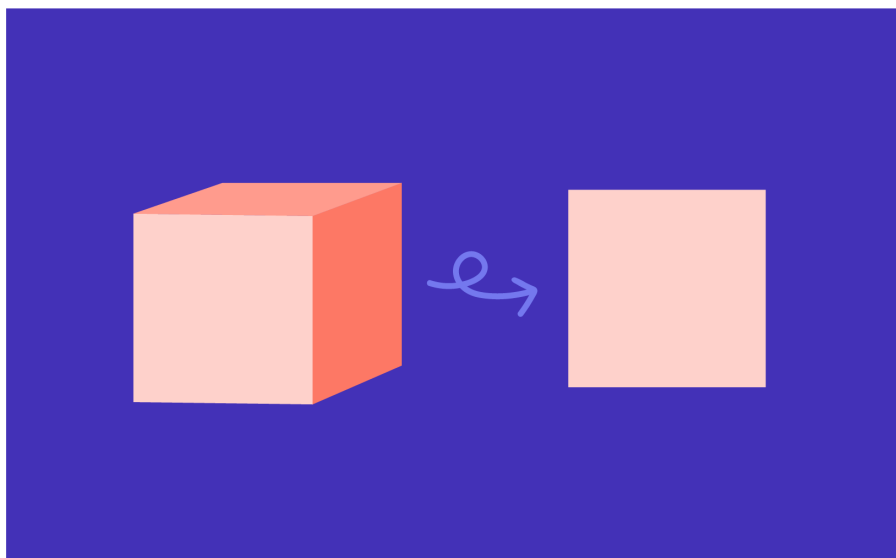


Comment évaluer la qualité d'une projection de données ?



Présentation du sujet

La projection de données est une façon de représenter des jeux de données sous la forme de nuages de points en 2 ou 3 dimensions. Lorsqu'un jeu de données à n-dimensions est projeté il doit passer par un processus de réduction de dimension ce qui cause des distorsions spatiales (autrement appelées "artefacts")



Notre objectif :

Mettre en place un logiciel qui permet d'évaluer la qualité d'une projection et de trouver la méthode la plus performante

Notre démarche :

- Recherche documentaire pour appréhender le sujet
- Lister les différentes méthodes de projections (PCA, MDS, t-SNE, ClassNeRV...)
- Lister les différents critères d'évaluation des méthodes (clustering, outlying...)
- Génération de jeux de données test
- Implémentation du programme en Python
- Test

	Czekanowski	Folkes	Hubert	Jaccard	Kulczynski	McNemar	Phi	Precision	Rand	Recall	Rogers Tanimoto	Russel Rao	SokalS1	SokalS2	Total score
MDS dataset1	0.12	0.16	-0.0	0.06	0.2	306.01	0.0	0.33	0.64	0.07	0.47	0.02	0.03	0.78	0/14
tSNE dataset1	0.02	0.06	0.0	0.01	0.17	394.03	0.0	0.34	0.66	0.01	0.5	0.0	0.01	0.8	5/14
pca dataset1	0.31	0.31	-0.0	0.18	0.31	55.63	-0.0	0.33	0.57	0.28	0.4	0.1	0.1	0.73	7/14
mds dataset2	0.08	0.12	-0.0	0.04	0.19	344.52	0.0	0.33	0.65	0.04	0.48	0.01	0.02	0.79	0/14
tSNE dataset2	0.02	0.06	0.0	0.01	0.17	393.52	0.0	0.34	0.66	0.01	0.5	0.0	0.0	0.8	5/14
pca dataset2	0.13	0.17	0.0	0.07	0.21	293.66	0.0	0.33	0.64	0.08	0.47	0.03	0.04	0.78	7/14
mds dataset3	0.04	0.08	-0.0	0.02	0.18	377.4	0.0	0.33	0.66	0.02	0.49	0.01	0.01	0.8	3/14
tSNE dataset3	0.02	0.06	0.0	0.01	0.18	391.69	0.0	0.35	0.66	0.01	0.5	0.0	0.01	0.8	5/14
pca dataset3	0.06	0.1	-0.0	0.03	0.18	362.47	-0.0	0.33	0.66	0.03	0.49	0.01	0.02	0.79	7/14

Résultats :

- L'ACP semble être la plus performante sur nos jeux de données de test
- Résultats à nuancer : il faudrait tester avec une plus grande variété de jeux de données
- Étudier plus en détail une méthode de projection prometteuse: ClassNeRV