

STAI: Geometric Learning Assignment 3

**Matei Bejan, Andrei Manolache, Mihaela
Nistor**



UNIVERSITATEA DIN
BUCUREȘTI
— VIRTUTE ET SAPIENTIA —

Table of contents

1 Notebook 1: Learning Tangent Data

2 Notebook 2: Fréchet Mean and Tangent PCA

3 Notebook 3: K-Means clustering on a Riemannian Manifold



Notebook 1: Learning Tangent Data

- Am modificat logica de evaluare, iar acum vom cauta hiperparametri cu Cross-Validare pentru fiecare metrica in cazul in care lucram pe varietate, si pentru cazul in care lucram in spatiul Euclidian.
- Am introdus inca doua modele: SVM si Random Forest.
- Obiectivul pentru CV e schimbat in ROC-AUC, care e mai robust la dezechilibrul intre clase.
- Rezultatele sunt mult mai apropiate dupa ce cautam hiperparametri, iar in unele cazuri sunt chiar mai bune decat pe spatiul Euclidian!



Notebook 1: Learning Tangent Data

- Implementam kernel-ul de la <https://ieeexplore.ieee.org/document/6247965>.
- Kernelul ne ajuta sa proiectam datele pe varietatea matricelor SPD si sa optimizam modelul pe spatiul tangent, tinand cont de metrica Log-Euclidiană.
- Operatiile ar trebui sa fie echivalente cu folosirea unui kernel liniar impreuna cu proiectia pe spatiul tangent si folosirea metricii LE, dar in practica, SVM-ul ce foloseste kernel-ul implementat are rezultate mai bune.
- O observatie interesanta e faptul ca daca proiectam datele de pe varietate pe spatiul tangent si folosim un kernel rbf care proiecteaza la randul lui datele intr-un spatiu infinit dimensional, rezultatele sunt considerabil mai bune.



Notebook 1: Learning Tangent Data

SVC	
Metric	CV Score
AI	0.809%
LE	0.812%
Euclidean	0.815%

Random Forest	
Metric	CV Score
AI	0.808%
LE	0.827%
Euclidean	0.799%



Notebook 1: Learning Tangent Data

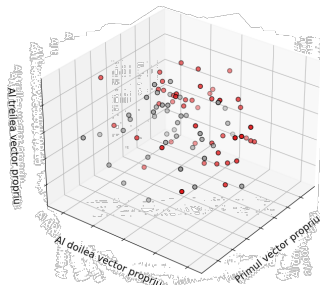
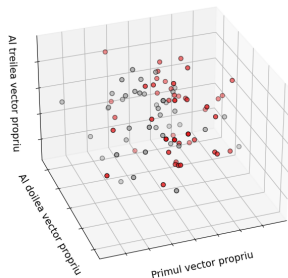
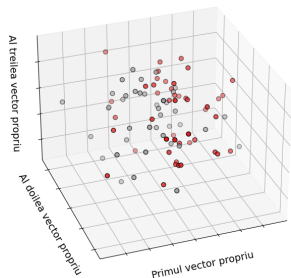
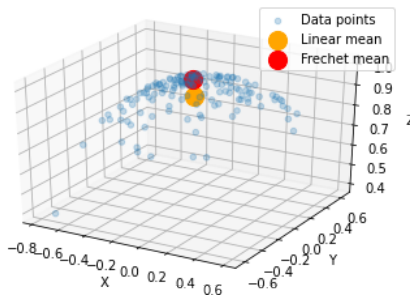


Figure: Folosim KernelPCA pentru a folosi kernel-ul definit anterior cu metoda PCA. Rezultatele cu KernelPCA, Proiectie, si TangentPCA sunt identice, dupa cum ne-am si asteptat.



Notebook 2: Fréchet Mean and Tangent PCA

- Notebook-ul 2 se concentreaza pe vizualizarea datelor in geomstats, cu precadere a decompozitiei datelor pe hipersfera.
- Am plicat decompozitia TangentPCA cu linear mean din geomstats, pe langa rezultatele deja existente cu Fréchet mean.



Notebook 2: Fréchet Mean and Tangent PCA

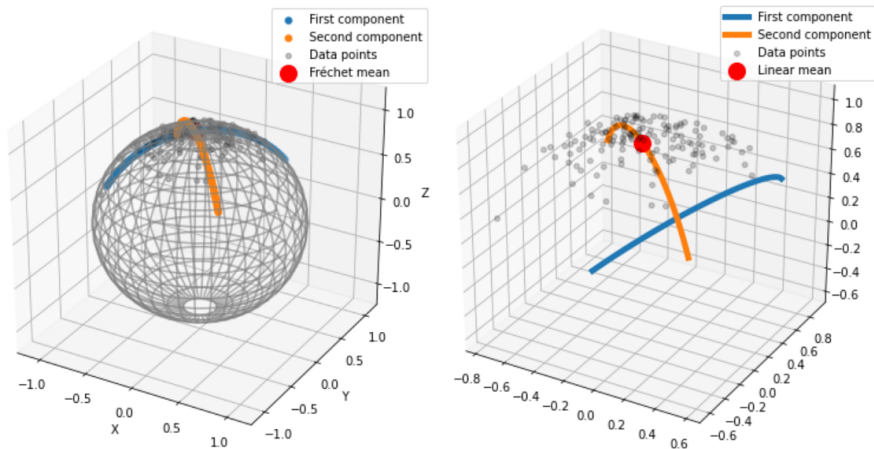


Figure: Fréchet and linear mean decompositions in 3D.



Notebook 2: Fréchet Mean and Tangent PCA

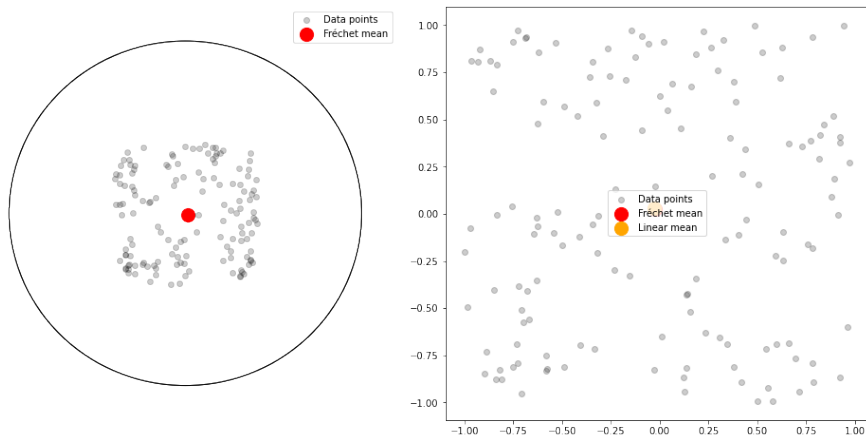


Figure: Data distribution in the (Y, Z) space and Fréchet and linear mean superposition.



Notebook 2: Fréchet Mean and Tangent PCA

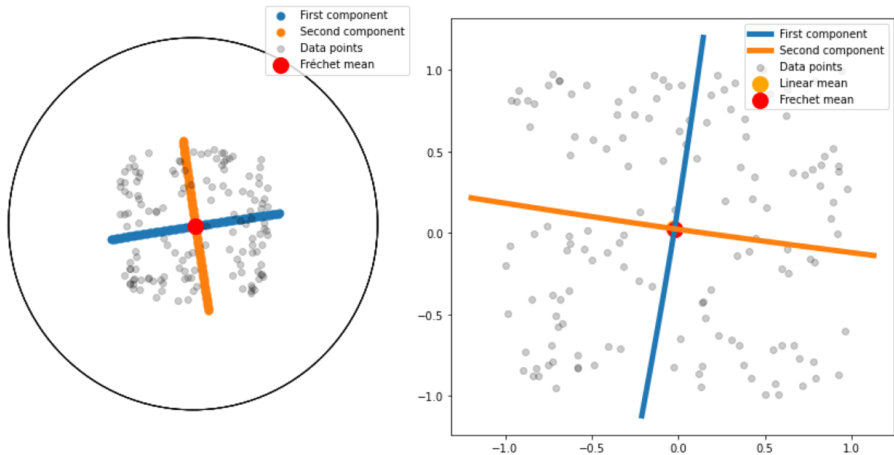


Figure: Fréchet and linear mean decompositions' projections in the (Y, Z) space.



Notebook 3: K-Means clustering on a Riemannian Manifold

- Notebook-ul 3 prezinta modul de folosire al algoritmului Kmeans din Geomstats pe 2 clustere de puncte generate pe o hipersfera.
- Desi algoritmul este unul nesupervizat, am folosit label-urile punctelor pentru a compara rezultatele dintre mai multi algoritmi de clusterizare din Geomstats cu cei din Sklearn, mutand punctele in spatiul tangent.

Method	Accuracy
Kmeans Geomstats	92.42%
Kmeans Sklearn	92.42%
KMedoids Geomstats	90.42%
KMedoids Sklearn	90.90%



Notebook 3: K-Means clustering on a Riemannian Manifold

Agglomerative Hierarchical Clustering Geomstats			
Euclidean	L2	L1	Cosine
49.5%	49.5%	84.5%	89.0%

Agglomerative Hierarchical Clustering Sklearn			
Euclidean	L2	L1	Cosine
83.5%	49.5%	88.0%	87.0%



Notebook 3: K-Means clustering on a Riemannian Manifold

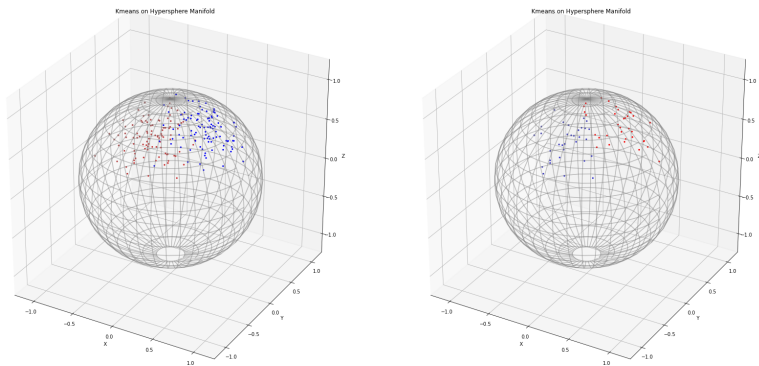


Figure: Geomstats K-Means and K-Medoids.



Notebook 3: K-Means clustering on a Riemannian Manifold

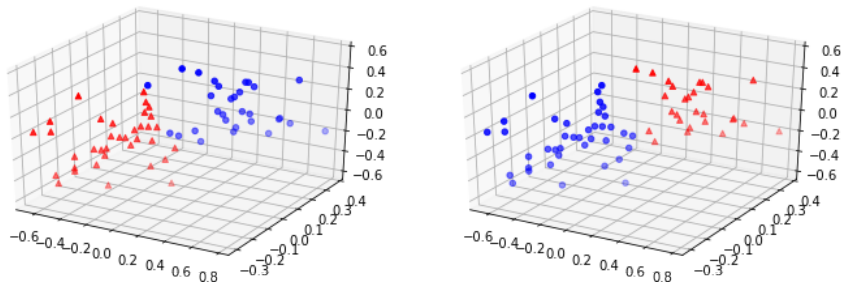


Figure: Sklearn K-Means and K-Medoids.



Notebook 3: K-Means clustering on a Riemannian Manifold

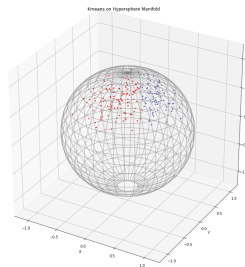
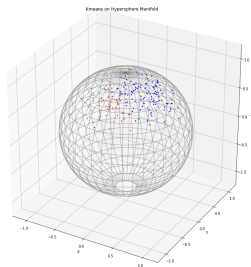
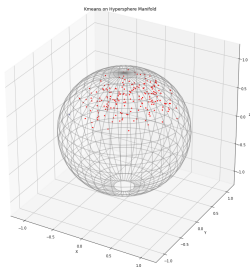
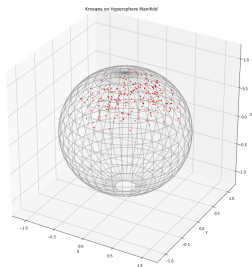


Figure: Agglomerative Hierarchical Clustering Geomstats results: Euclidean, L2, L1, Cosine



Notebook 3: K-Means clustering on a Riemannian Manifold

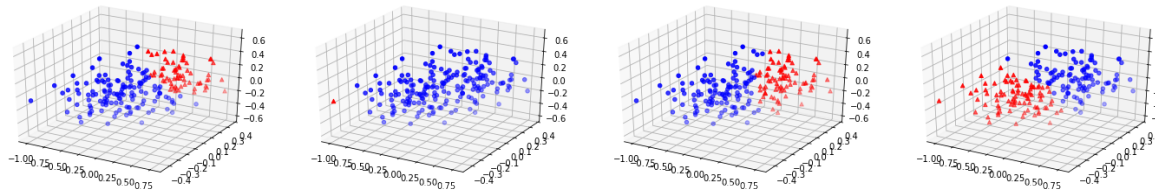


Figure: Agglomerative Hierarchical Clustering Sklearn results: Euclidean, L2, L1, Cosine.

