Dimension reduction and manifold learning

ECTS : 4

Volume horaire : 24

Description du contenu de l'enseignement :

Modern machine learning typically deals with high-dimensional data. The fields concerned are very varied and include genomics, image, text, time series, or even socioeconomic data where more and more unstructured features are routinely collected. As a counterpart of this tendency towards exhaustiveness, understanding these data raises challenges in terms of computational resources and human understandability. Manifold Learning refers to a family of methods aiming at reducing the dimension of data while preserving certain of its geometric and structural characteristics. It is widely used in machine learning and experimental science to compress, visualize and interpret high-dimensional data. This course will provide a global overview of the methodology of the field, while focusing on the mathematical aspects underlying the techniques used in practice.

Compétence à acquérir :

- Curse of dimensionality, manifold hypothesis and intrinsic dimension(s)
- Multidimensional scaling
- Linear dimension reduction (random projections, principal component analysis)
- Non-linear spectral methods (kernel PCA, ISOMAP, MVU, Laplacian eigenmaps)
- Ad-hoc distance-preserving methods (diffusion maps, LLE)
- Probabilistic dimension reduction and clustering (SNE, UMAP)
- Neural network-based dimensionality reduction

Bibliographie, lectures recommandées :

- Ghojogh, B., M. Crowley, F. Karray, and A. Ghodsi (2023). Elements of dimensionality reduction and manifold learning
- Lee, J. A., M. Verleysen, et al. (2007). Nonlinear dimensionality reduction