DIMENSION REDUCTION & MANIFOLD LEARNING

Examination outline

When/Where: Wednesday, March 11th morning and afternoon / rue de l'Estrapade
Format: Oral presentation, and turn in the "Reading report" sheet available here: (link)
Support: Slides only.
Slides + Reading report (pdf only) must be sent by email at eddie.aamari@ens.fr the day before the presentation.
Duration: 15 minutes for the presentation, 5 minutes for questions.
What: Reading, summarizing, and criticizing your chosen paper.
Include some simulation / experimental part
Choose one (or a few) point to elaborate/develop.
Show maths proofs if you think that it is relevant
Make sure to work in groups
Make sure to rehearse before the exam: 15 minutes is quite short, you must compress the information and talk only about what is necessary
You should pretend that your audience knows everything about the course material, but nothing about your specific article: introduce and explain everything necessary

Instructions for Choosing the Article

- Vote **yes** individually for only one slot in the following poll:
  https://beta.framadate.org/polls/0b64cc1e68f39f743d46
- **Note**:
  - A vote for more than one article will be deemed invalid and subsequently deleted.
  - Assignment will follow the "first-come, first-served" rule.
  - Contact me in advance for any questions about the paper you are reading.

List of articles

1. **Non-linear manifold learning in fMRI uncovers a low-dimensional space of brain dynamics**
   Gao, S., Mishne, G. & Scheinost, D. (link)
   Keywords: fMRI, brain dynamics, nonlinear manifold learning, low-dimensional latent space, neuroscience.

2. **Minimax Estimation of Distances on a Surface and Minimax Manifold Learning in the Isometric-to-Convex Setting**
   Arias-Castro, E., & Alain Chau, P. (link)
   Keywords: minimax rates, geodesic distance estimation, surfaces/manifolds, isometry, statistical guarantees.

3. **On the Convergence of Maximum Variance Unfolding**
   Arias-Castro, E., & Pelletier, B. (link)
   Keywords: MVU, semidefinite programming, consistency, manifold learning, convergence analysis.

4. **Infinite multidimensional scaling for metric measure spaces**
   Kroshnin, A., Stepanov, E., & Trevisan, D. (link)
   Keywords: MDS, metric measure spaces, infinite-dimensional embeddings, stability, probabilistic geometry.

5. **Confidence Sets for Multidimensional Scaling**
   Vishwanath, S. & Arias-Castro, E. (link)
   Keywords: uncertainty quantification, confidence sets, MDS, inferential guarantees, visualization reliability.

6. **A kernel-based analysis of Laplacian Eigenmaps**
   Wahl, M. (link)
   Keywords: Laplacian Eigenmaps, kernel methods, spectral convergence, graph Laplacian, manifold approximation.

7. **On the Selection of Tuning Parameters for Patch-Stitching Embedding Methods**
   Chau, P.A. & Arias-Castro, E. (link)
   Keywords: patch-based embeddings, parameter selection, local-to-global stitching, stability, manifold learning practice.

8. **On Sparsity and Sub-Gaussianity in the Johnson-Lindenstrauss Lemma**
   Garivier, A. & Pilliat, E. (link)
   Keywords: Johnson–Lindenstrauss, sparse projections, sub-Gaussian tails, dimension reduction, concentration inequalities.

9. **A theory of stratification learning**
   Aamari, E. & Berenfeld, C. (link)
   Keywords: stratified spaces, singular data, learning theory, geometric inference, guarantees/consistency.

10. **Statistical exploration of the Manifold Hypothesis**
    Whiteley, N., Grey, A. & Rubin-Delanchy, Patrick (link)
    Keywords: manifold hypothesis, statistical testing/diagnostics, model misspecification, geometry of data, empirical study.

11. **Estimating the intrinsic dimension of datasets by a minimal neighborhood information**
    Facco, E., d'Errico, M., Rodriguez, A., & Laio, A. (link)
    Keywords: intrinsic dimension, kNN statistics, local neighborhoods, scaling laws, practical estimator.

12. **Beyond the noise: intrinsic dim. estimation with optimal neighbourhood identification**
    Di Noia, A., Macocco, I., Glielmo, A., Laio, A., & Mira, A. (link)
    Keywords: intrinsic dimension under noise, neighborhood selection, robustness, bias–variance tradeoff, adaptive methods.

13. **Clustering with t-SNE, provably**
Linderman, G.C. & Steinerberger, S. (link)
Keywords: t-SNE theory, provable clustering, separation conditions, visualization guarantees, embedding behavior.

14. **Explaining t-SNE Embeddings Locally by Adapting LIME**
Bibal, A., Vu, V.M., Nanfack, G. & Frénay, B. (link)
Keywords: explainable embeddings, local surrogate models, LIME, interpretability, t-SNE diagnostics.

15. **TriMap: Large-scale Dimensionality Reduction Using Triplets**
Amid, E. & Warmuth, M. K. (link)
Keywords: triplet constraints, global structure, scalable visualization, neighbor preservation, dimensionality reduction.

16. **PHATE: A Dim. Reduc. Method for Visualizing Trajectories & Transitions**
Moon, K. R. *et al.* (link)
Keywords: trajectory inference, diffusion geometry, transitions, denoising, single-cell visualization.

17. **Assessing single-cell transcriptomic variability through density-preserving data visualization (den-SNE / densMAP)**
Narayan, A. *et al.* (link)
Keywords: density preservation, UMAP/densMAP, single-cell RNA-seq, variability, visualization fidelity.

18. **Understanding How Dimension Reduction Tools Work: An Empirical Approach to Deciphering t-SNE, UMAP, TriMap, and PaCMAP for Data Visualization**
Wang, Y., Huang, H., Rudin, C., & Shaposhnik, Y. (link)
Keywords: empirical diagnostics, t-SNE/UMAP/TriMap/PaCMAP, failure modes, interpretability, practical guidance.

19. **Towards a comprehensive evaluation of dimension reduction methods for transcriptomic data visualization**
Huang, H. *et al.* (link)
Keywords: benchmarking, transcriptomics, visualization metrics, stability, method comparison.

20. **Isumap: Manifold Learning and Data Visualization leveraging Vietoris-Rips filtrations**
Barth, L. S., Joharinad, P., Jost, J., & Keck, J. (link)
Keywords: Vietoris–Rips filtration, persistent homology, topology-aware embedding, geodesic structure, TDA.

21. **Deep generative modeling for single-cell transcriptomics (scVI)**
Lopez, R. *et al.* (link)
Keywords: variational inference, latent variable models, single-cell RNA-seq, batch correction, generative modeling.

22. **Latent Space Oddity: On the Curvature of Deep Generative Models**
Arvanitidis, G., Hansen, L. K., & Hauberg, S. (link)
Keywords: latent space geometry, Riemannian metrics, curvature, geodesics, VAEs/generative models.

23. **Unpicking Data at the Seams: Understanding Disentanglement in VAEs**
Allen, C. (link)
Keywords: disentanglement, VAEs, identifiability, representation learning, evaluation/limitations.