Unsupervised Learning HW4

Due: Tue Dec 06, 2022 at 11:59pm

All homeworks (including this one) should be typesetted properly in pdf format. Late homeworks or handwritten solutions will not be accepted. You must include your name and UNI in your homework submission. To receive credit, a typesetted copy of the homework pdf must be uploaded to Gradescope by the due date. You must show your work to receive full credit. Discussing possible solutions for homework questions is encouraged on piazza and with your peers, but everyone must write their own individual solutions. You must cite all external references you used (including the names of individuals you discussed the solutions with) to complete the homework.

1 Readings

Pick and read any four papers of your choice from the following list. Summarize the main results of your assigned paper, discuss their significance and provide a short proof sketch of their technical results.

List of papers:

Emebddings / Dimensionality Reduction

- Fakcharoenphol, Rao and Talwar. A tight bound on approximating arbitrary metrics by tree metrics.
- Brinkman and Charikar. On the Impossibility of Dimension Reduction in L1.
- Larsen and Nelson. Optimality of the Johnson-Lindenstrauss lemma.
- Verma. A note on random projections for preserving paths on a manifold.
- Verma. Distance preserving embeddings for general n-dimensional manifolds.
- Clarkson. Tighter bounds for random projections of manifolds.
- Indyk and Naor. Nearest neighbor preserving embeddings
- Arora, Hu, Kothari. An Analysis of the t-SNE Algorithm for Data Visualization.
- Arias-Castro. Some theory for ordinal embedding.

Density Estimation

- Dasgupta and Schulman. A two-round variant of EM for Gaussian mixtures.
- Dasgupta. Learning mixtures of Gaussians.

- Dasgupta and Kpotufe. Optimal rates for k-NN density and mode estimation.
- Arora, Ge, Halpern, Mimno, Moitra, Sontag, Wu, Zhu. A Practical Algorithm for Topic Modeling with Provable Guarantees.
- Arora, Ge, Liang, Ma, Zhang. Generalization and Equilibrium in Generative Adversarial Nets
- Arora, Risteski, Zhang. Do GANs learn the distribution? Some theory and empirics

Structure discovery

- Diaconis, Goel, Holmes. Horseshoes in multidimensional scaling and local kernel methods.
- Fefferman, Mitter, Narayanan. Testing the Manifold Hypothesis.
- Chazal and Michel. An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists.

Misc. Topics

- Dasgupta and Sinha. Randomized partition trees for nearest neighbor search.
- Chaudhuri and Dasgupta. Rates of convergence for the cluster tree.
- Dasgupta and Freund. Random projection trees and low dimensional manifolds.
- Dasgupta, Hsu, Verma. A concentration theorem for projections.
- Clarkson. Nearest-neighbor searching and metric space dimensions.
- Niyogi, Smale, Weinberger. Finding the Homology of Submanifolds with High Confidence from Random Samples.
- Niyogi, Smale, Weinberger. A Topological View of Unsupervised Learning and Clustering.
- Gionis, Indyk, Motwani. Similarity search in high dimensions via hashing.
- Indyk and Motwani. Approximate nearest neighbors: towards removing the curse of dimensionality.