

COS General Exam - Andy Jones

May 10, 2021, 1-3pm

Committee: Barbara Engelhardt (adviser), Ryan Adams, Ben Raphael

Abstract

Dimension reduction is useful for exploratory data analysis. In many applications, it is of interest to discover variation that is enriched in a "foreground" dataset relative to a "background" dataset. In this talk, I propose two new models for this "contrastive learning" task. First, I propose probabilistic contrastive principal component analysis (PCPCA), a model-based alternative to contrastive PCA. I discuss several of PCPCA's advantages, including greater interpretability, uncertainty quantification, robustness to noise and missing data, and the ability to generate data from the model. Second, I propose the contrastive Poisson latent variable model (CPLVM). The CPLVM is an extension of PCPCA that is designed to model count-based data, such as single-cell RNA-seq data. Using this model, I propose a hypothesis testing framework that is able to detect global contrastive changes across features, as well as changes specific to subsets of features. I demonstrate the performance of these two models through a series of simulations and experiments with datasets of gene expression, protein expression, and images.

Textbook

[Pattern Recognition and Machine Learning](#). Christopher Bishop.

Papers

1. [The EM Algorithm for Mixtures of Factor Analyzers](#). Zoubin Ghahramani, Geoffrey E. Hinton. Technical report (1996).
2. [EM Algorithms for PCA and SPCA](#). Sam Roweis. NeurIPS (1998).
3. [Learning the parts of objects by non-negative matrix factorization](#). Daniel D. Lee, H. Sebastian Seung. Nature (1999).
4. [A Probabilistic Interpretation of Canonical Correlation Analysis](#). Francis R. Bach, Michael I. Jordan. Technical report (2005).
5. [Stochastic Variational Inference](#). Matthew D. Hoffman, David M. Blei, Chong Wang, John Paisley. JMLR (2013).
6. [Contrastive Learning Using Spectral Methods](#). James Zou, Daniel Hsu, David Parkes, Ryan Adams. NeurIPS (2013).
7. [Linear Dimensionality Reduction: Survey, Insights, and Generalizations](#). John P. Cunningham, Zoubin Ghahramani. JMLR (2015).
8. [Exploring patterns enriched in a dataset with contrastive principal component analysis](#). Abubakar Abid, Martin J. Zhang, Vivek K. Bagaria, James Zou. Nature Communications (2018).

9. [Deep generative modeling for single-cell transcriptomics](#). Romain Lopez, Jeffrey Regier, Michael B. Cole, Michael I. Jordan, Nir Yosef. Nature Methods (2018).
10. [Unsupervised learning with contrastive latent variable models](#). Kristen A. Severson, Soumya Ghosh, Kenney Ng. AAAI (2019).