Lecture 1. Introduction and Syllabus

Yuan Yao

Hong Kong University of Science and Technology

January 30, 2019

Outline

Geometric Data Analysis

Topological Data Analysis

Planned Schedule

Part I. Geometric Data Analysis

- A duality in linear dimensionality reduction
  - Principal Component Analysis (PCA)
  - Multidimensional Scaling (MDS)
  - Random matrix theory and phase transitions
  - Random projection and restricted isometry property
- Extended PCA/MDS via SDP
  - Robust PCA
  - Sparse PCA
  - Graph Realization or Sensor Network Localization
- Manifold Learning: nonlinear dimensionality reduction via spectral method on graphs
  - Locally Linear Embedding (PCA+), Isomap (MDS+)
  - Laplacian LLE, Diffusion Map, LTSA

Part I. Geometric Data Analysis (continued)

- Supervised PCA
  - Ridge Regression and PCA
  - Slice Inverse Regression and Linear Discriminant Analysis
- *Other topics in representation learning
  - tSNE
  - Steerable PCA
  - Dictionary learning and Matrix Factorization
  - Deep learning

Outline

Geometric Data Analysis

Topological Data Analysis

Planned Schedule

Part II. Topological Data Analysis

- Clustering method (0-homology)
  - k-center
  - k-means
  - hierarchical linakge
- Topological Data Analysis and Morse Theory
  - Reeb graph and mapper
  - Persistent homology and discrete Morse theory
  - *Critical nodes and graphs
- *Euler Calculus and signal processing

Part I. Topological Data Analysis (continued)

- Hodge Theory: a bridge connecting geometry and topology
  - Spectral clustering and graph Laplacian
  - Statistical ranking and graph Helmoholtzian/Hodge Laplacian
    - Experimental design and random graph theory
    - Online ranking and stochastic algorithms
    - Budget control and information maximization
    - Individual learning vs. social choice theory
  - Game theory
    - Finite game flow and combinatorial Hodge Theory
    - Differentiable games (GANs), stochastic games and Markov decision process (reinforcement learning)

Outline

Geometric Data Analysis

Topological Data Analysis

Planned Schedule

Planned Schedule

- The course runs for about 13 weeks.
- Week 1:
  - Jan 30: Introduction
  - Feb 1: seminar by Ruohan ZHAN (Stanford University) with title "Safety masked reinforcement learning"
- Week 2: spring festival break (Feb 8 will be rescheduled to later)
- Week 3:
  - Feb 13: PCA
  - Feb 15: MDS

Planned Schedule (continued)

- Feb 20 - May 8: to-be-announced on courseweb
  - `https://yao-lab.github.io/2019_csic5011/`
- Occasionally invited speakers from academia or industry will present
- Discussions on piazza (by invitation only):
  - `https://piazza.com/ust.hk/spring2019/csic5011/home`