Lecture 1. Introduction and Syllabus

Yuan Yao

Hong Kong University of Science and Technology

February 18, 2020

Outline

Course Information

Part I: Geometric Data Analysis

Part II: Topological Data Analysis

Course Information

- The course runs for about 13 weeks, every Wednesday 3-6pm.
  - Courseweb
    - `https://yao-lab.github.io/2020.csic5011/`
  - Zoom webinar link:
    - `https://hkust.zoom.com.cn/j/237158553`
  - Occasionally invited speakers from academia or industry will present
  - Projects based evaluation, no final exam!

Outline

Course Information

Part I: Geometric Data Analysis

Part II: Topological Data Analysis

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
- Supervised PCA
  - Ridge Regression and PCA
  - Slice Inverse Regression and Linear Discriminant Analysis

Part I. Geometric Data Analysis (continued)

- Manifold Learning: nonlinear dimensionality reduction via spectral method on graphs
  - Locally Linear Embedding (PCA+), Isomap (MDS+)
  - Laplacian LLE, Diffusion Map, LTSA
- *Other topics in representation learning
  - tSNE
  - Steerable PCA
  - Dictionary learning and Matrix Factorization
  - Deep learning

Part I. Geometric Data Analysis (continued)

- Manifold Learning: nonlinear dimensionality reduction via spectral method on graphs
  - Locally Linear Embedding (PCA+), Isomap (MDS+)
  - Laplacian LLE, Diffusion Map, LTSA
- *Other topics in representation learning
  - tSNE
  - Steerable PCA
  - Dictionary learning and Matrix Factorization
  - Deep learning
- Summary: all these methods are based on spectral methods -
  - Can you hear the shape of drum? (by Hermann Weyl and Mark Kac)

Part I. Geometric Data Analysis (continued)

- Manifold Learning: nonlinear dimensionality reduction via spectral method on graphs
  - Locally Linear Embedding (PCA+), Isomap (MDS+)
  - Laplacian LLE, Diffusion Map, LTSA
- *Other topics in representation learning
  - tSNE
  - Steerable PCA
  - Dictionary learning and Matrix Factorization
  - Deep learning
- Summary: all these methods are based on spectral methods -
  - Can you hear the shape of drum? (by Hermann Weyl and Mark Kac)
  - Can you hear the shape of data? (in this course)

Outline

Course Information

Part I: Geometric Data Analysis

Part II: Topological Data Analysis

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)