TOPOLOGICAL DATA ANALYSIS
COURSE SYLLABUS

Instructor: Brian Hepler, Ph.D.

Course Description

This course introduces the core concepts of algebraic topology used in data analysis, with a particular focus on persistent homology and its computational aspects. You will learn how to build and analyze simplicial complexes, understand stability and statistical issues in persistence, and explore applications in machine learning (including geometric deep learning) and quantum computing. A hands-on component using Python libraries (e.g. GUDHI, ripser, scikit-tda) reinforces theoretical insights with practical projects.

Lecture Breakdown

Lecture 1: Intro to TDA and Motivations

- Overview of TDA: What it is an why it matters.
- Historical context and motivations: from shape analysis to data science.
- Comparison with traditional statistical methods.
- Examples of datasets where topology provides unique insights (e.g., sensor networks, biological data, complex systems).
- *Interactive Notebook:* Visual exploration of simple point clouds and clustering phenomena using Python (e.g., Matplotlib, GUDHI/Ripser).

Lecture 2: Basic Topology, Simplicial Complexes, and the Nerve Theorem

- Quick review of basic topological concepts: open sets, coverings, and continuity.
- Definition and properties of simplicial complexes.
- Construction techniques: Čech, Vietoris-Rips, and witness complexes.
- The nerve theorem and its role in connecting combinatorial complexes to the topology of underlying spaces.
- *Interactive Notebook:* Construct Čech and Vietoris–Rips complexes from synthetic datasets and visualize their evolution.

Lecture 3: Homology Theory and Persistent Homology

- Introduction to homology: chains, cycles, and boundaries.
- Betti numbers as topological invariants.
- Definition of persistence: filtrations, persistence modules, and barcodes.
- Intuition behind persistence diagrams and their interpretation.
- *Interactive Notebook:* Utilize libraries like Dionysus or GUDHI to compute persistence barcodes on point cloud data with adjustable parameters.

Lecture 4: Algorithms and Computational Aspects of TDA

- Algorithms for computing persistent homology (e.g.. matrix reduction, boundary operator computation).
- Discussion of algorithm complexity and practical implementations.
- Review of available software packages and libraries.
- Discussion on computational challenges and optimizations.
- *Interactive Notebook:* A step-by-step walkthrough of a persistent homology algorithm on a benchmark dataset, highlighting performance considerations.

Lecture 5: Stability, Metrics, and Statistical Inference in TDA

- The stability theorem for persistence diagrams.
- Metrics on diagrams: Wasserstein and bottleneck distances.
- Statistical tools and hypothesis testing within TDA.
- *Interactive Notebook:* Simulate noisy data to illustrate diagram stability and perform statistical comparisons.

Lecture 6: TDA in Machine Learning and Geometric Deep Learning

- Enhancing ML pipelines with TDA features.
- Integration with neural networks and geometric deep learning frameworks.
- Applications in image and signal processing.
- *Interactive Notebook:* Developing an end-to-end classification task where persistence diagram summaries serve as input features to a machine learning model (e.g., using scikit-learn or PyTorch).

Lecture 7: TDA Applications in Quantum Computing and Quantum Data Analysis

- Overview of topological approaches in quantum computing (e.g., topological quantum error correction, quantum state space analysis).
- Analyzing quantum state spaces and entanglement with TDA.
- Exploring topological signatures in quantum phase transitions.
- *Interactive Notebook:* Simulating simple quantum systems (using libraries like Qiskit) and applying TDA to study the structure of the resulting state spaces or measurement outcomes.

Lecture 8: Advanced Topics, Case Studies, and a Potential Capstone Project

- Advanced topics: Mapper algorithm, multi-parameter persistence, and emerging directions in TDA.
- In-depth case studies from industry and academia (e.g., neuroscience, material science, sensor networks).
- Course wrap-up: integrating theory and practice, discussion of current research frontiers.
- *Interactive Notebook:* A capstone project where students select a real-world dataset and perform an end-to-end TDA analysis—from constructing a complex to computing persistent homology, extracting features, and applying a learning algorithm. Visualizations using matplotlib or plotly will aid in interpretation.

Mathematical Prerequisites

Linear Algebra and Multivariate Calculus: Familiarity with vector spaces, eigendecomposition, and calculus.

Basic Topology and Geometry: Some exposure to topological concepts, though the course includes a review.

Algebraic Concepts: Basic knowledge of abstract algebra (groups, rings) is helpful.

References

- Carlsson, G. "Topology and Data"
- Edelsbrunner, H. & Harer, J. "Computational Topology: An Introduction".
- Edelsbrunner, H. & Harer, J. "Persistent Homology: A Survey".
- Ghrist, R. *Elementary Applied Topology*