

Course Outline for MATH 3622, Mathematics of Data Science

MATH 3622 Mathematics of Data Science (3 units)

Prerequisite: One of the following: MATH 206, MATH 244, [MATH 1151](#), or [MATH 2341](#); MATH 248 or [MATH 2031](#); and CSC 101 or [CSC 1001](#).

Introduction to mathematical foundations of data science including regression, data dimension reduction, clustering, community detection, and computational topology, with a focus on advanced linear algebra tools needed to establish the mathematical foundations of methods, to show convergence, and for error minimization. 3 lectures. Cross-listed as DATA/[MATH 3622](#).

Course Learning Objectives	
1	Prove theorems that guarantee convergence of popular algorithms in data science
2	Use linear algebra and calculus to state the mathematical formulation of common methods in data science
3	Solve linear and nonlinear systems of eigenproblems
4	Perform stability and complexity analysis of various data science algorithms
5	Apply algorithms for community detection on interaction graphs
6	Analyze point clouds of data using methods from topological data analysis and interpret the results
7	Apply mathematics of data science to data sets to solve real-life problems.

Expanded Course Content

Textbooks, materials, and/or other resources
Pattern Recognition and Machine Learning by Christopher M. Bishop
Optimization Algorithms for Data Analysis by Stephen Wright
Numerical Algorithms by Solomon
Topological Data Analysis with Applications by Gunnar Carlsson and Mikael Vejdemo-Johansson

Week	Readings	Topics of Discussion
1	Solomon 4.1, 5.4	What is Data Science? Regression: least squares, Gram-Schmidt, QR factorization
2	Solomon 6.3, 6.4	Regression: cross-validation, bootstrapping
3	Solomon 7.1, Instructor's Notes	Principal component analysis: spectral theorem, singular value decomposition
4	Solomon 7.2, Instructor's Notes	Principal component analysis: power iteration method, Rayleigh Quotient, PCA in higher dimensions
5	Bishop 9.1, Instructor's Notes	Clustering: k-means, optimization
6	Instructor's Notes	Clustering: convergence
7	Instructor's Notes	Linear and nonlinear classifiers
8	Wright 2.9, Instructor's Notes	Support vector machines
9	Instructor's Notes	Community detection: Stochastic block models, null model
10	Instructor's Notes	Community detection: modularity, Louvain algorithm, cluster coefficient
11	Carlsson 4.1-4.3	Simplicial complexes and simplicial homology
12	Carlsson 4.4-4.9	Filtered simplicial complexes and persistent homology
13	Carlsson 5.1-5.2	Feature maps and persistence landscapes
14	Carlsson 6.1-6.4	Topological data analysis
15	None	Final Presentations

Method of Assessment

The possible methods of assessment are written individual examinations, quizzes, in-class worksheets or presentations, projects, and homework. One or more midterms and a cumulative final exam are required.

Instructional Mode

The instructional mode for this class is face-to-face. This designation means that the essential academic standards of the course require in-class student engagement, and thus remote instruction is not appropriate.