MATH 161
Mathematical Foundations of Machine Learning

Course Description
Introduction to mathematical concepts in machine learning methods with emphasis on the theoretical tools needed for developing new machine learning algorithms. Topics include linear algebra and vector calculus in application to supervised learning, regression, classification, unsupervised learning, clustering, dimensionality reduction, and the optimization and probability theory used in machine learning algorithms.

Prerequisites
MATH 010A with a grade of C- or better, MATH 031 with a grade of C- or better; or equivalent; or consent of instructor.

Textbook
Mathematics for Machine Learning (2020) by Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong

Additional Resources
Convex Optimization: Algorithms and Complexity (Volume 8 No. 3-4, 2015) by Sébastien Bubeck

Matrix Methods in Data Mining and Pattern Recognition (2007) by Lars Elden

Suggested Lecture Schedule

Week #	Textbook Chapter(s)	Topic(s)
1	1, 2, 8	Data as vectors, systems of linear equations, vector spaces, linear independence, basis and rank, linear mapping Discussion/Lab – Introduction to Python
2	2, 3	Linear regression, affine spaces, inner products, lengths and distances, angles and orthogonality, orthonormal basis, orthogonal complement and projections Discussion/Lab – Implement linear regression by solving a linear system
3	4, 5	Determinant and trace, Cholesky decomposition, eigendecomposition, diagonalization, singular value decomposition (SVD), principal component analysis (PCA) Discussion/Lab – Implement PCA and analyze Olivetti face dataset

4	5, 7	Convex optimization, optimization using gradient descent, constrained optimization, Lagrange multipliers Discussion/Lab – Implement gradient descent for linear regression
5	6	Construction of probability space, discrete and continuous probabilities, Gaussian distribution, Bayes theorem, independence, maximum likelihood approach Discussion/Lab – Generate Gaussian samples and compute posterior probability using Bayes theorem
6	5, 9	Polynomial regression, classification, logistic and probit regression, neural networks and backpropagation Discussion/Lab – Compare classification using logistic regression, probit regression, and neural networks
7	9	Ridge regression, kernel ridge regression, feature functions, least absolute shrinkage and selection operator (LASSO), support vector machines (SVM), hinge loss Discussion/Lab – Polynomial regression to discuss overfitting and regularization
8	7 (Deisenroth, Faisal, Ong), 3 (Bubeck), 6 (Bubeck)	Legendre-Fenchel transform, convex conjugate, gradient and subgradient, proximal point algorithms, stochastic gradient descent (SGD) Discussion/Lab – Simulate stochastic gradient descent and other variations
9	10	Maximum variance perspective and projection perspective PCA, non-negative matrix decomposition, low-rank approximation, latent variable, k-mean clustering Discussion/Lab – Analyze Olivetti face dataset using PCA and non-negative matrix decomposition
10	7 (Deisenroth, Faisal, Ong), 3 (Bubeck), 6 (Bubeck)	Mini-batching, variance reduction Discussion/Lab – Final exam review and helpdesk for final project