

COMPSCI 589

Machine Learning

Short description:

Introduction to core machine learning models and algorithms for classification, regression, dimensionality reduction and clustering. The course will cover the mathematical foundations behind the most common machine learning algorithms, and the effective use in solving real-world applications. Requires a strong mathematical background and knowledge of one high-level programming language such as Python.

Detailed description:

This course will introduce core machine learning models and algorithms for classification, regression, clustering, and dimensionality reduction. On the theory side, the course will cover the mathematical foundations behind the most common machine learning algorithms. It will focus on understanding models and the relationships between them. On the applied side, the course will focus on effectively using machine learning methods to solve real-world problems with an emphasis on model selection, regularization, design of experiments, and presentation and interpretation of results.

The course will be held in a flipped-classroom manner, with students being assigned pre-recorded videos, and the lectures being reserved for discussions, including Q&A on the lecture topics, exercises, connecting the lecture abstractions to real-world application, implementation considerations and demos. The assignments will involve both mathematical problems and implementation tasks. Knowledge of a high-level programming language is absolutely necessary. Python is most commonly used, but languages such as Matlab, R, Scala, Julia would also be suitable. Strong foundations in linear algebra, calculus, probability and statistics are essential for the successful completion of this course.

Lectures: Monday & Wednesday 2:30-3:45pm.

Credit: 3 units

Instructor:

- [Ina Fiterau](#), mfiterau@cs.umass.edu, (413) 577-0064. Office hours: After class.

Teaching assistants:

- [Iman Deznabi](#), iman@cs.umass.edu.
- [Ke Xiao](#), kexiao@cs.umass.edu.

Textbooks:

- (Optional) Trevor Hastie, Robert Tibshirani, Jerome Friedman, [The Elements of Statistical Learning](#), 2nd Edition, Springer Series in Statistics.
- (Optional) Chris Bishop, [Pattern Recognition and Machine Learning](#), Springer 2006.

Grading:

- Homeworks: 50%
- Midterm: 30%. In class. Date TBA.
- Mini-Project: 10%. Assignment based on an open challenge.
- Checkpoint Quizzes: 10%
- Extra credit: participation, in class and on piazza.

Class materials will be posted to the [Moodle course](#).

Discussions will happen on [Piazza](#) or over Moodle.

Course topics:

1. Introduction to Machine Learning. Simple classifiers
 - Definition of Machine Learning
 - Relationship to other fields

- Course overview
- Learning problem formulation
- Regression vs classification; supervised vs unsupervised; parametric vs nonparametric models
- K-NN classifiers
- Decision trees

Reading: Bishop, Section 1.2.1-1.2.4 Probability Theory. ESL Section 2.3.2. ESL Section 2.5.

2. Probability and estimation

- Random variable independence
- Bayes rule
- Estimators
- Maximum likelihood estimator (MLE)
- Maximum a posteriori estimator (MAP)

Reading: Bishop, 2.1 Binary Variables, 2.2 Multinomial Variables

Advanced: Mitchell, [Estimating Probabilities](#)

3. Naive Bayes

- Bayes Optimal Classifiers
- Conditional Independence
- Naive Bayes
- Learning for Naive Bayes
- Gaussian Naive Bayes
- Naive Bayes use case: the Bag of Words model

Reading: Mitchell, 3.1 and 3.2, [Naive Bayes](#)

4. Linear Discriminant Analysis (LDA)

- Fitting linear responses
- Fitting by least squares
- Maximizing conditional likelihood
- LDA - model class conditional densities as multivariate Gaussians

Reading: ESL 4.1-4.3 (p. 101-102, 106-110). Bishop 4.1.1-4.1.4 Discriminant Functions. Bishop 4.2 Probabilistic Generative Models

5. Logistic Regression (LR)

- Generative vs discriminative classifiers
- Classification using the logistics function
- Gradient methods to solve LR: gradient descent, stochastic gradient descent
- MLE and MAP estimates for LR

Reading: ESL Section 4.4 (p. 119-120, 127-132)

Advanced: Mitchell, 3.3, [Logistic Regression](#)

6. Generalization and Evaluation

- Training error and generalization error
- Hypothesis space, model capacity
- Generalization, overfitting, underfitting, bias-variance trade-off
- Regularization, model selection, cross-validation

(Optional) Deep dive: Machine Learning Theory

- Theoretical model of ML
- Generalization bounds
- Consistent learning
- PAC learning
- Anostic learning. Relationship to bia/variance tradeoff
- Infinite hypothesis space. VC dimension. Sauer's lemma

Reading: Nina Balcan, [Notes on generalization guarantees](#).

7. Support Vector Machines

- Maximizing the margin
- Hinge loss vs logistic loss
- Basis expansions and kernels
- The kernel trick

Reading: ESL Section 12.3. ESL Section 12.3.6 (p. 434-438). Bishop 6.1, 6.2 (p. 291 - 299).

8. Ensemble Methods

- Introduction to ensembles
- Bagging

- Random forests
- Boosting. Adaboost
- Stacking
- (Optional) Deep dive: Analysis of Adaboost.

Reading: ESL Chapter 16 (p. 605-622). Bishop Sections 14.3,14.4 (p. 657 - 665).

9. Linear Regression, Ridge, and Lasso

- Regression intro
- Linear regression
- Ordinary least squares
- Regularization

Reading: ESL Sections 3.1, 3.2.1 (p. 43-51). ESL Sections 3.4.1-3.4.3 (p. 61-73).

10. Regression trees and smoothing

- Regression trees
- Feature selection
- Kernel smoothing

Reading: ESL 6.1 and 6.2 (p. 191-200). ESL 9.2.1, 9.2.2 (305-308).

11. Neural Networks and Deep Learning

- The Multilayer Perceptron (MLP)
- Nonlinear Activations
- Universal Function Approximation
- Convolutional Neural Networks (CNNs) for vision

Reading: ESL 11.3 Neural Networks

12. Backpropagation and Sequential Neural Networks

- Training neural networks
- Backpropagation
- Learning rates and acceleration
- Recurrent neural networks (RNN)
- Long-term Short-term Memory (LSTM)

Reading: ESL 11.4 Fitting Neural Networks. ESL 11.5 Some Issues in Training Neural Networks.

13. Linear Dimensionality Reduction and SVD

- Dimensionality reduction overview
- Linear dimensionality reduction
- Singular Value Decomposition (SVD)

Reading: ESL Section 14.15.1 (p.534-536).

14. Principal Components Analysis

- Eigenvalue decomposition
- Direction of maximum variance
- Principal Component Analysis (PCA)
- Connection between PCA and SVD

Reading: Bishop 12.1 Principal Component Analysis (p.559-569).

15. Sparse Coding, NMF, ICA and Kernel PCA

- Sparse coding
- Nonnegative matrix factorization
- Independent Component Analysis (ICA)
- Kernel PCA

Reading: ESL Section 14.6 (p.553-557). ESL Section 14.7 (p.557-570).

16. Clustering I

- K-Means
- Mixture models
- Expectation Maximization (EM)

Reading: ESL 14.3.4 - 14.3.11 (k=means). ESL 8.5 (EM).

17. Clustering II

- Exhaustive clustering
- Hierarchical clustering
- Spectral clustering

Exam exception policy: If you have any special needs/circumstances pertaining to an exam, you must talk to the instructor **at least 2 weeks before** the exam.

Late homework policy: If you cannot turn in a homework on time, you will need to discuss with the instructor **at least one day in advance**.

Regrade policy: Any requests for regrading must be submitted **within a week** of receiving the grade and preferably discussed during office hours. Each TA will be responsible for a different part of the homework, as indicated when the assignment is issued, so please direct questions appropriately. Only contact the instructors after discussing the issue with the TAs.

Copyright/distribution notice: Many of the materials created for this course are the intellectual property of the course instructors and of the professors whose courses served as a basis for some of the lectures. This includes, but is not limited to, the syllabus, lectures and course notes. Except to the extent not protected by copyright law, any use, distribution or sale of such materials requires the permission of the instructor. Please be aware that it is a violation of university policy to reproduce, for distribution or sale, class lectures or class notes, unless copyright has been explicitly waived by the faculty member.