ColumbiaX: CSMM.102x Machine Learning

Help

Bookmarks

- Machine Learning Course: Getting Started
  - Course Outline
  - Course Information and Support
  - Recommended Readings
  - Pre-Course Survey
- Week 1
- Week 2
- Week 3
- Week 4
- Week 5
- Week 6
- Week 7
- Week 8
- Week 9
- Week 10
- Week 11

Machine Learning Course: Getting Started > Course Outline > Machine Learning Course Outline

Machine Learning Course Outline

Bookmark this page

Synopsis

This course provides an introduction to supervised and unsupervised techniques for machine learning. We will cover both probabilistic and non-probabilistic approaches to machine learning. Focus will be on classification and regression models, clustering methods, matrix factorization and sequential models.

Methods covered in this course include linear and logistic regression, support vector machines, boosting, K-means clustering, mixture models, expectation-maximization algorithm, hidden Markov models, among others. Please see the course outline below for a more detailed description of the contents of this course.

Course Outline

Week 1

- **Lecture 1**: We will discuss the various perspectives of the course and machine learning in general. We will then cover the maximum likelihood problem for learning parameters of a probability distribution.
- **Lecture 2**: We move to our first supervised learning problem of linear regression. We discuss the least squares approach to linear regression and understand the geometric intuitions of the problem.

Week 2

- **Lecture 3**: We continue our discussion of least squares by thinking probabilistically about the problem, making connections to maximum likelihood. This will motivate the ridge regression approach to linear regression through a technique called regularization. We analyze and compare these two fundamental approaches to linear regression via the SVD.
- **Lecture 4**: We discuss the bias-variance trade-off using least squares and ridge regression as a motivating example. We then introduce Bayes rule and maximum a posteriori (MAP) inference as an

- Week 12
- Post-Course Survey
- Final Exam

alternative to maximum likelihood, making connections to ridge regression.

Week 3

- **Lecture 5**: We discuss Bayesian linear regression as a natural development of ridge regression. This leads to a discussion of forming predictive distributions and “active learning” as two features of the fully Bayesian approach.

- **Lecture 6**: We wrap up our focus on regression by considering cases where the dimensionality of the problem is much larger than the number of samples. We first discuss a minimum L2 approach, which is more useful for introducing two key mathematical tools in machine learning: analysis and optimization. We then discuss sparsity-promoting methods for linear regression.

Week 4

- **Lecture 7**: We shift to the supervised learning problem of classification. We cover simple nearest neighbor approaches and discuss what an optimal classifier looks like. This motivates the generic Bayes classification approach, an approximation to the optimal classifier.

- **Lecture 8**: We move to general linear classifiers. We discuss in detail the geometric understanding of the problem, which is crucial to appreciating what a linear classifier tries to do. We discuss the first linear classifier called the Perceptron. While this method has been improved upon, the Perceptron will provide us with our first occasion to discuss iterative algorithms for model learning.

Week 5

- **Lecture 9**: We discuss logistic regression, a discriminative linear classification model. We compare with the generative Bayes classification model via the log odds function. The likelihood distribution formed by the logistic regression model suggests matching it with a prior; through this example we discuss the general Laplace approximation technique for approximating a posterior distribution.

- **Lecture 10**: We make a “trick” we have been using more concrete by discussing feature expansions and their use in kernel methods. After discussing kernels, we look at a specific instance of a powerful nonparametric model that makes use of them for regression (and classification): the Gaussian process.

Week 6

- **Lecture 11**: We return to the geometric view of linear classification and remove all probabilistic interpretations of the problem. This inspires the maximum margin approach to binary classification. We discuss and analyze an optimization algorithm called the support vector machine (SVM) that achieves this max-margin goal. We show how kernels neatly fit into this model with no extra effort.

- **Lecture 12**: We shift to a radically different classification approach to the linear classifiers we have been discussing thus far. Tree classifiers attempt to find partitions of a space by which to classify data separately in each partition. We introduce a statistical technique called the bootstrap to “bag” these trees into a “random forest.”

Week 7

- **Lecture 13**: We discuss and analyze boosting, a method for taking any classifier and making it better. This is done by learning sequences of classifiers on various subsets of the data such that their weighted combination makes significantly better predictions than any individual classifier on its own. We prove the training error theorem of boosting, perhaps the most difficult part of the class, but well worth the effort!
- **Lecture 14**: This lecture marks the beginning of the unsupervised learning portion of the course. The first family of algorithms we consider are clustering algorithms. We present and derive the k-means algorithm, the most fundamental clustering algorithm.

Week 8

- **Lecture 15**: We discuss the expectation-maximization (EM) algorithm for performing maximum likelihood via an indirect route. The EM algorithm is a remarkable technique that makes many difficult problems much easier. We discuss EM in the context of a missing data problem, but it will reappear in Lectures 16, 19 and 21.
- **Lecture 16**: We compare hard and soft clustering models and cover a modified k-means algorithm. We then focus our discussion on a probabilistic approach to clustering called the Gaussian mixture model, deriving an iterative EM algorithm to learn its parameters.

Week 9

- **Lecture 17**: We move to another unsupervised learning problem of factorizing a matrix into the product of two smaller matrices. This is a widely-used technique for collaborative filtering, where we wish to recommend content to users. We motivate the model we discuss, called probabilistic matrix factorization, in the context of movie recommendation.
- **Lecture 18**: We discuss non-negative matrix factorization, which differs from Lecture 17 by restricting all values in the data and model to be greater than zero. This allows for “parts-based learning” from data, of which topic modeling is a prime example. We present the two standard NMF algorithms for this problem.

Week 10

- **Lecture 19**: We cover the fundamental matrix factorization technique called principle components analysis (PCA), a very useful dimensionality reduction approach. Extensions covered include probabilistic PCA for image denoising and inpainting, and kernel PCA for nonlinear dimensionality reduction.

- **Lecture 20**: We move to the unsupervised problem of designing and learning sequential models. Our first topic is the Markov model. We discuss two important properties of Markov chains and apply them to the problems of ranking and semi-supervised classification.

Week 11

- **Lecture 21**: We broaden the Markov model to the hidden Markov model (HMM). We clarify the important difference between the two and discuss an EM algorithm for learning HMMs. We give a high-level discussion of how HMMs can be used for speech recognition.

- **Lecture 22**: We discuss a final sequential model where all unknowns are continuous valued. We present the Kalman filter for object tracking and put all our Bayesian knowledge to use in deriving the filtering algorithm for real-time learning of this continuous-state linear Gaussian model.

Week 12

- **Lecture 23**: In the last week we shift gears to two problems that are very different from what we've previously discussed. In this lecture we cover association analysis, which is the problem of learning interesting highly probable combinations of the form A implies B. The clever and exact algorithm we cover makes this combinatorially “impossible” problem very possible.

- **Lecture 24**: In this final lecture we discuss model selection. We have made many modeling choices throughout this course without knowing exactly which is the best. This lecture discusses two basic techniques for choosing the final complexity of a model.

Grading

Quizzes (15%): The course has 11 quizzes.

Projects (40%): There will be 4 projects worth 10% each for a total of 40%. All projects count.

Final Exam (45%): There will be a final exam one week after the last lecture.

Passing Grade

To pass the course, you must score **60% or above**.

Honor Code

Academic Honesty Policy

You are required to read, and understand the following agreement regarding **Academic Honesty**. Each student is sole owner of his own code and work and **must NOT**:

- Submit work that is not original.
- Publish code or solutions online.
- Post the course questions on forums including stack overflow.
- Submit someone else's work, or a modification of that work, with or without that person's knowledge.
- Allow someone else to submit his/her work, or a modification of that work.
- Solve as a group a quiz or project. **All coursework is to be done by the student working alone.**
- Contract course work out to others.
- Plan or execute with another student a cooperative subterfuge during an exam.
- Make use of unauthorized material during an exam.

Project assignments will be checked with plagiarism detection software.

Thank you for abiding by these rules. Doing so will ensure the experience is fair to everyone taking this class or the future sessions of this class.