Syllabus

for MITA 22:544:631 Algorithmic Machine Learning

Last updated on 9/4/2018

Reference books:

1. C.M. Bishop. *Pattern Recognition and Machine Learning* Springer 2006
2. T. Hastie, R. Tibshirani and J. Friedman. *The Elements of Statistical Learning,* Springer, 2009 (available online in pdf format from authors web site)

Course Web site:

https://sakai.rutgers.edu

Student work:

There will be no tests in the course. Each student will have to do the following tasks:

- There will be six or seven homework assignments involving R or python programming, as well as problems solving and modeling business learning problems
- A term project, possibly in teams of up to 3 students, must be proposed and implemented. Teams are expected to write up a summary of the project and present them in class in about 20 minutes.

Course Overview:

In this course we will cover the broad topic of machine learning both from the point of view of computer science and the theory of algorithms, and from the point of view of statistics with emphasis on Bayesian approach. Applications in business will be emphasized.

On the statistical side we study the so-called parametric and non-parametric methods for both classification and regression. We also describe and distinguish the frequentist vs Bayesian approach. We cover variance reduction methods such as bagging, boosting, and over-fitting remedies such as regularization (ridge and lasso), cross-validation, Akaiki Information Criterion (AIC), and Schwarz's Bayesian Information Criterion (BIC).

For classification, we cover mostly two class case initially, and then generalize to multiple classes. We then go over many well-known techniques for classification and regression and analyze them both from statistical methods and the VC dimension/Rademacher approaches. Among general methods we look at will be the Maximum likelihood (ML) and the Bayesian Maximum a` posteriori (MAP) approaches.

On the algorithmic side, we will cover computational complexity and resource usage of various algorithms. We cover issues that arise in optimization algorithms that arise from large scale data and involve tens of thousands, to millions of variables.

Topics (Topics with star will be covered if there is time):

The specific learning algorithms which will be covered and analyzed thoroughly will be a subset of the following (we will not have time to cover all, but hope to cover as much as we can):

1. Basic review of probability, statistics, Bayes rule, Bayesian vs frequentist approach, Bayes decision rule

2. Simple models such as k-Nearest Neighbor (kNN), Naive Bayes method, and Decision Trees
3. Introduction to Unsupervised (unlabeled) learning: Cluster analysis, k-means and related methods, Principal Component Analysis and other spectral methods, Formulation as supervised learning through association rules,
4. The general concept of Maximum Likelihood (ML), and the Bayesian version, Maximum A Posteriori (MAP) approaches
5. Linear models, General linear models, additive models (e.g logistic regression, linear regression and some examples of ML and MAP fall in this category), both non-parametric and regularized approaches will be discussed
6. Support Vector Machines
7. Neural Networks (Note: for these topics along with ML and MAP methods we have to introduce the basic notions of optimization and algorithms)
8. *Markov random fields, and hidden Markov models, Bayesian networks (applications of directed and undirected graphs to learning)
9. *Theoretical foundations of computation Learning Theory (COLT), what is learning, or learnable, VC dimension/Rademacher approach