Syllabus

B555: Machine Learning
Spring 2018

Instructor C. Raphael (craphael@indiana.edu)

Classes TR 9:30 – 10:45 International Studies Bldg. (GA) 1122

Office Informatics West 315 856-1849

Office Hours MW: 4-5

AIs Yuan Xie (xieyuan@umail.iu.edu) Office Hours: 3-5 Tuesday, Luddy 3139A

Inhak Hwang (inhahwan@indiana.edu) Office Hours: 10-12 Monday, Luddy 3118J

Course Web page http://www.music.informatics.indiana.edu/courses/B555

Course Objective
Machine Learning combines ideas from statistics, computer science, optimization, and engineering resulting in approaches that *learn* from data. The meaning of the word "learn" will depend on the particular ML context, but includes function approximation, estimating the parameters of a model, and finding an optimal strategy or decision-making process. ML has a wide variety of applications whose reach continues to expand, including life sciences, medicine, security, commerce, finance, social sciences, language, the arts, etc. In fact, the challenge may be to find disciplines that are *not* touched by ML. Students will learn the basic ideas of ML, which are statistical in nature, and master general techniques for estimation, classification, and decision-making that apply broadly.

Prerequisites
Graduate standing or permission of instructor. We assume familiarity with linear algebra, multivariate calculus, undergraduate-level probability.

Textbooks
Will will be using two books for this course with a 3rd recommended text

1. *All of Statistics* by Larry Wasserman, Springer, 2010.
2. *Pattern Recognition and Machine Learning* by Christopher Bishop, Springer, 2006.
3. *The Elements of Statistical Learning* by Trevor Hastie, Robert Tibshirani, Jerome Friedman, Springer 2001 (recommended).

Grading
Homework 20%
Midterm 1 20% (Feb. 6 in class)
Midterm 2 25% (March 8 in class)
Final 35%

Homework Policy
Homework will be a mix of pen-and-paper and computing assignments. The computing assignments will be done in R, and will involve a mix of real and synthetic data. Homework must be submitted in class on the day it is due. Late homework will be penalized or not accepted.

Course Outline

1. Mathematical foundations of machine learning
    (a) Probability Theory
    (b) Important discrete and continuous distributions
    (c) Mixture Models
    (d) Simpson's Paradox
    (e) Probabilities on graphs (Graphical Models)
2. Parameter Estimation
    (a) Point estimation and various criteria
    (b) Bayesian statistics
    (c) Curse of Dimensionality
    (d) The Expectation-Maximization algorithm
3. Classification Algorithms
    (a) Logistic Regression and optimization
    (b) Naive Bayes Classifier
    (c) Decision Trees
    (d) Support Vector Machines
4. Regression approaches
    (a) Least squares and linear regression
    (b) Overfitting, Regularization, Model Selection
    (c) Neural networks
5. Graphical Models
    (a) Bayesian Networks
    (b) Markov Random Fields
    (c) Inference in Graphical Models
6. Special Topics
    (a) Hidden Markov models
    (b) Time series models