INTRODUCTION TO MACHINE LEARNING
Fall 2021

Ilya Shpitser
ilyas@cs.jhu.edu

Instructor:	Ilya Shpitser	**Office Hours:**	TBD
Course TA 1:	Gina Wong	**Office Hours:**	TBD
Course TA 2:	TBD	**Office Hours:**	TBD
Class Time:	MWF 1:30pm-2:45pm	**Class Location:**	Online

Course Links

class lectures	https://wse.zoom.us/j/93781516508?pwd=aHE3TnJTYmpMc09uZFJ2MlZ6dkdYUT09 (pw: 646123)
Ilya's office hours	https://wse.zoom.us/my/ilyashpitser
Gina's office hours	
TBD's office hours	
class piazza	https://piazza.com/jhu/fall2021/cs601475675/home
class gradescope	https://gradescope.com/courses/297335 (access code **GEG23D**)

Prerequisites

Math pre-requisites

- Linear algebra (vector spaces, orthogonality, singular value decomposition)
- Multivariate calculus (partial derivative, gradient, Hessian, Jacobian)
- Probability and statistics (random variables, probability distributions, expectations, mean, variance, covariance, conditional probability, law of large numbers, Bayes rule, MLE)

CS pre-requisites

- Algorithms (Dynamic programming, basic data structures, complexity)
- Programming (Python)

General requirement

- Ability to deal with "abstract mathematical concepts".

Course Description

This course serves as a survey introduction to the field of machine learning. The course will cover major classes of problems in machine learning: supervised, semi-supervised, and unsupervised learning, prediction problems (regression and classification), graphical models, dimension reduction, clustering, missing data, reinforcement learning, and causal machine learning. The course will cover both the methodological concepts necessary to address these problems, and provide hands-on experience via implementing algorithms in the Python programming language, and applying them to real data.

Course Topics

- Statistical inference: regression models, maximum likelihood.
- Regularization and model selection.
- Classification: linear classifiers, support vector machines, the perceptron.
- Ensemble methods: boosting, bagging, super learner, random forests.
- Neural networks: activation functions, backpropagation, convolutional and recurrent neural networks.
- Graphical models: factorization, Markov properties, Markov random fields, Bayesian networks, sum-product algorithms.
- Semi-supervised learning: missing data, reinforcement learning, causal machine learning.
- Unsupervised learning: clustering, dimension reduction, learning graphs from data.
- Advanced topics: machine learning in practice, algorithmic fairness, data dependence, learning theory, learning hidden variable graphs.

Course Activities

- Two lectures and one recitation each week (except during midterm weeks, and the last week of classes).
- Two in-class midterms (30% of the grade, 15% each).
- Six homework assignments (42% of the grade, 7% each).
- A final group project (28% of the grade).

Learning Outcomes

After finishing this class, the students will be able to:

- Understand most major machine learning methods in use today.
- Implement machine learning algorithms in Python, and apply them to data.
- Analyze and understand the behavior of machine learning algorithms.

Class Schedule

Lectures	Dates	Content
Lec 1	Mon., Aug. 30	**Introduction.** Administrative details. What is the class about? History and context for machine learning.
Lec 2 Rec 1	Wed., Sep. 1 Fri., Sep. 3	**Supervised Learning.** Introduction. The bias-variance tradeoff Loss functions. Overfitting. Inductive bias. Decision trees.
Lec 3 Rec 2	Wed., Sep. 8 Fri., Sep. 10	**Probability theory.** History. Joint, conditional, and marginal distributions. Conditional independence. Chain rule and Bayes rule. Expected value, variance, covariance. Density functions.
Lec 4	Mon., Sep. 13	**Probabilistic view of data analysis.** Models and hypothesis classes. The likelihood. Score equations. Maximum likelihood. Bayesian reasoning. Conjugate priors. Gibbs sampling. The Metropolis-Hastings algorithm. **(Hw 1)**
Lec 5 Rec 3	Wed., Sep. 15 Fri., Sep. 17	**Regression.** Parametric regressions. Local search. The Taylor expansion of the score. The Newton-Raphson algorithm.
Lec 6	Mon., Sep. 20	**Regression.** Model selection and regularization.
Lec 7 Rec 4	Wed., Sep. 22 Fri., Sep. 24	**Classification.** Decision boundaries. 0-1 loss. Bayes risk versus minimax risk. Optimal Bayes classifier. Generative versus discriminative models. Linear classifiers. Logistic regression. Linear discriminant analysis (LDA). **(Hw 2)**
Lec 8	Mon., Sep. 27	**Classification.** The perceptron and support vector machines (SVMs). Max-margin learning. Slack variables. The kernel trick.
Lec 9 Rec 5	Wed., Sep. 29 Fri., Oct 1	**Ensemble methods.** Bagging. Super learner. Random forests. Boosting. The AdaBoost algorithm.
Lec 10	Mon., Oct. 4	**Neural networks.** The projection pursuit model. Universal approximation. Activation functions. Backpropagation. Dropout learning. Data augmentation. Double descent. Convolutional and recurrent neural networks. **(Hw 3)**
Lec 11 Rec 6	Wed., Oct. 6 Fri., Oct. 8	**Graphical models.** History. Markov random fields (MRFs). Undirected graphs. Local Markov property. Global Markov property. Factorization. The Hammersley-Clifford theorem.
Lec 12	Mon., Oct. 11	**Graphical models.** Wright's path analysis models. Bayesian networks. Directed Acyclic Graphs (DAGs). Local Markov property. D-separation. Factorization. The fundamental theorem of Bayesian networks.
Lec 13	Wed., Oct. 13	**Graphical models.** Inference. Sum-product algorithm. Clique-trees.
Midterm 1	Fri., Oct. 15	
Lec 14	Mon., Oct. 18	**Unsupervised learning.** Clustering problems. The k-means algorithm. Gaussian mixtures. The EM algorithm. **(Hw 4)**
Lec 15 Rec 7	Wed., Oct. 20 Fri., Oct. 22	**Semi-supervised learning.** Missing data problems Missing completely at random (MCAR). Missing at random (MAR). Missing not at random (MNAR). Graphical models for missing data. Identification. Inference with missing data.
Lec 16	Mon., Oct. 25	**Unsupervised learning.** Learning graphs from data. Faithfulness. The Bayesian Information Criterion (BIC). The GES algorithm.
Lec 17 Rec 8	Wed., Oct. 27 Fri., Oct. 29	**Unsupervised learning.** Dimension reduction. Principal component analysis (PCA). Probabilistic PCA. Kernel PCA. **(Hw 5)**
Lec 18	Mon., Nov. 1	**Reinforcement learning.** Markov decision processes (MDPs). Value iteration. Q learning. The exploration-exploitation tradeoff. Model-based reinforcement learning.
Lec 19 Rec 9	Wed., Nov. 3 Fri., Nov. 5	**Causal machine learning.** Counterfactuals. Consistency. Ignorability.
Lec 20	Mon., Nov. 8	**Causal machine learning.** Conditional ignorability. Covariate adjustment. The g-formula. Inverse probability weighting (IPW). Augmented IPW. Heterogeneous treatment effects. **(Hw 6)**
Lec 21	Wed., Nov. 10	**Advanced topics: algorithmic fairness.**
Lec 22	Mon., Nov. 15	**Advanced topics: data dependence.**
Lec 23	Wed., Nov. 17	**Advanced topics: machine learning in practice.**
Lec 24	Mon., Nov. 29	**Advanced topics: learning theory.**
Lec 25	Wed., Dec. 1	**Advanced topics: learning hidden variable graphs.**
Midterm 2	Fri., Dec. 3	
Lec 26	Mon., Dec. 6	**Final Lecture.** Conclusions. Where to go from here? The future of ML.