Carnegie Mellon University
95-828 Machine Learning for Problem Solving

Spring 2017
Tentative Syllabus

Lecture 0:
Linear Algebra Review
Probability Review

Please see links to material at the 'Resources' tab of course web site.

Lecture 1: Introduction

Part I Preparation and Preliminary Analysis
Lecture 2: Data Preparation
Lecture 3: Exploratory Data Analysis

Part II Supervised Learning
Lecture 4: Linear Models
Lecture 5: Generalized Linear Models
Lecture 6: Model Selection
Lecture 7: Model Evaluation
Lecture 8: Tree-based Methods
Lecture 9: Kernels and Support Vector Machines
Lecture 10: Instance-based Learning
Lecture 11: Ensemble Learning
Lecture 12: Bayesian Networks
Lecture 13: Neural Nets and Deep Learning

Part III Unsupervised and Semi-supervised Learning
Lecture 14: Association Rules
Lecture 15: Clustering
Lecture 16: Outlier Analysis
Lecture 17: Semi-supervised and Active Learning

Part IV Learning with Complex Data Types
Lecture 18: Unstructured Data: ML for Text
Lecture 19: Dependent Data: ML for Time Series
Lecture 20: Dependent Data: ML for Sequences
Lecture 21: Dependent Data: ML for Networks

[Jan 17, 19]
Lecture 1: Introduction

- What is ML? ML applications
- Machine learning paradigms
  - Supervised learning (classification, regression, feature selection)
  - Unsupervised learning (density estimation, clustering, dimensionality reduction)
- Data mining concepts & tasks
  - Association rules, similarity search, cluster analysis, outlier analysis
- Basic data types
  - (Mixed) attribute data, text, time series, sequence, network data
- The problem solving process:
  - Business understanding, data preparation, data understanding, modeling, evaluation

Readings:

Witten & Frank	Chapter 1.1-1.3
Provost & Fawcett	Chapter 2

PART I: DATA PREPARATION AND PRELIMINARY ANALYSIS

[**Jan 19, 24**]
Lecture 2: Data Preparation

- Data types
- Data cleaning
  - Missing and inaccurate values
- Feature extraction
  - Feature types and conversion
  - Scaling and normalization
- Data reduction
  - Principal Component Analysis
  - Random projections
  - Non-linear dimensionality reduction
  - Feature subset selection
- Sampling (static and streaming)

Readings:

Aggarwal	Chapter 2

Witten & Frank	Chapter 2, 7.1-7.4
Shalizi	Chapter 18

[Jan 26]
Lecture 3: Exploratory Data Analysis

- Getting to know your data
- Histogram, Kernel Density Estimation
- Charts, graphs, infographics
- Interactive visualization

Readings:

Hastie	Chapter 6
Shalizi	Chapter 14.1-14.5

PART II: SUPERVISED LEARNING

[Jan 31, Feb 2]
Lecture 4: Linear Models

- Linear Regression
- Robust Regression
- Sparse Linear Models
  - Feature subset selection: revisited
  - Shrinkage methods: ridge regression and Lasso
  - Group Lasso, elastic net
- Logistic Regression

Readings:

Hastie	Chapter 3.1-3.4, 4.4
Shalizi	Chapter 2, 11
Murphy	Chapters 1.4, 7.1-7.5, 13.3-13.5
Provost & Fawcett	Chapter 4
Witten & Frank	Chapter 7.5

[Feb 7]
Lecture 5: Generalized Linear Models

- Generalized Linear Models (GLMs)
- Generalized Additive Models (GAMs)
    - Motivation: medical data analysis
    - Basis expansions
    - Generalizations, shape functions

Readings:

Hastie	Chapter 9.1, 9.3, 9.6
Shalizi	Chapter 12
https://web.stanford.edu/~hastie/Papers/gam.pdf	

[Feb 9]
Lecture 6: Model Selection

- What is a good model?
- Overfitting
- Decomposition of error
- Bias-Variance tradeoff
- Cross Validation
- Regularization
- Information Criteria (AIC, BIC, MDL)

Readings:

Provost & Fawcett	Chapter 5
Hastie	Chapter 7

[Feb 14]
Lecture 7: Model Evaluation

- Performance measures for Machine Learning
- Creating baseline methods for comparison
- Visualizing model performance

Readings:

Witten & Frank	Chapter 5

Provost & Fawcett	Chapter 7, 8, 11
Shalizi	Chapter 3, 10

[Feb 16]
Lecture 8: Tree-based Methods

- Regression trees
- Classification trees
- Missing values and pruning
- From trees to rules

Readings:

Hastie	Chapter 9.2
Witten & Frank	Chapter 4.3-4.4, 6.1-6.2
Provost & Fawcett	Chapter 3
Shalizi	Chapter 13
Murphy	Chapter 16.2

[Feb 21]
Lecture 9: Kernels and Support Vector Machines

- SVM intuition, formulation, and the dual
- Slack variables, Hinge loss
- The Kernel trick
- Kernel functions

Readings:

Murphy	Chapter 14.2, 14.5
http://www.cs.cornell.edu/courses/cs578/2007fa/slides_sigir03_tutorial.pdf	
Witten & Frank	Chapter 6.4

[Feb 23, 28]
Lecture 10: Instance-based Learning

- k-Nearest Neighbor Classifier
- Finding nearest neighbors efficiently

- Kernel NN classification
- Kernel Regression
- Kernel PCA
- Locally-weighted Linear Regression

Readings:

Murphy	Chapter 14.1-14.4, 14.7
Witten & Frank	Chapter 4.7-4.8, 6.5
Shalizi	Chapter 7.1, 7.5

[Feb 28, Mar 2]
Lecture 11: Ensemble Learning

- Combining multiple models
- Bagging
- Boosting
- Random Forests
- Stacking

Readings:

Witten & Frank	Chapter 8
Hastie	Chapter 16, 15
Murphy	Chapter 16.6

[Mar 7]
Lecture 12: Bayesian Networks

- Naïve Bayes classification
- Conditional independence
- Representation
- Inference
- Structure and parameter learning

Readings:

Provost & Fawcett	Chapter 9
Witten & Frank	Chapter 6.7
Murphy	Chapter 3.5, 26

[Mar 9] **Midterm Exam (in class)**

[Mar 21]
Lecture 13: Neural Nets and Deep Learning

- The Perceptron
- Multi-layer perceptrons
- Training and Prediction
- Deep neural networks
- Applications of deep networks

Readings:

Murphy	Chapter 16.5, 28
Aggarwal	Chapter 10.7

PART III: UNSUPERVISED AND SEMI-SUPERVISED LEARNING

[Mar 23]
Lecture 14: Association Rules

- Frequent itemsets
- Association rule generation
- Interesting patterns
- Applications

Readings:

Witten & Frank	Chapter 4.5, 6.3
Aggarwal	Chapter 4, 5.4
Provost & Fawcett	Chapter 12

[Mar 28, 30]
Lecture 15: Clustering

- Distance functions

- Fast similarity search
- Hierarchical clustering
- K-means clustering
- Kernelized k-medoids clustering
- The EM algorithm
- Mixture models
- Spectral clustering
- (Biclustering, subspace clustering)
- Applications

Readings:

Provost & Fawcett	Chapter 6, 12 (part)
Witten & Frank	Chapter 6.8
Murphy	Chapter 14.4.2, 25
Aggarwal	Chapter 7.8

[Apr 4, 6]
Lecture 16: Outlier Analysis

- One-class SVM
- LOF and LOCI
- Ensemble methods: feature bagging, iForest
- Applications of Outlier Mining

Readings:

Witten & Frank	Chapter 7.5
Aggarwal	Chapter 8, 9.4, 9.5

[Apr 6, 11]
Lecture 17: Semi-supervised and Active Learning

- Assumptions (smoothness, cluster, manifold)
- Semi-supervised learning
  - Self-training,
  - Multi-view learning,
  - Co-training
- Active learning
  - Uncertainty sampling,

- Query-by-committee,
- Expected model change,
- Density-weighted methods

Readings:

Witten & Frank Chapter 6.9
https://mitpress.mit.edu/sites/default/files/titles/content/9780262033589_sch_0001.pdf
http://burrsettles.com/pub/settles.activelearning_20090109.pdf
http://pages.cs.wisc.edu/~jerryzhu/pub/sslicml07.pdf
http://www.pami.sjtu.edu.cn/rg/papers/intro.pdf

PART IV: LEARNING WITH OTHER DATA TYPES

[Apr 13]
Lecture 18: Unstructured Data: ML for Text

- Representing text
- Named entity extraction
- Novelty and first-story detection
- Topic models
- Applications

Readings:

Provost & Fawcett	Chapter 10
Aggarwal	Chapter 13
Witten & Frank	Chapter 9.5, 9.6

[Apr 18]
Lecture 19: Dependent Data: ML for Time Series

- Time series preparation and similarity
- Trends and Anomalies
- Forecasting with ARMA, ARIMA models
  - De-trending and seasonal components
- Change-point detection

    - Monitoring the learning process: SPC algorithm
    - CUSUM, Minimum MSE
- Multi-variate forecasting with VAR

Readings:

Aggarwal	Chapter 14
Shalizi	Chapter 21

[Apr 25]
Lecture 20: Dependent Data: ML for Sequences

- (Hidden) Markov Models
- (Hidden) Semi-Markov Models
- Hurdle models
- (Marked) Point processes
- Self-exciting and self-correcting processes
- (Multi-variate) Hawkes process

Readings:

Aggarwal Chapter 15
https://ideas.repec.org/p/pra/mprapa/7675.html
https://www.cs.ubc.ca/~murphyk/Papers/segment.pdf
https://arxiv.org/pdf/1011.1788.pdf
http://lamp.ecp.fr/MAS/fiQuant/ioane_files/HawkesCourseSlides.pdf
http://www.eriklewis.com/AFOSR-MURI.pdf
https://arxiv.org/pdf/1203.3680.pdf

[Apr 27]
Lecture 21: Dependent Data: ML for Networks

- Transductive learning
- Learning in networks with and without attributes
- Graph-regularized classification

Readings:

http://eliassi.org/papers/ai-mag-tr08.pdf
http://pages.cs.wisc.edu/~jerryzhu/pub/sslicml07.pdf

http://www.pami.sjtu.edu.cn/rg/papers/intro.pdf
http://graph-ssl.wdfiles.com/local--files/blog%3A_start/graph_ssl_acl12_tutorial_slides_final.pdf

[May 2] **Project Presentations I**
[May 4] **Project Presentations II**

BOOKS

- Data Mining: Practical Machine Learning Tools and Techniques, 3rd Edition, Morgan Kaufmann Series
  Ian H. Witten, Eibe Frank and Mark A. Hall

- Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking, O'Reilly
  Foster Provost and Tom Fawcett

- The Elements of Statistical Learning: Data Mining, Inference, and Prediction, FREE!
  Trevor Hastie, Robert Tibshirani, Jerome Friedman

- Data Mining, The Textbook, Springer 2015.
  Charu C. Aggarwal

- Machine Learning: a Probabilistic Perspective, The MIT Press 2012.
  Kevin P. Murphy

- Advanced Data Analysis from an Elementary Point of View, Cambridge U. Press
  Cosma R. Shalizi

** You can find links to the books at the course front page.*