Machine Learning Syllabus

Graduate Program in Software
SEIS 763: Machine Learning
Dr. Chih Lai

SEIS 763: Syllabus

Machine Learning (ML)

Instructor

Chih Lai, Ph.D.

E-mail: clai@stthomas.edu

(please include "**763-01**" in the subject field of your e-mail)

WWW: **http://www.linkedin.com/pub/chih-lai-ph-d/3/2b6/193**

Canvas: **canvas.stthomas.edu**

Office: OSS 308

Office Hours: 4:00 – 5:00 PM Monday, Wednesday, also by prior appointment.

Class Rooms / Hours

5:45 – 9:00 PM Monday, Wednesday, Room OWS

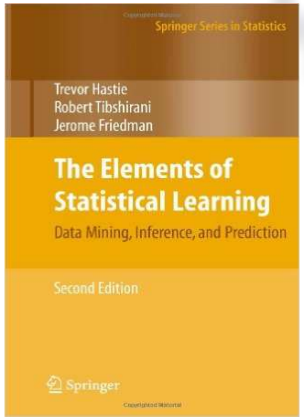
Textbooks

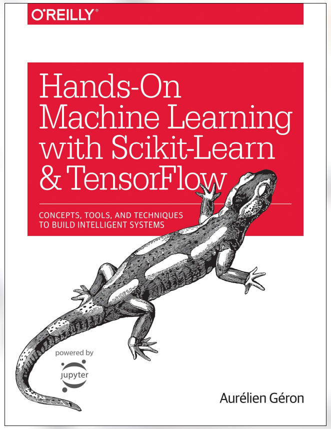
The Element of Statistical Learning, by T. Hastie, R. Tibshirani, J. Friedman, Springer, 2016.

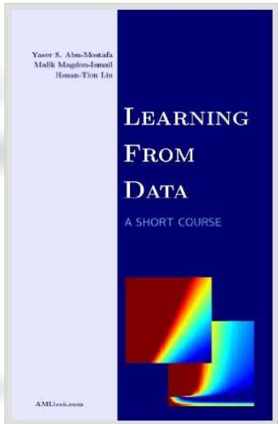
Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, 2017

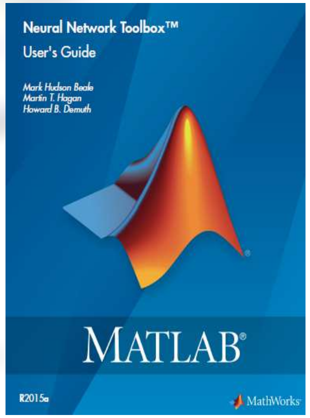
Learning From Data, by Y. S. Abu-Mostafa, M. Magdon-Ismail, H. T. Lin, AMLBook, 2012.

Matlab, Statistics and Machine Learning Toolbox, http://www.mathworks.com/help/stats/









Chapter 1, 2, 3, 4, 5, 6, 11, 12, 14.3, 14.4, 14.5, 16.

Course Description

- Machine Learning (ML) builds computational systems that learn from and adapt to the data presented to them. ML has become one of the essential pillars in information technology today and provides a basis for several applications we use daily in diverse domains such as engineering, medicine, finance, and commerce. This course covers widely used supervised and unsupervised machine learning algorithms used in industry in technical depth, discussing both the theoretical underpinnings of machine learning techniques and providing hands-on experience in implementing them. Additionally, students will also learn to evaluate effectiveness and avoid common pitfalls in applying machine learning to a given problem.

- (1) the fundamental learning methods used by machines, (2) problems, solutions, and advantages of machine learning, (3) regularization of both learning processes to avoid machine over-learning, (4) advanced machine learning methods that works in an infinite dimension space, (5) semi-supervised learning from large amount of low-quality data, (6) boosting of learning, (7) self-organizing map for clustering and classification, (8) case studies, and (9) integration of machine learning and deep learning.

Prerequisite➔ **CEIS 631 or related knowledge / training.**

Note: This is ****NOT**** a high-level end-user tool-training class.

Tentative Class Schedule– Session 1~4

No	Date	Topics
1	5 / 30	* Introduction, Machine Learning (ML), * Predictive vs. Descriptive analytics, Big Data Analytics, Parallel Processing, * Matlab Introduction, Python sklearn package, Visualization, Statistics
2	6 / 4 LR	* Linear Regression (LR), Least Square Errors, RMSE, Residuals, $\boldsymbol{R}^2$, Quadratic LR, * LR Diagnoses (Leverage, Cook Distance), Outlier Removal, Matlab / Python LR,
3	6 / 6 LR GD	* Dummy variables for Categorical variables, **Feature Scaling** * How Machine Learns? Gradient descent (GD), Learning Rate, Local Minimum, Speed-up GD, * Minimize LR Cost function, Estimating θ, Stochastic GD,
4	6 / 11 REG LGI	* Regularization, Overfitting and Control overfitting, Balance of MSE & Penalty * Regularization and L1 / L2 Isosurface, Lasso / Ridge regularization & visualization (plot), * Elastic Net, Cross-Validation, Example – Predicting student grades using LR + regularization. * Logistic Regression introduction

Tentative Class Schedule– Session 5~8

	Date	Topics
5	6 / 13 LGI	* Logistic Regression introduction, hypothesis formula & visualization, * Overall Confusion Matrix, Prediction **ACCURACY** * Probability and Cost, Maximum Likelihood, Negative Log Likelihood (NLL), Cross Entropy * Logistic for Multiclass Problems, One-vs-All Prediction, Outlier Impact
6	6 / 18 CFM ROC PRC AUC	* Lasso for Logistic Regression, Feature Selection, * ***k***-fold Cross-Validation, Confusion Matrix, Prediction **QUALITY**, Skewed Classes, Bias and Variance * Recall / Precision / F-Measure, ROC Curve, AUC (Area Under Curve) for **EACH** class * Plot ROC Curves for Binary and Multiple Classes, Parallel Coordinate Plot and Entropy * Precision-Recall Curve (PRC), ROC AUC vs. PRC AUC, Classification or Regression??
7	***6 / 20***	***Mid-term exam** (5:45—8:00pm) * Cover lecture 1—6 ****Project plan due*** * Project discussion
8	6 / 25 SVM	* Support Vector Machine (SVM), Hyperplane, Margin of Prediction Error, * SVM = Large Margin Classifier, Support Vectors, SVM Margin, Cost Function, * SVM Hinge Loss Interpretation & Visualization, Less Outlier Impact in SVM, * Soft Margin SVM, Slack, SVM Regularization w/ Box Constraints

Tentative Class Schedule– Session 9~10

	Date	Topics
9	6 / 27 SVM Kernel	* SVM Kernel, Kernel Trick, RBF Kernel, Infinite Dimensionality, * Dot Product & Similarity, SVM RBF Kernel with Sigma Regularization, Landmark Points in RBF * SVM Kernel and ***knn*** Classification / Search, Range Query, Speedup kNN, Lazy Classification * SVM One Class Classification, Skewed Class Distribution, Minority Class Classification
10	**7 / 2** Boost NB	* Naïve Bayes (NB), Zero-Frequency Problem and Solution, Numeric Values for NB * Concept Drifting and Dynamic NB * Ensemble Learning, Data Bagging, Weak Learner * Bootstrap, Boosting, AdaBoost, Overfitting in Ensemble * Linear Discriminant Analysis (LDA), Within-class / between-class scatter, LDA Regularization
11	7 / 9 GMM	* Transductive Learning for Un-Labeled Data, Unsupervised Learning * kMeans Clustering, Digit Recognition Example * Clustering as Classification, Classification Rules ≈ Cluster Means + STDV? * Clustering, Gaussian Mixture Model (GMM), Expectation Maximization (EM), EM vs. k-means, * Clustering with Membership, GMM Regularization, GMM for Anomaly Detection * Multi-Dimensional Scaling (MDS), Principle Component Analysis (PCA)

Tentative Class Schedule– Session 11~14

	Date	Topics	
12	7 / 11 SOM	* Self-Organizing Map (SOM), SOM Clustering, SOM Clustering Digits (unsupervised learning), * Self-Organizer Map (SOM), SOM Clustering, SOM Weight Plot, SOM Hit Map * Use SOM for Classification, Creating SOM Prediction Labels and Prediction Confidence * SOM Up-Sampling for Minority Class? SOM ≈ Transductive Learning for Unlabeled Data? * Introduction to Deep Learning / Artificial Intelligence / Relation between ML / DL / AI * Kurtosis Measure for Non-Gaussianity, Central Limit Theory, Independent Component Analysis * Recommender system, collaborative filtering, Matrix factorization, CF / MF * Uncorrelated or Independent Data, ISOMAP,	
13	7 / 16	*Project presentation	(*Notes prepared by teams, **slides due after presentation**)
14	**7 / 18**	*** Final exam**	*** Comprehensive exam** (*Final Project due 5/9)

Exams and Grading

Exams

The exams will be based primarily on the materials covered in class.

Grading

Homework + Quizzes + Participation	15%	➔ **hardcopy required, no late submission**
Project	30%	
Midterm exam	25%	
Final exam	30%	

Letter grade will be assigned **approximately** **as follows:**

80% — 100%	A, A-	(truly exceptional, exceed expectations)
70% — 80%	B+, B, B-	(meet expectations)
60% — 70%	C+, C, C-	
Below 60%	F	

** Final distribution may be adjusted based on the class performance.

** Students who do **NOT** take exam(s) or miss project presentation will receive an **"F"** grade.

** Individual exceptions to the scheduled exams should be both "**individual and rare**".

Course Project

Course Research Project

You will conduct a machine learning project in a team of **5** people.

Check the class schedule for the due dates of project plan and final report.

- Suggestions on Presentations
  - Motivations– Give examples why the problem is interesting and important.
  - Technical contents– Use examples to show how the techniques work.
  - Discussion– Pros & cons, pitfalls.
  - Performance studies (precision and speed)
  - References– Background and related work
  - Issues/impacts related to information ethics and privacy

Project Plan and Final Project Submission

- **Each team must submit a final zipped file w/ the following sections:**

1. **Description of data source and web link(s).**
2. **Size/# of records of the dataset or files.**
3. **# of attributes of the dataset.**
4. **Description of each attribute.**
5. **Some general statistics of the dataset.**
6. **Tools / methods you plan to use in your study.**
7. **Exactly what problems/questions your team plans to predict / study.**
8. **Project presentation slides**
9. **Project WORD documents** with a (short) User Manual.
10. **Project processed datasets.**
11. **Project code / programs**

Do **NOT** expect unrealistic prediction accuracy.

- Share your ****single**** zipped file with the instructor on **UST OneDrive**.

- DO ****NOT**** send your zipped file via email to the instructor.

Checklist Before Presentation

- **If your team plans to use the classroom PC for your presentation, please login to that PC at least few hours before your team's presentation.**
- **Otherwise, it will take a while to login for the first time before your presentation, and it will delay other team's presentation.**
- **If your team plans to use your own laptop for your presentation, please test the connection between your laptop and the classroom projector at least few weeks before your team's presentation.**
- **You may need to find an appropriate connector to connect to the projector.**
- **Please carefully control your team's presentation time.**
  - **Consider possible questions that may interrupt your presentation.**
  - **Please do NOT "READ" your slides.**

Resources

Computing Resources

OSS 327 Computer Lab,

Please check your UST e-mail account regularly.

Support Staff

Instructor Chih Lai for questions regarding the materials covered in class, design and implementation clarification.

GPS Lab assistant Marius Tegomeh (962-5517, mntegomoh@stthomas.edu) for questions on using the equipment in Room 327.

Attendance Policy

Course attendance is expected, but no grade is given for it. Students who miss sessions are responsible for all information in that session. Students who need to miss **presentations** or **exams** due to unavoidable conflicts must arrange in advance to make up the session with the instructor.

Course Assignments

Homework will be assigned from time to time during the semester in order to reinforce the concepts/techniques discussed in the class. Assignments will be collected on the specified due dates. **NO** late submission will be accepted without proper reasons.

Mandatory College Policy Statement

ENHANCEMENT PROGRAM FOR DISABILITY

- In compliance with the University of St. Thomas policy and disability guidelines, your instructors are available to discuss appropriate academic accommodations that you may require as a student with a disability. Requests for academic accommodation should be made during the first week of the semester so timely arrangements can be made. Students are encouraged to register with the Enhancement Program for disability verification and for determination of reasonable accommodations. Classroom accommodations will be provided to students with documented disabilities. Information about the Enhancement Program is available at http://www.stthomas.edu/enhancementprog/ or (651) 962-6315.

ACADEMIC INTEGRITY

- Students are obliged to refrain from acts that they know or, under the circumstances, have reason to believe, will impair the integrity of the university. Violations of academic integrity include, but are not limited to, cheating, plagiarism, unauthorized multiple submissions, knowingly furnishing false or incomplete information to any agent of the university for inclusion in academic records, and falsification, forgery, alteration, destruction, or misuse of official university documents or seal. Any of these violations may cause the expulsion of the student from the program.

DISCLAIMER

- This syllabus is not a contract, but a plan for action. The instructors reserve the right to alter its stipulations, upon prior notification to students, if and when educational circumstances warrant changes.

Chih Lai, Ph.D. Professor

- **Teaching Experience:**

  Dr. Lai is a professor with GPS. He has taught courses in Machine Learning and Deep Learning, Data Mining and Predictive Analytics, Healthcare Analytics. He was also a visiting professor of the Informatics Department at Trier University of Applied Science in Germany in 2010. Dr. Lai also taught an Operating System course at the Computer Science Department of Oregon State University.

- **Research and Publications:**

  Dr. Lai's research interests include Machine Learning and Deep Learning on multimedia data (numerical, images, videos, text). Dr. Lai has published many technical papers on IEEE and ACM conferences / journals. Dr. Lai received teaching and research grants from State of Minnesota, Amazon, Microsoft, and university of St. Thomas. Please visit http://www.linkedin.com/pub/chih-lai-ph-d/3/2b6/193 for details.

- **Industry Experience:**

  Before joining UST, Dr. Lai was a principal software engineer, working on a new aircraft collision avoidance system (ADS-B) which FAA has mandatory installation on most aircraft by 2020. Dr. Lai received three U.S. patents and three European patents, all related to aircraft collision avoidance algorithms. Dr. Lai also worked with Medtronic and has pending patents on monitoring and evaluating Parkinson patients. Other industry experience includes building a network gateway between IBM / Novell networks.