SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN
Established in collaboration with MIT

01.112 Machine Learning (Undergraduate) September 2017

Instructors	Shaowei Lin Lu Wei	shaowei_lin@sutd.edu.sg wei_lu@sutd.edu.sg	1.502.11 1.302.10	Weeks 01-07 Weeks 08-14
Teaching Assistants	Zhang Yan Gary Phua Jie Zhanming	yan_zhang@mymail.sutd.edu.sg gary_phua@mymail.sutd.edu.sg zhanming_jie@mymail.sutd.edu.sg		Class 1 Class 2 Class 2
Office Hours	Shaowei Lin Lu Wei Zhang Yan Gary Phua Jie Zhanming	Tue 10:30 - 11:30am, or by appointment Tue 10:30 - 11:30am, or by appointment Thu 03:00 - 04:00pm, or by appointment Thu 03:00 - 04:00pm, or by appointment Thu 03:00 - 04:00pm, or by appointment		LT4 (2.404) Office (1.302-10) TT9 (1.415) TT9 (1.415) TT9 (1.415)
Lessons	Class 1 Class 2	Mon 01:00pm - 03:00pm Mon 09:00am - 11:00am	Thu 08:30am - 10:30am Thu 01:00pm - 03:00pm	
Classroom	Think Tank 9 & 10 (1.415 & 1.416)			
Prerequisites	A strong background in Linear Algebra, Probability and Python Programming.			
Assessment	Homework (30%), Project (20%), Midterm Exam (25%), Final Exam (25%)			
Schedule	Homework 1-6 Project Data Project Report Midterm Exam Final Exam	Due Fridays 05:00pm Due 30 Oct (Mon) 05:00pm Due 06 Dec (Wed) 05:00pm 03 Nov (Fri) 02:30pm - 04:30pm 15 Dec (Fri) 09:00am - 11:00am		eDimension LT1 (1.102) ISH / MPH ?
Optional	Linear Alg. Review Probability Review Midterm Review I Midterm Review II DEEP I DEEP II DEEP III DEEP IV	14 Sep (Thu) 03:00pm 21 Sep (Thu) 03:00pm 31 Oct (Tue) 10:30am - 12:30pm 02 Nov (Thu) 03:00pm - 05:00pm 11 Oct (Wed) 02:00pm - 04:00pm 08 Nov (Wed) 02:00pm – 04:00pm TBD TBD		TT9 (1.415) TT9 (1.415) LT4 (2.404) TT9 (1.415) TT11 (1.503) TT11 (1.503)

Description

Machine learning is the study of algorithms that improve their performance at a task with experience (Mitchell). In this course, students will learn how machine learning has led to many innovative real-world applications. The students will also gain an in-depth understanding of a broad range of machine learning algorithms from basic to state-of-the-art, such as: naïve Bayes, logistic regression, neural networks, clustering, probabilistic graphical models, reinforcement learning and SVMs.

Learning Objectives

1. Recognize the characteristics of machine learning that make it useful to real-world problems.
2. Understand the basic underlying concepts for supervised discriminative and generative learning.
3. Understand the concepts of cross-validation and regularization, be able to use them for estimation of algorithm parameters.
4. Characterize machine learning algorithms as supervised, semi-supervised, and unsupervised.
5. Have heard of a few machine learning toolboxes.
6. Be able to use support vector machines.
7. Be able to use regularized regression algorithms.
8. Understand the concept behind neural networks for learning non-linear functions.
9. Understand and apply unsupervised algorithms for clustering.
10. Understand the foundation of generative models.
11. Understand the inference and learning algorithms for the hidden Markov model.
12. Understand the learning algorithm for hidden Markov model with latent variables.
13. Understand algorithms for learning Bayesian networks.
14. Understand reinforcement learning algorithms.

Measurable Outcomes

1. List useful real-world applications of machine learning.
2. Implement and apply machine learning algorithms.
3. Choose appropriate algorithms for a variety of problems.

Textbooks

No required textbooks as notes will be provided. For further reading, we recommend

1. C. Bishop: *Pattern Recognition and Machine Learning*. Springer, 2006.
2. G. James, D. Witten, T. Hastie, R. Tibshirani: *An Introduction to Statistical Learning*. Springer, 2009.
3. R. Duda, P. Hart, D. Stork: *Pattern Classification*, 2nd Ed. John Wiley & Sons, 2001.
4. T. Mitchell: *Machine Learning*. McGraw-Hill, 1997.
5. R. Sutton, A. Barto: *Reinforcement Learning: An Introduction*. MIT Press, 1998.

Syllabus

Weeks 01 - 07	Weeks 08 - 14
Introduction	From Generative Models to Graphical Models
Regression - Single Variable	Hidden Markov Models - Introduction
Regression - Multiple Variables	Hidden Markov Models - Supervised Learning
Classification - Logistic	Hidden Markov Models - Inference
Classification - Perceptron	Hidden Markov Models - Unsupervised Learning
Clustering	Bayesian Networks - Learning, Inference
Recommendation	Bayesian Networks - Independence
Support Vector Machines	Bayesian Networks - Structured Learning
Deep Learning	Undirected Graphical Models (Tentative)
Generative Models - Maximum Likelihood	Reinforcement Learning I
Generative Models - Expectation-Maximization	Reinforcement Learning II
Major Themes	Guest Lecture

Software
We will be using the Python programming language for the project and the homework assignments. Python 3.x is preferred over Python 2.x. For the first half of the course, Jupyter Notebook will be the preferred programming environment for homework assignments, because it outputs the results and graphs together with the code. You will print out the notebooks for your homework submissions.

To install Python, Jupyter and other scientific computing packages, do try Continuum's Anaconda distribution. Please avoid Enthought's Canopy distribution, because it does not support Python 3.x. For more information, please see https://jupyter.readthedocs.io/en/latest/install.html.

If you are interested in software development with Python beyond this course, I recommend using the PyCharm IDE instead of Jupyter Notebook. The first tool is more suited for project management, while the second tool is better for data experiments.

Project
The course project consists of two components: the data collection and the final report, and may require a fair bit of programming. More details will be announced in Week 5. Until then, please form groups of 2-3 members for the project.

Homework
Assignments are to be submitted online in eDimension, with solutions combined into a single file. No deadline extensions will be granted, except in extenuating circumstances (e.g. family emergencies).

Honor Code
You are encouraged to form study groups to work on the homework together. However, please write out the solutions on your own, without referring to notes from other students, online solutions or answers from past courses. Reproducing the solution from scratch will help you to find out if you fully understand the material. Please also list your group members on the first page of your submission.

Piazza
This term we will be using Piazza for class discussion. The system is highly catered to getting you help fast and efficiently from classmates, the TAs, and myself. Rather than emailing questions to the teaching staff, I encourage you to post your questions on Piazza. If you have any problems or feedback for the developers, email team@piazza.com. To sign up, please visit https://piazza.com/configure-classes/fall2017/01112 .

Linear Algebra, Probability Review
The teaching assistants will conduct two optional 30min reviews of linear algebra and probability, on the dates listed above. For further readings, we recommend the following textbooks.

D. Poole, *Linear Algebra: A Modern Introduction*. 3rd edition, 2010.
J. L. Devore, *Probability and Statistics for Engineering and the Science*. 8th edition, 2011.

DEEP Option
For students interested in learning advanced machine learning, there will be four additional DEEP lectures which will be taught at a faster pace. At the end of the term, we will offer an optional take-home exam, and students who perform well will received an official DEEP certificate for the course.