MA4270: Data Modeling and Computation (Spring 2017)

Vincent Y. F. Tan

1 Instructor

The instructor for this class is Vincent Y. F. Tan.

- Office: S17-05-20 (or E4-06-06)
- Office Hours: TBD
- Email: vtan@nus.edu.sg

- Lecture time: 2pm to 4pm every Monday and Thursday at S16-0307
- Tutorial: 10am every Wednesday at S17-04-04

2 Information

This is a course on modeling of data from a statistical perspective. This introductory course gives an overview of many concepts, techniques, and algorithms in machine learning, beginning with topics such as classification and linear regression and ending up with more recent topics such as boosting, support vector machines, hidden Markov models, and Bayesian networks. The course will give the student the basic ideas and intuition behind modern machine learning methods as well as a bit more formal understanding of how, why, and when they work. The underlying theme in the course is statistical inference as it provides the foundation for most of the methods covered. The topics covered include:

1. The Perceptron and the Perceptron Convergence Theorem
2. Maximum Margin Classifiers
3. Maximum Likelihood Estimation and Logistic Regression
4. Active Learning
5. Kernels and the Support Vector Machine
6. Model Selection Criteria
7. Feature Selection and Combination of Classifiers (Boosting)
8. Mixture Models and the Expectation Maximization Algorithm
9. Clustering Methods including Spectral Clustering
10. Markov and Graphical Models

All the information for this class can be found on IVLE.

3 Assessments

- Two in-class quizzes ($2 \times 15\%$)
- Computational exercise (10%)
- Project (10%)
- Final Exam (50%)

4 Quizzes and Exam

There will be two quizzes during the course of the semester (see schedule for the timing). The exam is scheduled on the finals week and will be comprehensive, i.e., covering all topics in the course.

For each of the two quizzes, you are allowed 1 A4-cheat sheet (double-sided). For the final, you are allowed also allowed 1 A4-cheat sheet.

5 Computational Exercise

Details about the computational exercse will be provided in week 3. In the computational exercise, which is due after recess week, you are expected to implement the perceptron, logistic regression, and support vector machines algorithm on some datasets. You are expected to produce a well-written report.

6 Project

Details about the project will be provided later on (before recess week). Essentially there are two options: First, you can download a dataset, analyze it numerically using an algorithm proposed in the class by using your favorite program, come up with a conclusion and write a 8-page report. The second option consists of a critical review of a research paper (suggested by the instructor) in the area of machine learning and data modeling. You are expected to have a thorough investigation of the paper and come up with a detailed report. In the report, you need to include the background of the work, the proposed major ideas, the understanding of how machine learning ideas are applied to solve a practical problem, and any possible extension of the work.

7 Prerequisites

You are expected to have a firm grasp of probability, statistics and linear algebra. You should know what is a random variable, probability mass functions, joint and conditional distributions, Markov chains, probability bounds (such as Markov and Chebyshev inequalities), laws of large numbers, conditional independence, etc. You should also know what is a norm, inner products and eigenvectors and eigenvalues.

8 References

Lecture notes will be provided. Other references, such as papers, will be posted on the course website. The following book is also useful.

- C. M. Bishop: Pattern Recognition and Machine Learning, Springer, 2007

9 Detailed Syllabus

See other document for the Schedule.

Week	Lecture/ Tutorial Number	Date	Topic	Other notes
1	L1	09-01-2017	Introduction, Linear classification, Perceptron update rule	
1	L2	12-01-2017	Perceptron convergence theorem and generalization	
2	L3	16-01-2017	Maximum margin classification	
2	L4	19-01-2017	Classification errors, regularization, logistic regression	
3	L5	23-01-2017	Linear regression, estimator bias and variance, active learning	
3	T1	25-01-2017		PS1
3	L6	26-01-2017	Non-linear predictions, kernels	
4		30-01-2017	Chinese new year holiday	
4	T2	01-02-2017		PS2
4	L7	02-02-2017	Kernel regression, kernels	
5	L8	06-02-2017	Support vector machine (SVM)	
5	T3	08-02-2017		PS3
5	L9	09-02-2017	Model selection	
6	L10	13-02-2017	Quiz 1	
6	T4	15-02-2017		PS4
6	L11	16-02-2017	Feature selection	Prof. W. Haskell to take over
6.5			Recess week	
7	L12	27-02-2017	Combining classifiers and boosting	Comput. Ex. Due
7	T5	28-02-2017		PS5
7	L13	02-03-2017	Boosting	
8	L14	06-03-2017	Mixture models	
8	T6	08-03-2017		PS6
8	L15	09-03-2017	Expectation Maximization algorithm	Term paper proposal due
9		13-03-2017	Quiz 2	
9	T7	15-03-2017		PS7
9	L16	16-03-2017	EM Theory (MM procedure)	
10	L17	20-03-2017	Clustering (k-means)	
10	T8	22-03-2017		PS8
10	L18	23-03-2017	Spectral Clustering	
11	L19	27-03-2017	Markov models	Lec19
11	T9	29-03-2017		PS9
11	L20	30-03-2017	Hidden Markov models	
12	L21	03-04-2017	Bayesian networks	
12	T10	05-04-2017		PS10
12	L22	06-04-2017	Learning Bayesian networks: Parameter Learning	
13	L23	10-04-2017	Learning Bayesian networks: Structure Learning	

			(Chow-Liu algorithm)	
13	T10	12-04-2017		PS11
13	L24	13-04-2017	TBD	Term paper due
		05-05-2017 9am-12pm	Final exam	