STA314: Statistical Methods for Machine Learning I

Overview

Machine learning (ML) is a set of techniques that allow computers to learn from data and past experience, rather than requiring humans to specify the desired behaviour by hand. ML has become increasingly central both in statistics as an academic discipline, and in the data science industry. This course provides a broad introduction to commonly used ML methods, as well as the key statistical concepts underlying ML. It serves as a foundation for more advanced courses, such as STA414 (Statistical Methods for Machine Learning II).

We will cover popular statistical methods for supervised and unsupervised learning from data as well as important concepts used in the field, including: training error, test error and cross-validation; classification, regression, and logistic regression; variable selection; penalized regression; principal components analysis; stochastic gradient descent; decision trees and random forests; k-means clustering and nearest neighbour methods. Computational tutorials will support effective application of these methods.

Annoucement

- No class on Sep 30th. We do have tutorials on that day.
- The first midterm is on Wednesday, Sep 25th.
- Both the second problem set (with solution) and the midterm of last year are out.
- The first problem set is out. Its solution is to be posted in 3 days.

Course information

Course email and textbooks

- Course email: sta314@course.utoronto.ca (for requests and communication with the teaching group)
- Textbook: An Introduction to Statistical Learning (ISL)
- References:
    - Elements of Statistical Learning (ESL).

    - Pattern Recognition and Machine Learning (PRML).
    - Convex Optimization by Boyd and Vandenberghe (ConvOpt).
- Piazza: sign-up link.
- Syllabus.

Staff

Instructor	email	office	office hours (OH)	OH mode
Xin Bing	xin.bing@utoronto.ca	UY 9192	Mon, 12:45pm-2:45pm	Zoom link

TAs

Section	TA	OH	OH mode
LEC0101	Haochen Song	Tue (5pm-6pm)	Zoom link
	Xiaochuan Shi	Mon (5pm-6pm)	Zoom link
	Jorge Arturo Esquivel Fuente	Tue (11am-12am)	Zoom link
	Junhao Zhu	Fri (9am-10am)	in-person, SS621
LEC0201	Liam Welsh	Thu (9am-10am)	Zoom link
	Luis Sierra Muntané	Tue (10am-11am)	in-person, SS621
	Konstantinos Christopher Tsiolis	Thu (4pm-5pm)	Zoom link
	Rafael Alexander Valencia Sanchez	Thu (12:30pm-1:30pm)	Zoom link

Lectures

Section	Time	Location
LEC0101	Mon (10am-11am), Wed (11am-1pm)	MP103

Section	Time	Location
LEC0201	Mon (3pm-4pm), Wed (3pm-5pm)	BR200

Tutorials

Section	Session	Time	Location	TA
LEC0101	101	Mon, 11am-12pm	HA403	Haochen
	102		HS106	Xiaochuan
	103		MS3278	Arturo
	104		MS4279	Junhao
LEC0201	201	Mon, 4pm-5pm	MS3278	Luis
	202		MS2173	Haochen
	203		MS4279	K.C.
	204		HS106	Alex

Grading scheme

	Date	Time & Location	Credit	Content	Solution
Quizzes	NA	During tutorials	5%		
Midterm one	Sep 25	LEC0101: 11am-1pm, EX100 LEC0201: 3pm-5pm, EX100	25%	Weeks 1-4	
Midterm two	Oct 23	LEC0101: 11am-1pm, EX200	25%	Weeks 1-7	

	Date	Time & Location	Credit	Content	Solution
		LEC0201: 3pm-5pm, EX100			
Final exam	TBA	TBA	25%	Weeks 1-12	
Course project	Sep 26, 11:59pm	Group sign-up: by Nov 8, 11:59pm, Quercus Prediction submission: by Dec 6, 11:59pm, Kaggle Report due: Dec 8, 11:59pm, Quercus	20%	[Documentation]	

Practical problem sets

	File	Solution
Problem set 1	[Problem set 1]	[Sol:Q1-Q4], [Sol:Q5]
Problem set 2	[Problem set 2]	[Sol:Q1], [Sol:Q2-Q4]
Midterm 23Fall	[Midterm 23Fall]	Not available
Problem set 3	[Problem set 3]	[Sol:Q1-Q3], [Sol:Q4Q5]
Problem set 4		

Lectures

This is a preliminary schedule; it may change throughout the term.

Week	Dates	Lecture Topic	Lecture Slides	Suggested Readings
1	Wed, Sep 4	Course logistics Introduction to Statistical Learning The bias-variance tradeoff	[Lec00.pdf], [Lec01.pdf]	ISL 1, 2.1 - 2.3
2	Mon, Sep 9	Linear regression	[Lec02.pdf]	[Linear algebra & Probability] ISL 3.1, 3.2, 3.3, 3.6
		Tutorial 1 (linear regression)	[Notes], [R code], [Python code]	ISL 3.6
	Wed, Sep 11	Cross-validation Model selection under linear models	[Lec03.pdf]	ISL 5.1, 6.1.1, 6.1.2, 6.1.3 ESL 7.10
3	Mon, Sep 16	Shringkage regression	[Lec04.pdf]	ISL 6.2, 6.4 ESL 7.10
		Tutorial 2 (cv, subset selection)	[Notes], [R code], [Python code]	ISL 5.3.1-5.3.3, 6.5.1
	Wed, Sep 18	Move beyond linearity	[Lec05.pdf]	ISL 7.1-7.4, 7.6, 7.7
4	Mon, Sep 23	Introduction to classification	[Lec06.pdf]	ISL 4.1, 4.2
		Tutorial 3 (Shrinkage regression)	[Notes], [R code], [Python code]	ISL 6.5.2
	Wed, Sep 25	**Midterm one**		
5	Mon, Sep 30	No class		

Week	Dates	Lecture Topic	Lecture Slides	Suggested Readings
		Tutorial 4 (Move beyond linearity)	[R code], [Python code]	ISL 7.8
	Wed, Oct 2	Review of midterm one		
6	Mon, Oct 7	Logistic regression	[Lec07.pdf]	ISL 4.3, 4.4 ESL 4.3, 4.4
		Tutorial 5 (Problem set 3, LR)	[Notes], [R code], [Python code]	ISL 4.7
	Wed, Oct 9	Gradient descent Multi-class logistic regression	[Lec08-GD.pdf], [Lec08-LR.pdf]	ISL 4.3, 4.4 ESL 4.3,4.4 PRML 4.1, 4.3 ConvOpt 2.1-2.3, 3.1, 3.2, 4.1, 4.2
7	Mon, Oct 14	Thanksgiving, no class / tutorial		
	Wed, Oct 16	Discriminant analysis	[Lec09.pdf]	ISL 9.1-9.5 ESL 12.1, 12.2
8	Mon, Oct 21	No class		
		Tutorial 6 (TBA)		
	Wed, Oct 23	**Midterm two**		