

CS 145: Introduction to Data Mining

News

[10/2/2017] First day of class.

[10/1/2017] Book refers to: Jiawei Han, Micheline Kamber, and Jian Pei, [Data Mining: Concepts and Techniques](#), 3rd edition.

Class Schedule

(Future lectures and events are tentative.)

Week#	Date	Topic	Further Reading	Discussion Session	Homework
1	Oct. 2	Introduction [slides] and Math Review	<ul style="list-style-type: none"> Book Chapter 1, 2, 3 Review of probability from a course by David Blei from Princeton U. Machine Learning Math Essentials by Jeff Howbert from Washington U. http://cs229.stanford.edu/section/cs229-prob.pdf optimization 		
1	Oct. 4	Linear Regression [slides] ;	<ul style="list-style-type: none"> http://cs229.stanford.edu/notes/cs229-notes1.pdf 	<ul style="list-style-type: none"> Machine Learning Math Essentials by Jeff Howbert from Washington U. Week 1 Slides 	
2	Oct. 9	Logistic Regression [slides] ; Course Project Introduction [slides]	<ul style="list-style-type: none"> http://cs229.stanford.edu/notes/cs229-notes1.pdf 		
2	Oct. 11	Decision Tree; Regression Tree; Random Forest [slides]	<ul style="list-style-type: none"> Decision Tree: Book Chapter 8.1, 8.2 Regression Tree: http://www.stat.cmu.edu/~cshalizi/350-2006/lecture-10.pdf Random Forest: https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm 	<ul style="list-style-type: none"> Week 2 Slides 	HW1 out
3	Oct. 16	SVM [slides]	<ul style="list-style-type: none"> Notes on SVM by Andrew Ng: http://cs229.stanford.edu/notes/cs229-notes3.pdf SMO: http://cs229.stanford.edu/materials/smo.pdf 		
3	Oct. 18	Neural Network [slides]	<ul style="list-style-type: none"> http://neuralnetworksanddeeplearning.com/ http://www.deeplearningbook.org/ 	<ul style="list-style-type: none"> Week 3 Slides 	HW1 due HW2 out
		Similarity			

4	Oct. 23	measure and KNN [slides]	<ul style="list-style-type: none"> http://scott.fortmann-roe.com/docs/BiasVariance.html 		
4	Oct. 25	Classification Evaluation; Other Practical Issues [slides]	<ul style="list-style-type: none"> Book Chapter 8.5 	<ul style="list-style-type: none"> Week 4 Slides 	
5	Oct. 30	Clustering Basics: K-means; Hierarchical Clustering; DBSCAN [slides]	<ul style="list-style-type: none"> Book Chapter 10.1-10.4 		HW2 due HW3 out
5	Nov. 1	Density Estimation; [slides]	<ul style="list-style-type: none"> https://www.homeworkhelponline.net/blog/math/tutorial-kde http://bebi103.caltech.edu.s3-website-us-east-1.amazonaws.com/2015/tutorials/r3_kde.html http://users.monash.edu/~webb/Files/LiuYangWebbBoughton09.pdf 	<ul style="list-style-type: none"> Week 5 Slides 	
6	Nov. 6	Mixture Models [slides]	<ul style="list-style-type: none"> http://www.stat.cmu.edu/~cshalizi/350/lectures/29/lecture-29.pdf http://www.cs.ubc.ca/~murphyk/Teaching/CS340-Fall06/reading/mixtureModels.pdf 		
6	Nov. 8	Clustering Evaluation; Other Practical Issues [slides]	<ul style="list-style-type: none"> Book Chapter 10.6, 2.4 	<ul style="list-style-type: none"> Midterm Review 	HW3 due
7	Nov. 13	Midterm Exam (in-class)			
7	Nov. 15	Frequent Pattern Mining and Association Rules I [slides]	<ul style="list-style-type: none"> Book Chapter 6 	<ul style="list-style-type: none"> Midterm Paper Review 	
8	Nov. 20	Frequent Pattern Mining and Association Rules II	<ul style="list-style-type: none"> Book Chapter 6 		HW4 out
8	Nov. 22	Sequential Pattern Mining [slides]	<ul style="list-style-type: none"> http://web.engr.illinois.edu/~hanj/cs512/bk2chaps/chapter_8.pdf 	<ul style="list-style-type: none"> Thanksgiving Holiday 	
9	Nov. 27	Sequential Similarity Search [slides]			
9	Nov. 29	Naive Bayes for Text [slides]	<ul style="list-style-type: none"> http://www.ccs.neu.edu/home/yzsun/classes/2014Fall_CS6220/Slides/NB.pdf 	<ul style="list-style-type: none"> Week 9 Slides 	HW4 due
10	Dec. 4	Topic Model [slides]	<ul style="list-style-type: none"> pLSA tutorial: http://arxiv.org/pdf/1212.3900.pdf topic modeling tutorial: https://www.cs.princeton.edu/~blei/kdd-tutorial.pdf 		

10	Dec. 6	Final review [slides]		<ul style="list-style-type: none">• Go over course material	
11	Dec. 13	Final Exam			