Shanghai Jiao Tong University

CS280 Elements of Data Processing

Instructor Information:	Xiangdong An Home Institution: University of Tennessee at Martin Email: xan@utm.edu Office Hours: Determined by Instructor		
Term:	December 16, 2019 - January 7, 2020	**Credits:**	4 units
Class Hours:	Monday through Friday, 120 minutes per teaching day		
Discussion Sessions:	2 hours each week, conducted by teaching assistant(s)		
Total Contact Hours:	64 contact hours (1 contact hour = 45 mins, 2880 mins in total)		
Required Texts (with ISBN):	Recommended texts: J. Han, M. Kamber and J. Pei, Data Mining: Concepts and Techniques, 3rd ed., Morgan Kaufmann, 2012. ISBN: 978-0-12-381479-1. Bing Liu, Web Data Mining, Springer, 2011. ISBN: 978-3-642-26891-5.		
Prerequisite:	*Foundation of Computing*		

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

Course Overview

This course will provide the concepts and techniques in processing gathered date or information in various aspects. It covers the topics of database management, data representation, transformation and analysis, information retrieval, visualization and reporting, privacy issues, web mining and classification methods. It is aimed to help students understand data mining and the tools used in discovering knowledge from the collected data.

Course Goals

1. Understand the basic concept of data mini and web data mining;
2. Understand the fundamental principle of data, Classification Methodologies, clustering and link analysis;
3. Grasp the common algorithms and assessment techniques.

Grading Policy

Participation	10 %
Assignments	40%
Mid-term Exam	20%
Final Exam	30%

Grading Scale

Number grade	Letter grade	GPA
90-100	A	4.0
85-89	A-	3.7
80-84	B+	3.3
75-79	B	3.0
70-74	B-	2.7
67-69	C+	2.3
65-66	C	2.0
62-64	C-	1.7
60-61	D	1.0
≤59	F (Failure)	0

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

Class Schedule

Date	Lecture	Readings
Day 1	Why Processing Data, Type of Attributes, Basic Statistical Description of Data	HKP: 3.1, 2.1, 2.2
Day 2	Data Preprocessing and Cleaning: Missing Values and Outlier Detection and Removal	HKP: 3.2, 12.1, 12.2
Day 3	Entropy and Information Gain	HKP: 8.2.2
Day 4	Transformation by Normalization, Discretization by Binning	HKP: 3.5.1, 3.5.2, 3.5.3
Day 5	Data Dimension Reduction	HKP: 3.4
Day 6	Assessing Correlations and Recommender Systems	HKP: 2.4.7 L:12.4
Day 7	**Mid-term Exam**	
Day 8	Data Visualization, Clustering and Clustering Visualization	HKP: 2.3, 10.1, 10.2 L: 4.2
Day 9	Classification Methodologies: Decision Trees, K-Nearest Neighbor, Naïve Bayes	HKP: 8.2, 9.5.1, 8.3 L: 3.9
Day 10	Experimental Design and Evaluations	HKP: 8.5.1, 8.5.2, 8.5.3, 8.5.5 L: 6.4
Day 11	Text Preprocessing and Information Retrieval	L: 6.1, 6.2, 6.3, 6.5, 6.6
Day 12	Link Analysis & Social Network Analysis	L: 7.1, 7.2, 7.3
Day 13	Data Preprocessing and Web Usage Mining	L: 12.1, 12.2, 12.3
Day 14	Data Linkage, Privacy and Bloom Filters, Social and Ethical Implications of Big Data Analytics	HKP: 13.4
Day 15	**Final Exam**	