CS 513 / SOC 550 Knowledge Discovery and Data Mining Syllabus

The syllabus below describes a recent offering of the course, but it may not be completely up to date. For current details about this course, please contact the course coordinator. Course coordinators are listed on the course listing for undergraduate courses and graduate courses.

Text Books

Required
Daniel T. Larose , *Discovering Knowledge in Data* , 2014

Recommended
, , Lecture notes and handouts

Week-by-Week Schedule

Week	Topics Covered	Reading	Assignments
1	1. What is Data Mining & Knowledge Discover? 2. The Six Phases of Data Mining	Ch. 1	
2	Five Business and Operations Applications	handout	
3	1.Data Cleaning 2.Handling Missing Data 3.Identifying Misclassifications	Ch. 2	case study 1
4	1. Graphical Methods for Outliers 2. Data Transformation: Min-Max Normalization; Z-Score Standardization	Handout	
5	1. Supervised and Unsupervised Learning 2. Methodology for Supervised Learning 3. k-Nearest Neighbor Algorithm 4. Distance Function 5. Database Considerations	Ch. 5, handout.	case study 2
6	1. k-Nearest Neighbor Algorithm for estimation and prediction 2. Choosing k 3. Case Study	Ch 5	
7	1. C4.5 Algorithm 2. Classifications and Regression Trees (CART) Algorithm	Ch. 6	case study 3
8	1. Decision Rules 2. Comparison of the C4.5 and CART Algorithms Applied to Real Data 3. Case Studies	Ch. 6	
9	1. Human Braine 2. Input and Output 3. Neural Network for Estimation and prediction 4. Summation Function 5. Sigmoid Activation Function	Ch. 7	case study 4
10	1. Back-Propagation Algorithm 2. Terminating Criteria 3. Learning Rate 4. Applications of ANN 5. Case Study	Ch. 7	
11	1. Clustering Task 2. Hierarchical Clustering Methods 3. k-Means Clustering	Ch. 8	case study 5
12	1. Applications of k-Means Clustering 2. Applications of k-Means Clustering Using SAS Enterprise Miner 3. Case Study	Ch. 8	
13	Model Evaluation Techniques	Handout	case study 6

Week	Topics Covered	Reading	Assignments
14	Projects and Papers Presentations. An End-to-End Knowledge Discovery and Data Mining Project developed and executed during the semester by each students using a real world data set. The result is documented as a research project and presented at the class.		