

Claremont Graduate University

Introduction to Data Mining

Faculty

Ashish Bhan (KGI and CGU)

Course Overview

This course provides a broad introduction to the concepts, techniques and algorithms commonly used in data mining. The first half of the course will cover the basic principles of data measurement, exploratory data analysis and visualization, model structures, scoring and evaluation, as well as brief reviews of techniques for classification, regression, and clustering. The remaining half of the course will cover specific application areas in greater depth: some possible topics are Text Mining and Information Extraction, Web data mining, Search Engines, Credit Scoring and Spam Filters (depending on student interests).

Pre-requisites

Basic understanding of elementary Probability and Statistics, Calculus, Linear Algebra, and some knowledge of computing, i.e., basic knowledge of MATLAB.

Curriculum Topics

Data Measurement
Exploratory Data analysis and visualization
Dimension reduction
Regression Modeling
Classification algorithms
Clustering algorithms
Information retrieval
Search Engines
Topic models
Web user data
Credit Scoring

Learning Objectives

It is expected that students who have taken the course will have mastered the basic set of ideas required in order to carry out further study and research in data mining methods and algorithms.

Assessment Tools

Homework – 60%; Final Project – 40%

Textbook

Izenman, Alan – Modern Multivariate Statistical Techniques, Springer 2008