Introduction to Data Mining

Instructor: Dr. Xiaoou Li

Course description

The quantity and variety of online data is increasing very rapidly. The data mining process includes data selection and cleaning, machine learning techniques to "learn" knowledge that is "hidden" in data, and the reporting and visualization of the resulting knowledge. This course will cover these issues and will illustrate the whole process by examples of practical applications from the life sciences, computer science, and commerce. Several machine learning topics including classification, prediction, and clustering will be covered.

Course objectives and contents

- To introduce basic applications, concepts, and techniques of data mining, including
  - **data and data pre-preprocessing,**
  - **classification,**
  - **machine learning and deep learning,**
  - **association analysis,**
  - **cluster analysis,**
  - **anomaly detection**,
  - **other issues of data analysis**.
- To develop basic skills of data science for solving practical problems in a variety of disciplines.
- To train students' independent study ability and research thinking.
- To practice data analysis and machine learning software, such as Weka, RapidMiner, R, Matlab, Tensorflow.

Grading elements

- Homework (exercises) 30%
- Reading 10%
- Projects 50%
- Final exam 10%

Class requirements

No database knowledge is needed, modest background in statistics or mathematics such as **linear algebra, probability and statistics**, **discrete mathematics, regression,** and **optimization**, etc.

Textbook

Introduction to Data Mining (2nd edition), by Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, and Vipin Kumar, Pearson**,** 2018.