Machine Learning in Big Data Era

Course outline: This course provides a broad introduction to topics in supervised machine learning, including k-nearest neighbor classifier, regression, decision trees, boosting, perceptrons, Gaussian random fields, and unsupervised learning such as k-means, PCA, and Gaussian mixture models. The business application of these models will be addressed with real data, especially in the context of mobile data analysis. Field trips in local mobile data analytics company and one or two guest talks from data science experts will be included in this course.

Pre-requisites: basic linear algebra and statistics, basic understanding of computer programming

Tentative course schedule (Details will be completed before the end of February)

Week 1:

- What is big data / How to identify big data applications / What are the available tools for large dataset analysis
- Case study: How to collect mobile data from vast amount of users

Week 2:

- unsupervised machine learning algorithms (PCA, k-means, Gaussian mixture model)
- Field trip: Local e-commerce company field trip.

Week 3:

- Supervised machine learning algorithms (k-nearest neighbor, regression, decision tree, boosting, perceptron, Gaussian random fields)
- Case study: Human activity recognition data (from UCI data repository)

Week 4:

- tools available to apply machine learning algorithms (KNIME)
- Guest speaker: TBD

Assessment:

- Reading assignment: 10%
- Participation: 5%
- Homework assignment: 35%
- Midterm exams: 20%
- Final exam: 30%