**Algorithms for Data Analysis and Data Mining**

A 60-hour courses with 30-hour lectures and 30-hour lab practice/tutorials

to be taught by Xin-She Yang

Data analysis and data mining are important topics for many science disciplines such as engineering, life science and computer science. As the volumes of data are increasing rapidly, driven by social media, the internet, wireless sensors, numerical simulation, and the artificial intelligence (AI)-based applications, an important skill as a university student is to make the sense of the data (either from experiments or simulations) and to draw proper conclusions when given a set of data. Algorithms for data analysis and data mining and their implementations are the essential contents for this course. Successful students of this course will be able to

1. Take an analytical approach to provide statistical summaries for a given data set or the observations/data from an experiment, and to interpret the results correctly.
2. Demonstrate an understanding of different algorithms and apply a range of algorithms and techniques in different applications.
3. Identify appropriate algorithms and software tools for use in data analysis and data mining in practice.
4. Explain the outputs from the statistical analysis and data mining, and present/report the results to a wider audience.

The essential syllabus covers a wider range of algorithms and techniques, including

1. Basic mathematical and statistical foundations for data analysis and data mining (e.g., basic probability theory, normal distributions, norms, derivatives, eigenvalues and eigenvectors)
2. Introduction to statistical software, such as R and R Studio (free) and basic programming
3. Summary statistics for data sets (such as means, variances, standard deviations, etc)
4. Regression analysis (best-fit line, generalize linear regression, nonlinear regression, goodness of the fit, p-value, hypothesis testing, and optimization)
5. Classification of data sets (such as nearest neighbour algorithm, k-mean clustering, support vector machine)
6. Principal component analysis (PCA) and logistic regression
7. Time series and time series forecasting (basic trends and seasonal features, moving average, ARIMA, forecasting techniques)
8. More advanced algorithms and techniques for data mining (support vector regression, Lasso method, elastic net method, neural networks and deep learning)

Resources

* Detailed lecture notes and slides will be provided
* Some demo R codes and tutorials will be given to students
* Chapters from the book https://www.sciencedirect.com/book/9780128172162/introduction-to-algorithms-for-data-mining-and-machine-learning