Information on DSA-Coded Modules

DSA1101 Introduction to Data Science

Pre-requisite: H2 pass in Mathematics or equivalent

This module is designed to provide a basic introduction to data science along with real examples and case studies from both academic and industrial sources, in areas as diverse as finance, biological sciences, physics and pharmacy.

DSA2101 Essential Data Analytics Tools: Data Visualisation

Pre-requisites: DSA1101 and MA1101R and ST2131/MA2216

Data visualisation is an essential tool for data analytics. This module is an introduction to data cleaning, exploration, analysis and visualisation. Students will learn how to take raw data, extract meaningful information, use statistical tools, and make visualisations. Topics include: programming in R, introduction to data storage systems, data manipulation, exploratory data analysis, dimension reduction, statistical graphics for univariate, multivariate (high-dimensional), temporal and spatial data, basic design principles and critical evaluation of visual displays of data.

DSA2102 Essential Data Analytics Tools: Numerical Computation

Pre-requisites: MA1101R and MA1102R

This module aims at introducing basic concepts and well-established numerical methods that are very related to of the computing foundation of data science and analytics. The emphasis is on the tight integration of numerical algorithms, implementation in industrial programming language, and examination on practical examples drawn from various disciplines related to data science. Major topics include: computer arithmetic, matrix multiplication, numerical methods for solving linear systems, least squares method, interpolation, concrete implementations in industrial program language, and sample applications related to data science.

DSA3101 Data Science in Practice

Pre-requisites: DSA2101 and ST2132

This module is designed to be a continuation of DSA1101 Introduction to Data Science. It focuses on data science methodology and the ability to apply such methodology to practical applications. Real-world problems will be provided by both industrial and academic partners in domains such as transportation, consulting, finance, pharmaceutics, life sciences and physics.

DSA3102 Essential Data Analytics Tools: Convex Optimisation

Pre-requisites: CS1010 and MA1101R and {MA1104 Multivariate Calculus or MA2311}

Convex optimisation is an indispensable technique in dealing with high-dimensional structured problems in data science. The module covers modelling examples; basic concepts for convex functions and sub-gradients; gradient and sub-gradient methods; accelerated proximal gradient methods; stochastic block coordinate descent methods; Lagrangian duals; splitting algorithms and implementations

DSA4211 High-Dimensional Statistical Analysis

Pre-requisite: ST3131

Dimensionality is an issue that can arise in many scientific fields such as medicine, genetics, business

and finance, among others. The statistical properties of estimation and inference procedures must be carefully established when the number of variables is much larger than the number of observations. This module will discuss several statistical methodologies useful for exploring voluminous data. They include principal component analysis, clustering and classification, tree-structured analysis, neural network, hidden Markov models, sliced inverse regression, multiple testing, sure independent screening (SIS) and penalized estimation for variable selection. Real data will be used for illustration of these methods. Some fundamental theory for high-dimensional learning will be covered.

DSA4212 Optimisation for Large-Scale Statistical Inference

Pre-requisites: MA1101R and {MA1104 Multivariate Calculus or MA2311} and ST2132

Computational optimisation is ubiquitous in statistical learning and machine learning. The module covers several current and advanced topics in optimisation, with an emphasis on efficient algorithms for solving large-scale data-driven inference problems. Topics include first and second order methods, stochastic gradient type approaches and duality principles. Many relevant examples in statistical learning and machine learning will be covered in detail.