

DATA - DATA

DATA601 Probability and Statistics (3 Credits)

Provides a solid understanding of the fundamental concepts of probability theory and statistics. The course covers the basic probabilistic concepts such as probability space, random variables and vectors, expectation, covariance, correlation, probability distribution functions, etc. Important classes of discrete and continuous random variables, their inter-relation, and relevance to applications are discussed. Conditional probabilities, the Bayes formula, and properties of jointly distributed random variables are covered. Limit theorems, which investigate the behavior of a sum of a large number of random variables, are discussed. The main concepts random processes are then introduced. The latter part of the course concerns the basic problems of mathematical statistics, in particular, point and interval estimation and hypothesis testing.

Prerequisite: Undergraduate courses in calculus and basic linear algebra. Cross-listed with: MSML601.

Credit Only Granted for: DATA601 or MSML601.

DATA602 Principles of Data Science (3 Credits)

An introduction to the data science pipeline, i.e., the end-to-end process of going from unstructured, messy data to knowledge and actionable insights. Provides a broad overview of what data science means and systems and tools commonly used for data science, and illustrates the principles of data science through several case studies.

Restriction: Must be in one of the following programs: (Data Science Post-Baccalaureate Certificate, Master of Professional Studies in Data Science and Analytics, or Master of Professional Studies in Machine Learning). Cross-listed with: MSML602.

Credit Only Granted for: DATA602, MSML602 or CMSC641.

Formerly: CMSC641.

DATA603 Principles of Machine Learning (3 Credits)

A broad introduction to machine learning and statistical pattern recognition. Topics include: Supervised learning: Bayes decision theory, discriminant functions, maximum likelihood estimation, nearest neighbor rule, linear discriminant analysis, support vector machines, neural networks, deep learning networks. Unsupervised learning: clustering, dimensionality reduction, PCA, auto-encoders. The course will also discuss recent applications of machine learning, such as computer vision, data mining, autonomous navigation, and speech recognition.

Restriction: Must be in one of the following programs: (Data Science Post-Baccalaureate Certificate, Master of Professional Studies in Data Science and Analytics, or Master of Professional Studies in Machine Learning). Cross-listed with: MSML603.

Credit Only Granted for: DATA603, MSML603 or CMSC643.

Formerly: CMSC643.

DATA605 Big Data Systems (3 Credits)

An overview of data management systems for performing data science on large volumes of data, including relational databases, and NoSQL systems. The topics covered include: different types of data management systems, their pros and cons, how and when to use those systems, and best practices for data modeling.

Prerequisite: DATA602.

Restriction: Must be in the Data Science Post-Baccalaureate Certificate of Professional Studies or Master of Professional Studies in Data Science and Analytics program.

Credit Only Granted for: DATA605 or CMSC642.

Formerly: CMSC642.

DATA606 Algorithms for Data Science (3 Credits)

Provides an in-depth understanding of some of the key data structures and algorithms essential for advanced data science. Topics include random sampling, graph algorithms, network science, data streams, and optimization.

Prerequisite: DATA602.

Restriction: Must be in the Data Science Post-Baccalaureate Certificate of Professional Studies or Master of Professional Studies in Data Science and Analytics program.

Credit Only Granted for: DATA606 or CMSC644.

Formerly: CMSC644.