

DATA - DATA

DATA400 Applied Probability and Statistics I (3 Credits)

Random variables, standard distributions, moments, law of large numbers and central limit theorem. Sampling methods, estimation of parameters, testing of hypotheses.

Prerequisite: 1 course with a minimum grade of C- from (MATH131, MATH141); or students who have taken courses with comparable content may contact the department. Cross-listed with: STAT400.

Credit Only Granted for: DATA400, ENEE324, or STAT400.

Additional Information: Not acceptable toward graduate degrees in MATH/STAT/AMSC.

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 of 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: BIOI601, MSML601.

Credit Only Granted for: BIOI601, 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: BIOI602, MSML602.

Credit Only Granted for: BIOI602, 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: BIOI603, MSML603, MSQC603.

Credit Only Granted for: BIOI603, DATA603, MSML603, MSQC603 or CMSC643.

Formerly: CMSC643.

DATA604 Data Representation and Modeling (3 Credits)

An introductory course connecting students to the most recent developments in the field of data science. It covers several fundamental mathematical concepts which form the foundations of Big Data theory. Among the topics included are Principal Component Analysis, metric learning and nearest neighbor search, elementary spectral graph theory, minimum and maximum graph cuts, graph partitions, Laplacian Eigenmaps, manifold learning and dimension reduction concepts, clustering and classification techniques such as k-means, kernel methods, Mercer's theorem, and Support Vector Machines. Some relevant concepts from geometry and topology will be also covered.

Prerequisite: DATA601 or MSML601.

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.

DATA607 Communication in Data Science and Analytics (3 Credits)

Expected learning outcomes include that, in the context of data science and analytics, students should be able to: summarize, report, organize prose, statistics, graphics, and presentations; explain uncertainty, sensitivity/robustness, limitations; describe model generation and representation; discuss interpretations and implications; communicate effectively to diverse audiences within a business organization, and possibly other outcomes.

Prerequisite: DATA602.

DATA612 Deep Learning (3 Credits)

Provides an introduction to the construction and use of deep neural networks: models that are composed of several layers of nonlinear processing. The class will focus on the main features in deep neural nets structures. Specific topics include backpropagation and its importance to reduce the computational cost of the training of the neural nets, various coding tools available and how they use parallelization, and convolutional neural networks. Additional topics may include autoencoders, variational autoencoders, convolutional neural networks, recurrent and recursive neural networks, generative adversarial networks, and attention-based models. The concepts introduced will be illustrated by examples of applications chosen among various classification/clustering questions, computer vision, natural language processing.

Prerequisite: DATA603 or MSML603. Cross-listed with: MSML612.

Credit Only Granted for: DATA612 or MSML612.

DATA641 Natural Language Processing (3 Credits)

Introduces fundamental concepts and techniques involved in getting computers to deal more intelligently with human language. Focused primarily on text (as opposed to speech), the class will offer a grounding in core NLP methods for text processing (such as lexical analysis, sequential tagging, syntactic parsing, semantic representations, text classification, unsupervised discovery of latent structure), key ideas in the application of deep learning to language tasks, and consideration of the role of language technology in modern society.

Prerequisite: DATA603 or MSML603. Cross-listed with: MSML641.

Credit Only Granted for: DATA641 or MSML641.

DATA650 Cloud Computing (3 Credits)

Presents the state of the art in cloud computing technologies and applications. Topics will include: telecommunications needs, architectural models for cloud computing, cloud computing platforms and services. Data center networking, server, network and storage virtualization technologies, and containerization. Cloud operating and orchestration systems. Security, privacy, and trust management; resource allocation and quality of service; interoperability and internetworking. Cross-listed with: MSML650.

Credit Only Granted for: MSML650 or DATA650.

DATA698 Research Methods and Study Design (3 Credits)

Expected learning outcomes include that students should be able to: compose problem specifications relevant to work environment, create project descriptions, determine data and resource requirements, propose appropriate methods analytical methods, construct research plans; determine reporting requirements appropriate to various employment situations, identify intended audiences and uses, propose supporting documentation, and possibly other outcomes. Includes ethical and legal considerations in data science.

Restriction: Must have completed at least 4 courses in the program.