

# CUNY School of Professional Studies

## DATA 609 COURSE TITLE

CUNY SPS Master of Science in Data Science

Fall 2021

**Instructor Name:** Dr. Harry Wang **Instructor Email:** haiyuan.wang@sps.cuny.edu **Class Meetup:**

**Office Hours:**

**Degree Program:** M.S. in Data Science

**Credits:** 3 graduate credits

**Prerequisites:** DATA 605 DATA 606

**Type of Course:** Elective course

### Course Description:

Many problems in data mining and machine learning can be formulated as optimization problems. Therefore, optimization and algorithms are closely related to data mining and machine learning. In this course, students will learn mathematical and optimization algorithms for data mining and machine learning. Major data mining and machine learning algorithms will be studied from optimization perspective. These major algorithms include data fitting and regression, decision trees and random forest, logistic regression, clustering and classification, support vector machines, neural networks, and deep learning. In addition, students will be expected to explain their models, reports, and analyses in plain and easy-to-understand language.

### Course Learning Outcomes:

By the end of the course, students should be able to:

- Understand major data mining and machine learning algorithms from optimization perspective
- Be able to formulate these optimization problems; understand and know solution algorithms.
- Learn how to use optimization tools.
- Program Learning Outcomes addressed by the course:
- Business Understanding. Apply frameworks and processes to build out data analytics solutions from understanding of business goals.
- Solid foundational data programming skills, using industry standard tools, essential algorithms, and design patterns for working with structured data, unstructured data and big data.
- Solid foundational math and statistics skills, with emphasis on linear algebra, probability, Bayesian statistics, and numerical methods.
- Data understanding. Collect, describe, model, explore and verify data.
- Data preparation. Selecting, cleaning, constructing, integrating, and formatting data.
- Optimization Modeling. Selecting optimization modeling techniques, generating test designs, building and assessing models.
- Model implementation and deployment.
- Presentation. Evaluating and communicating results.

### Required texts and materials:

Introduction to Algorithms for Data Mining and Machine Learning, 1st Edition. Xin-She Yang. ISBN: 978-0-12-817216-2

## Relevant Software Hardware and other tools:

Course will be using R for some of the computational homework and projects.

## Please download and install R from here:

<http://cran.r-project.org/>

Use below two links as reference for your R programming

<https://www.edx.org/course/introduction-to-r-for-data-science-2>

<https://r4ds.had.co.nz/index.html>

You are encouraged to ask me questions on the “Ask Your Instructor” forum on the course discussion board where other students will be able to benefit from your inquiries

## How is this course relevant for data analytics professionals?

Math modeling and optimization technics are essential for data analytics problems. It is the foundation for some problems that we will come across in linear regression, machine learning, and data mining fields. With understanding of math modeling techniques, IS and Data Analytics professionals would be able to solve many more real problems more efficiently and effectively.

## Grading

### Grade Distribution

Quality of Performance	Letter Grade	Range %	GPA
Excellent - work is of exceptional quality	A	93 - 100	4
Excellent	A-	90 - 92.9	3.7
Good - work is above average	B+	87 - 89.9	3.3
Satisfactory	B	83 - 86.9	3
Below Average	B-	80 - 82.9	2.7
Poor	C+	77 - 79.9	2.3
Poor	C	70 - 76.9	2
Failure	F	< 70	0

## Schedule

*Note: Schedule is subject to change.*

Dates	Topic
Aug-25 to Sep-12	Introduction to Optimization in Data Mining and Machine Learning
Sep-06 to Sep-19	Mathematical Foundations
Sep-13 to Sep-26	Optimization Algorithms
Sep-20 to Oct-03	Data Fitting and Regression
Sep-27 to Oct-10	Logistic Regressions PCA LDA and ICA
Oct-04 to Oct-17	Data Mining Techniques
Oct-11 to Oct-24	Support Vector Machine and Regressions
Oct-18 to Oct-24	Neural Networks and Deep Learning

## Accessibility and Accommodations

The CUNY School of Professional Studies is firmly committed to making higher education accessible to students with disabilities by removing architectural barriers and providing programs and support services necessary for them to

benefit from the instruction and resources of the University. Early planning is essential for many of the resources and accommodations provided. Please see: [http://sps.cuny.edu/student\\_services/disabilityservices.html](http://sps.cuny.edu/student_services/disabilityservices.html)

### **Online Etiquette and Anti-Harassment Policy**

The University strictly prohibits the use of University online resources or facilities, including Blackboard, for the purpose of harassment of any individual or for the posting of any material that is scandalous, libelous, offensive or otherwise against the University's policies. Please see: [http://media.sps.cuny.edu/filestore/8/4/9\\_d018dae29d76f89/849\\_3c7d075b32c268e.pdf](http://media.sps.cuny.edu/filestore/8/4/9_d018dae29d76f89/849_3c7d075b32c268e.pdf)

### **Academic Integrity**

Academic dishonesty is unacceptable and will not be tolerated. Cheating, forgery, plagiarism and collusion in dishonest acts undermine the educational mission of the City University of New York and the students' personal and intellectual growth. Please see: [http://media.sps.cuny.edu/filestore/8/3/9\\_dea303d5822ab91/839\\_1753cee9c9d90e9.pdf](http://media.sps.cuny.edu/filestore/8/3/9_dea303d5822ab91/839_1753cee9c9d90e9.pdf)

### **Student Support Services**

If you need any additional help, please visit Student Support Services: [http://sps.cuny.edu/student\\_resources/](http://sps.cuny.edu/student_resources/)