



Syllabus & Course Overview

Machine Learning and
Big Data - DATA622

Fall 2023

CUNY School of Professional Studies

Agenda

1. Introductions (30 minutes)
2. Syllabus (5 minutes)
3. Grading (5 minutes)
4. Calendar (5 minutes)
5. Textbooks (5 minutes)
6. Week 1 – Tasks

Introductions



What are you most proud of?

What brought you here?

What do you want to take away from the course?

Syllabus

Week #	Week of	Topics	Key Task(s)	Due
1	Fri Aug-25	Introduction to 622, Intro to Machine Learning	Discussion 1	Sun Sep-03
2	Mon Sep-04	Linear & Logistic Regression	Discussion 2	Sun Sep-10
3	Mon Sep-11	Classification: Discriminant Analysis (LDA, QDA)	Discussion 3	Sun Sep-17
4	Mon Sep-18	Classification: kNN, Naïve Bayes	Discussion 4	Sun Sep-24
5	Mon Sep-25	Tree based methods: Decision Trees	Discussion 5	Sun Oct-01
6	Mon Oct-02	Tree based methods: Bagging, Random Forests, Boosting	Discussion 6 Homework 1	Sun Oct-08
7	Mon Oct-09	Tree based methods: Boosting and Additive Trees	Discussion 7	Sun Oct-15
8	Mon Oct-16	Support Vector Machines	Discussion 8	Sun Oct-22
9	Mon Oct-23	Support Vector Machines	Discussion 9 Homework 2	Sun Oct-29
10	Mon Oct-30	Unsupervised Learning: Clustering	Discussion 10	Sun Nov-05
11	Mon Nov-06	Unsupervised Learning: PCA and Distance metrics	Discussion 11	Sun Nov-12
12	Mon Nov-13	Resampling and model selection and other topics	Discussion 12	Sun Nov-19
-	Mon Nov-20	Thanksgiving Break	-	-
13	Mon Nov-27	Bias Variance Tradeoff	Discussion 13 Homework 3	Sun Dec-03
14	Mon Dec-04	Neural networks continued & Foundation models	Discussion 14	Sun Dec-10
15	Mon Dec-11	Ethics and Responsible AI	Discussion 15 Homework 4	Sun Dec-17

Class Calendar

2023	September					
MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY	SUNDAY
28	29	30	31	01	02	03 Due: Disc#1
04 Labor Day	05 Week 2	06	07	08	09	10 Due: Disc#2
11 Week 3	12	13	14	15	16	17 Due: Disc#3
18 Week 4	19	20	21	22	23	24 Due: Disc#4
25 Week 5	26	27	28	29	30	01
02	03					

2023	October					
MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY	SUNDAY
25	26	27	28	29	30	01 Due: Disc#5
02 Week 6	03	04	05	06	07	08 Due: D6/Asst# 1
09 Columbus Day	10 Week 7	11	12	13	14	15 Due: Disc#7
16 Week 8	17	18	19	20	21	22 Due: Disc#8
23 Week 9	24	25	26	27	28	29 Due: D9/Asst# 2
30 Week 10	31					

2023	November					
MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY	SUNDAY
30	31	01	02	03	04	05 Due: Disc#10
06 Week 11	07	08	09	10	11	12 Due: Disc#11
13 Week 12	14	15	16	17	18	19 Due: D12/Asst# 3
20	21	22	23	24	25	26
27 Week 13	28	29	30	01	02	03
04	05					

2023	December					
MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY	SUNDAY
27	28	29	30	01	02	03 Due: Disc#13
04 Week 14	05	06	07	08	09	10 Due: Disc#14
11 Week 15	12	13	14	15	16	17 Due: D15/Asst#4
18	19	20	21	22	23	24
Final Examinations		Fall Semester Ends		Holiday observed - CUNY SPS is closed		
25	26 Final Grades Due	27	28	29	30	31
01	02					

Course Grading

Course assignments	Number of deliverables	Points for each deliverable	Total available points
Homework Assignments	4 x	100 points	= 400 points
There will be 4 homework assignments: 10% each (100 points each). These are used to used to re-enforce course concepts.			
Quiz	1 x	100 points	= 100 points
Multiple choice quiz that covers the course work (end of semester)			
Discussion Board	15 x	10 points	= 140 points
A student is required to participate (by introducing a question).			

640points (100%)

Weekly calls

- Class calls will be recorded and posted online
- Thursday's 7PM ET
Call to go over Weekly tasks and concepts.

Zoom is on invitations sent to everyone, and on Blackboard,

Textbooks

- Our text book is: Practical Machine Learning in R (2020), Fred Nwanganga, et al
- The course, syllabus and examples will focus on R
- R is the language many of you are familiar with from prerequisites
- Python is not part of this course, but I will try and accommodate you

R
(preferred)
↑
OR
↓
Python

	<p>TITLE: Practical Machine Learning in R (2020) AUTHORS: Fred Nwanganga, Mike Chapple URL: https://www.wiley.com/en-us/Practical+Machine+Learning+in+R-p-9781119591535</p>		<p>TITLE: Machine Learning Projects in R SOURCE: ProjectPro URL: https://www.projectpro.io/projects/data-science-projects/machine-learning-projects-in-r</p>
	<p>TITLE: Machine Learning with Python Tutorial (2021) AUTHORS: Bernd Klein URL: https://python-course.eu/books/bernd_klein_python_and_machine_learning_a4.pdf</p>		<p>TITLE: Python Machine Learning Projects AUTHORS: Brian Boucheron & Lisa Tagliaferri, DigitalOcean URL: https://assets.digitalocean.com/books/python/machine-learning-projects-python.pdf</p>

Week 1

1. Discussion Board Week 1 (introduce yourselves) – Due Sunday EOD
2. Reading: Practical Machine Learning, chapters 1 & 2
3. Concepts:
 - Supervised and unsupervised learning; machine learning vs. statistical learning
Watch: <https://youtu.be/5N9V07Elfig> (18 minutes)
Watch: <https://www.youtube.com/watch?v=Z0v9QMkA3dA> (12 minutes)
 - A quick intro to reinforcement learning (20 minutes)
Read: <https://deepsense.ai/what-is-reinforcement-learning-the-complete-guide/>

Week 2

1. Discussion Board Week 2
2. Reading (Practical Machine Learning with R):
 - Linear Regression: chapter 4
 - Logistic regression: chapter 5
3. Concepts:
 - Helpful lecture slides (these slides go into the topics we will be covering next week. Up to slide #28 covers this week's topic):
<https://web.stanford.edu/~hastie/MOOC-Slides/classification.pdf>
 - Multinomial logistic regression:
Please see: <https://stats.idre.ucla.edu/r/dae/multinomial-logistic-regression/>
4. Helpful videos:
 - Statistical learning and models (the whole playlist):
<https://www.youtube.com/watch?v=p9n2w236B48&list=PL5-da3qGB5IDvuFPNoSqheihPOQNJpzyy&index=1>