Lahore University of Management Sciences

EE 212 – Mathematical Foundations for Machine Learning and Data Science

Fall 2021

Instructors	Dr. Zubair Khalid
Room No.	9-251
Office Hours	Tuesday, Thursday 3 pm to 4:30 pm
Email	Zubair.khalid@lums.edu.pk
Telephone	042 3560 8477
Secretary/TA	
TA Office Hours	
Course URL (if any)	Current: https://www.zubairkhalid.org/ee240_2021.html Past: https://www.zubairkhalid.org/ee240_2020.html

Course Teaching Methodology (Please mention following details in plain text)

- Teaching Methodology: A blend of both synchronous and asynchronous
- Lecture details: 30% recorded or notes, 70% live interactions (recorded and to be made available to the students)

Live Sessions Rules:

- Lecture video will be recorded and uploaded after the session
- During the lecture you are encouraged to keep your video on but may keep your video off and mute your microphones. You can unmute your microphone and speak when you have to ask a question or answer a question posed by the instructor(s).
- Any discontent with the indicated online mode of teaching should be communicated and discussed with the instructor in advance (else consent is assumed)
- Attendance is not mandatory but maintaining a good record will help students in many ways. Students not frequently attending the lecture will find difficult to cope with the course. We may take attendance during the session and monitor your presence in the class.

Course Basics				
Credit Hours	3			
Lecture(s)	Nbr of Lec(s) Per Week	2	Duration	1 hour and 15 minutes
Recitation/Lab (per week)	Nbr of Lec(s) Per Week		Duration	
Tutorial (per week)	Nbr of Lec(s) Per Week	1	Duration	1 hour

Lahore University of Management Sciences

Course Distribution	
Core	
Elective	Elective Course for Electrical Engineering
Open for Student Category	BS students
Close for Student Category	

COURSE DESCRIPTION
Machine Learning and Data Science are being used these days in a variety of applications including, but not limited to, forecasting in economics and finance, predicting anomalies or signal analysis in engineering, identification of speaker in acoustics, detection of cosmic bubbles in astrophysics and diagnosis in medical imaging. While machine learning and data science have enabled many success stories, and tools are readily available to analyse data or design machine learning systems, the strong mathematical foundations in these areas are of significant importance to understand, review, analyse and evaluate the technical details of the machine learning systems and data science algorithms that are usually abstracted away from the user. This course focuses on the mathematical foundations that are essential to build an intuitive understanding of the concepts related to Machine Learning and Data Science. Topics covered are - Linear Algebra: vectors and matrices, vector spaces, system of linear equations, eigen-value decomposition, singular value decomposition, regression, least-squares, regularization - Calculus: Multivariate calculus and differentials for optimization, gradient descent - Probability: probability axioms, Bayes rule, random variable, probability distributions - Statistics: descriptive stats, inferential stats, statistical tests - Introduction to Neural Networks: single and multi-layer perceptron(s), feedforward and feedback networks - Application to machine learning and data science: principal component analysis (PCA), time series forecasting, clustering etc - Hands-on exercises: Implementation of the exercises will be carried out in Python

COURSE PREREQUISITE(S)	
•	Pre-requisites: None
•	Co-requisites: None

COURSE OBJECTIVES	
	The goal of this course is to provide mathematical foundations of Machine Learning and Data Science. In broad brush, this course aims to: • Provide a thorough introduction to both fundamental and advanced topics of linear algebra necessary for machine learning and data science • Build mathematical foundations of calculus, probability and statistics • Provide an appreciation for applications of ML and Data Science • Equip the students with the basics of Python to enable them to implement and evaluate Machine Learning and Data Science algorithms

Lahore University of Management Sciences

Learning Outcomes	
EE212-	The students should be able to:
CLO1:	Understand the core theoretical concepts serve as foundations of Machine Learning and Data Science
CLO2:	Understand the core theoretical concepts of calculus, probability theory and statistics that serve as foundations of Machine Learning and Data Science
CLO3:	Formulate and implement basic problems in machine learning and data science

Grading break up: Component Details and weightages

Assignments, 20 %
Programming Assignments, 10 %
Quizzes, 15 %
Mid-Exam and Mid-Viva, 25 %
Final Exam and Final Viva, 30 %

Online Assessment Details:

Students are advised to prepare themselves for online assessment (oral viva exam). It is expected that you have a reasonably stable internet connection and you have pre-prepared and familiarized yourself with the indicated online modalities (like Zoom).

Plagiarism policy details:

Usual LUMS plagiarism policy will apply; Following the honor code is expected from students while being assessed in online mode. They are expected to work on their own without consultation from their fellow students for any assessment component except where group work is explicitly indicated; The discussion partners, website, and other sources used in assignments that have contributed to the solution must be acknowledged. Instructions regarding close book task have to be strictly observed; You are advised to work regularly and target consistency in performance. Any abnormal inconsistency in performance in an individual assessment task with the ongoing general performance can be further scrutinized for plagiarism.

Disciplinary Action policy:

Clear cases of noncompliance with regard to violation of honor code, above instructions and plagiarism may also be sent for disciplinary actions. Similarly any other non-serious behavior disrupting the smooth execution of online course may also be referred to DC.

Lahore University of Management Sciences

Examination Detail	
Midterm Exam	Yes/No: Yes Combine Separate: Combined Duration: 120 minutes Preferred Date: TBA Exam Specifications: TBA
Final Exam	Yes/No: Yes Combine Separate: Combined Duration: 180 minutes Exam Specifications: TBA

Textbook(s)/Supplementary Readings

Books:

- S.Boyd and L. Vandenberghe. **Introduction to Applied Linear Algebra – Vectors, Matrices, and Least Squares.** Cambridge University Press, 2019
- M. P. Deisenroth, A. A. Faisal and Cheng Soon Ong. **Mathematics for Machine Learning**. Cambridge University Press, 2019
- G. Strang. **Introduction to Linear Algebra.** 2016
- J. A. Gubner, **Probability and Random Processes for Electrical and Computer Engineers**, Cambridge University Press, 2006.
- S. L. Miller and D. Childers, **Probability and Random Processes: With Applications to Signal Processing and Communications.**
- A. Papoulis and S.U. Pillai, **Probability, Random Variables, and Stochastic Processes.**
- Class notes will be provided to supplement these readings

Lahore University of Management Sciences

Course Topics		
Module	Topic	Additional Remarks
Basic Linear Algebra	Course Overview, notation, vectors and matrices, basic operation on vectors	**Tutorial 1:** Basic matrix and vector operations **Lab exercise 0:** Intro to Python **Lab exercise 1:** Linear independence, basis, matrix rank
	Advanced operations on vectors, norm, angle, inner product	
	Operations on matrices	
	Linear independence, basis, matrix rank	
	Matrix vector product interpretation	
Advanced Linear Algebra	Systems of linear equations, matrix inverses, pseudo inverse	**Tutorial 2:** Advanced matrix and vector operations **Tutorial 3:** Solving a system of linear equations, least-squares regularization **Lab exercise 2:** Least-squares application: data-fitting **Lab exercise 3:** Eigen value decomposition
	Least-squares, constrained least squares regularization	
	Least-squares application: data-fitting, clustering	
	Eigenvalue decomposition plus geometric interpretation	
	Singular-value decomposition (SVD) plus geometric interpretation	
Calculus	Intro to Calculus, interpretation of derivative and first-order difference of the data	**Tutorial 4:** Interpretation: derivative, integration, weighted average, moving average
	Interpretation of integration, weighted average, moving average of time-series	
Probability & Statistics	Introduction to probability theory, axioms of probability	**Tutorial 5:** Axioms of probability **Lab exercise 4:** Probability distributions and Statistical Inference, Statistical tests interpretation of output
	Bayesian analysis overview	
	Random variables and probability distributions	
	Introduction to statistical inference	
	Statistical tests	
Introduction to Neural Networks	Introduction to neural network. Single layer perceptron	**Tutorial 6:** Hands-on working: single layer perceptron example
	Multi-layer perceptron, feedforward and feedback networks, back propagation	
Applications	Dimensionality Reduction using Principal Component Analysis	**Lab exercise 5:** Applications: PCA and Classification
	Linear Regression, Time-series forecasting	
	Classification: Perceptron classifier, Logistic Regression	
	Clustering: k-means clustering	