Foundations of Data Science and Machine Learning

Where: Online on Zoom.

Time: Saturday 2:30 PM – 4:00 PM; 4:30 PM-6:00 PM
Class Format: 1.5 hours of Lecture; 1.5 hours of Problem Solving

Syllabus

Probability and Statistics Module: Probability axioms; Conditional Probability; Bayes' Theorem; Independence; Counting Problems; Discrete and Continuous Random Variables; Expectation; Iterated Expectation; Total Law of Probability; Covariance; Correlation; Entropy; Mutual Information; Frequentist Inference; Bayesian Inference.

Calculus Module: Functions; Derivatives; Multivariate Calculus; Jacobian; Hessian.

Linear Algebra Module: Vectors; Matrices; Basis; Norms; Orthonormality; Linear System, Rank, and Solution; Linear Transformation; Matrix Multiplication; Matrix Decomposition: QR Factorization; Cholesky Decomposition; Singular Value Decomposition; Eigen Decomposition.

Main Textbooks

1. Bertsekas, Dimitri P., and John N. Tsitsiklis. Introduction to Probability. Vol. 1. Belmont, MA: Athena Scientific, 2002.
2. Gibert Strang. Linear Algebra and Learning from Data. Wellesley-Cambridge Press, 2019

References

3. Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong. Mathematics for Machine Learning. Cambridge University Press, 2020. (https://mml-book.github.io)
4. Gibert Strang. Linear Algebra for Everyone, Wellesley-Cambridge Press, 2020

Instructor

Prof. Deepak Subramani; deepakns@iisc.ac.in;

Instruction Plan:

- Theory
  - The theory slot (2:30-4) will have condensed notes in the form of slides.
  - At least one week before each class, optional pre-reading slides will be shared.
- Problem Solving
  - A set of topics discussed over two or three weeks will have a problem set (P-Set).
  - The P-Set will have starred and unstarred problems in each topic.
  - The starred problems will be worked out live in class during the 4:30-6 slot.
  - The unstarred problems can be solved similar to the starred, and will be graded homework exercises.
  - After submission deadline, the solution for unstarred will be shared.
- The class will be over Zoom and whiteboard (iPad Screen) will be shared.
- After each class, the recording and whiteboard writing will be shared.
- All material and recordings will be shared on Google Drive, strictly for your personal consumption (and not for onward transmission).

Class Schedule

Week	Date	Topic	PS	Quiz
Week 1	7 Aug	Introduction, Data Science Fundamentals		
Week 2	14 Aug	Probability Axioms	PS1	
Week 3	21 Aug	Conditional Probability, Bayes Theorem		
				Q1
Week 4	28 Aug	Independence		
Week 5	30 Aug	Counting Problems	PS2	
			PS1 Due	
Week 6	4 Sep	Random Variables: Discrete and Continuous		
				Q2
Week 7	11 Sep	Expectation, Total Law of Probability		
Week 8	18 Sep	Covariance, Correlation, Entropy	PS3	
			PS2 Due	
Week 9	25 Sep	Frequentist and Bayesian Inference		
				Q3
Week 10	3 Oct	Midterm Weekend – Open Book Exam Not Timed		
Week 11	9 Oct	Univariate and Multivariate Calculus, Taylor Series, Finite Difference Method	PS4	
Week 12	16 Oct	Automatic Differentiation		
				Q4
Week 13	23 Oct	Vectors, Matrices, Tensors, Basic Operations, Norms	PS5	
			PS4 Due	
Week 14	30 Oct	Fundamental Spaces, Basis, Orthogonality		
Week 15	6 Nov	Principal Components and Low-Rank Approximation – Eigen, SVD		
				Q5
Week 16	13 Nov	Matrix and Tensor Factorizations	PS6	
			PS5 Due	
Week 17	20 Nov	Numerical Solution of Linear Equations		
Week 18	27 Nov	Final Exam Weekend – Open Book Not Timed	PS6 Due	

Grading Scheme

Item	Type	Total	Best	Score	Total
Timed Quiz Every Third Sunday	MCQ	5	3	5	15
Problem Sets (PS)	Numerical	6	4	5	20
Midterm Exam	MCQ+Descriptive	1	1	25	30
Final Exam	MCQ+Descriptive	1	1	30	30

Only the Best (out of Total) for each item will be used for grading. So you have the flexibility to miss a few deadlines (except midterm and final) based on your schedule.