Course: Topics in Statistics, ML, and Data Science Class Meeting Time: Mondays 11:45 am – 1:45 pm Class Meeting Place: GC 4419 (occasionally online via Zoom) Instructor: Shirshendu Chatterjee Email: shirchat1@gmail.com Office: GC 4305 Office Hour: Mondays 2 – 3 pm

Course Description:

This course introduces the fundamental concepts and mathematical methods used in data science, modern statistics, and machine learning, including the description and theoretical analysis of several current algorithms, their theoretical basis, and associated mathematical frameworks. Many of the algorithms that will be discussed have been successfully used in various areas of real-world products and services. The main topics that we aim to cover will be chosen from the following references.

Expectations from Enrolled students:

Students will be expected to

- read relevant portions of the references,
- participate in class discussion,
- prepare detailed note and, necessary background materials (along with codes, figures, and data analysis if needed) using latex for one of the topics covered in the class and one of the relevant topics (a list of possible topics will be provided) that will not be covered in the class because of time limitation,
- present the last topic mentioned above in a class presentation towards the end of the course.

Grading Policy

The grade will be based on the three deliverables mentioned above.

Reference books:

- Foundations of Data Science By Avrim Blum, John Hopcroft, and Ravindran Kannan
- Foundations of Machine Learning By Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar
- The Elements of Statistical Learning By Trevor Hastie, Robert Tibshirani, and Jerome Friedman
- High-Dimensional Probability : An Introduction with Applications in Data Science By Vershynin
- Understanding Machine Learning: From Theory to Algorithms By Shai Shalev-Shwartz and Shai Ben-David