

## GRS MA881 A1: FALL 2023

# Empirical processes and concentration inequality in modern learning theory

### INSTRUCTOR INFORMATION

*Name:* Debarghya Mukherjee

*Office:* 509, CDS Building (665 Commonwealth Ave.)

*Email:* mdeb@bu.edu

*Office Hours:* Tuesday and Thursday 11:00 AM - 12:30 AM (509)

### CLASS INFORMATION

*Days:* Tuesday and Thursday

*Time:* 9:30 AM - 10:45 AM

*Classroom:* CDS B63

### COURSE DESCRIPTION

In this course, our focus will be on mastering fundamental techniques of the empirical process theory and a selection of concentration inequalities. These valuable tools enable us to analyze the out-of-sample performance of statistical estimators, ranging from simple regressors/classifiers to state-of-the-art deep neural networks.

At its core, the performance of an estimator depends on two crucial factors: the complexity of the model and the amount of available data. A skilled estimator strikes a balance between these aspects. If we employ an overly complex model, it may end up capturing noise and consequently overfitting. This means that while it performs remarkably well on the training data at hand, its performance in the future, unseen data will suffer.

Throughout this course, we will delve into various tools, such as Rademacher complexity, VC dimension, and covering number, among others. By employing these tools and examining numerous examples in both classical and modern learning problems, we will develop an understanding of the complexity inherent in different models.

Moreover, we will learn how to determine the rate of convergence of an estimator by making optimal choices concerning complexity, tailored to the underlying data-generating mechanism. This allows us to strike the right balance and achieve desirable convergence rates.

Lastly, we will explore essential concentration inequalities (e.g. Hoeffding's inequality, Bernstein's Inequality, Chernoff's bound, Talagrand inequality, etc.) that help us bound the out-of-sample performance of estimators. These concentration inequalities provide valuable insights into how fast the estimators converge toward the true underlying values. We will not go very deep into the probabilistic techniques, rather our focus will be to understand the basic ideas frequently used in this area.

### TEXTBOOKS

There are three books which are often used in most research works:

1. *Weak Convergence and Empirical Processes* by AW van der Vaart and Jon Wellner

2. *Empirical processes in M-estimation* by Sara van de Geer

3. *High-Dimensional Statistics: A Non-Asymptotic viewpoint* by Martin Wainwright

We will mostly follow the third book and a few parts of the first two books. Our focus mainly will be on chapters 2-5 and chapters 13-14 of the third book.

### **Evaluation**

The evaluation of the course will be through scribing the lecture notes and presentations. Details will be mentioned in the class.

**Students are expected to be familiar with the academic code of conduct, which can be found [here](#).**