GR 6203, (Fall 2025) Theoretical Statistics III

Vignettes in Statistical Learning Theory

Schedule

Time: Tuesdays 10:10 am to 12:20 pm (with a 5-10 mn break as needed), **Location:** SSW 1025.

Instructor: Samory Kpotufe. *email:* skk2175@columbia.edu

Office hours: After class or upon request:

Description

This course is a PhD level introduction to **Statistical Machine Learning Theory**, with a focus on convergence analysis and relevant technical tools. I will lecture for 8-10 weeks followed by (group) presentations by the class. Lectures will (try and) follow the schedule of topics below.

• We will start with *classical generalization analysis* for parametric classes via the lens of empirical processes, while emphasizing notions of model complexity that naturally arise in problems such as classification and regression. *Some keywords:* uniform concentration, chaining, localization, dimension and entropy numbers, VC dimension, Rademacher and Gaussian complexity, noise conditions, structural risk minimization. (3 weeks)

• We will then delve into *nonparametric analysis*, which require somewhat different intuition on generalization (as we're then dealing with infinite-dimensional classes). *Some keywords:* nearest neighbor methods, local polynomial estimates, smoothness, noise conditions, minimaxity and adaptivity. (2 weeks)

• This will be followed by common lower-bounding techniques that serve to understand the extent to which our generalization upper bounds are tight, and *no-free-lunch* theorems. *Some keywords:* Minimax lower-bounds, Packing numbers, Assouad, Fano. (2 weeks)

• We will then turn to how such classical analysis yields insights into modern non-i.i.d. settings, such as **active learning, transfer learning, multi-task learning**, etc. *Some keywords:* Disagreement metrics, divergences and metrics on probability spaces, sufficient dimension reduction. (3 weeks)

• Class presentations on related topics, papers, or projects (2-3 weeks). I will meet repeatedly with every group to work on presentations. *Possible Topics:* PAC-Bayes Analysis, Model Aggregation vs Model Selection, Implicit Regularization and Min-norm Solutions, etc.

Basic background

While we'll try to have self-contained discussions, familiarity with the following will be helpful.

- Basic probability concepts, e.g., measurability, integration, characterictic functions, $\mathcal{L}_p(\mu)$ spaces, ...
- Basic Linear Algebra, e.g., vector spaces, Spectral and Singular Value theorems, ...
- Basic Real Analysis, e.g. completeness, compactness, forms of continuity, ...
- Basic Statistical concepts, e.g., ℓ_p Risks, Regularization, basic concentration inequalities, ...

Useful Reading: I'll be giving out recommendations on papers and books as class progresses. Some authors of books on the subject (to get a sense): Aan van der Vaart and Jon Wellner, Gabor Lugosi, Luc Devroye, and Lászlo Györfi, Alexandre Tsybakov, etc ...

Grading:
The idea is to mostly base evaluation on class participation and engagement, along with projects, group homeworks, and or paper presentations.