My Seminar Notes

October 29, 2026

09/10/26: FOLDS seminar, Morgane Austern

What happens to double descent when the standard assumptions are violated?

Here, we consider synthetic data and dependent data, e.g., we consider the effects of data augmentation on this phenomenon.

[SJ: Is there a unified model of synthetic / data augmentation (former is stochastic form of augmentation?). Yes, Morgane will give some this talk.]

[SJ: Ask Morgane about Manit + my AI paper on leveraging synthetic data via soft filtering.]

Model 1: hetero data, independent, each $X_i \sim N(0, \Sigma_i)$, simultaneously diagonalizable.

Model 2: dependent data but identically distributed, block dependent, low rank

Model 3: "variance profile" (independent entries, with arbitrary pattern on the diagonal?) [SJ: See pic.]

Thm: formula for limiting risk. Define the function $r(\lambda)$ (λ is the ridge parameter) that satisfies some fixed-point equation.

Intuitively, r is coming from Stieltjes transform.

Idea: represent zero pattern of variance profile S by bipartite graph; blowup is related to graph theoretic properties of this graph

Key notion: degenerate column set, all black vertices that are left unmatched by some maximal matching (these are vertices without enough neighbors) [SJ: See pic.]

White vertices = observations, black vertices = directions in feature space / $\mathcal{X}$-space (these are the vertices we were talking about above)

[SJ: Not necessarily at $p/n = 1$! My caveat to her logic, beyond the $K_{p,n}$ setting!]

[SJ: Ask about random design setting? How well do we understand $|J|$ in this setting? What is known about such properties of random bipartite graphs?]

Her theorems explain when variance / bias(-squared) blow up (either implies a blowup of the test risk)

[SJ: Follow-up work by Morgane on variance profiles whose nonzeros aren't bounded away from zero (i.e., currently we assume). Also she is relaxing the simultaneously diagonalizable condition. But not relaxing the independence across observations assumption.]

[SJ: Gaussian universality beyond independent observations?]

[SJ: Canonical theoretical tool when there is dependence across]

$\rightarrow$ Gaussian universality of high dim'l logistic regression under dependence

Back to data augmentation; classify images, augment the cat by rotating it and augmenting the dataset

However the symmetries are not known and need to be learned

How well does data augmentation work when you don't know the symmetries? [SJ: What does this mean, only augment according to some subgroup $H \leqslant G$ as opposed to the full group G, and see how things work out?]

In independent setting, we have a recipe: Step 1: pretend data is gaussian; Step 2: simplying opt-based estimators via CGMT [SJ: Ask Hamed based on his adv robust regression paper.]

In high dim'l log reg setting, when $p/n \to \kappa$, allow the block dep / m-dep / mixing conditions, show gaussian universality holds. [SJ: Rate of convergence?]

Typically, do Lindeberg swapping via OU or polar or swapping, bound the rate of change along the path for independence case. [SJ: Based on a clever coupling in the β-mixing setting; ask Morgane and Alan for more, what exactly is this "nice" coupling / embedding / Matias' paper.]

Recall CGMT: conv gauss opt qf goes to auxiliary low dim'l opt problem (PO to AO). Use rotational invariance. Study data augmentation under random permutations, sign flipping, cropping.

[SJ: Talk to Kyle about GMM setting, does it correspond to my question about random design setting (marginal?), depends on the weights of the mixtures, exchangeable setting is doable? Block exchangeability would imply that we can just study the expected design.]

[SJ: Ask Morgane about the sampling GMM direction that Jiaoyang encouraged me to look into. Is Lester's work on sampling related to this? Is Sinho's?]