Lecture 21: Minimum norm solutions

Plan for rest of semester:

- GD prefers low/minimum norm solutions
- Project concentration
- Deep network generalization
- Double descent

Goal of course: we can get low test error.

\[ R(\hat{f}) - R(g) = \hat{R}(\hat{f}) - \hat{R}(g) \]

- Opt
- \( \hat{R}(\hat{f}) - \hat{R}(f) \)
- \( \hat{R}(f) - K(f) \)
- \( K(f) - e(g) \)

Generalization by "returning generalization: ..." - Zhang et al.
- "..."

One solution to issue:

- \( \hat{f} \) is adapted to data.
  (our approach: \( \hat{f} \) is norm bounded, & the norm we need depends on data).

In optimization:

1. Near-Initialization ("modally convex" / "neural tangent kernel")

   \[ \frac{1}{2} \| w_k - z \|^2 + \hat{R}(w_0) \leq \frac{1}{2} \| w_0 - z \|^2 + \hat{R}(z) \]

2. Far from initialization

   \( \| \text{norm preservation} \)

   \( \times \) Next two lectures
We'll consider some cases.  (a) New cases are expanded first (case 1a, in group 1, case 1a)

\[ \text{Lemma:} \quad \text{Hybridization center} \]

\[ \text{Theorem:} \quad \text{Transition} \]

\[ \text{Definition:} \quad \text{Function} \]

\[ \text{Algorithm:} \quad \text{Procedure} \]

\[ \text{Proposition:} \quad \text{Property} \]

\[ \text{Proof of (A)} \]

\[ \text{Proof of (B)} \]

\[ \text{Proof of (C)} \]