Lecture 8: signal-to-noise; kernels

Announcements:

* hw1 postponed
* lecture plan

Apx
* near initialization / overparameterization
* other topics

Optimization
* starting 9/27, don't bother with old notes.

Theorem. Let "teacher" $g(x) = \sum_{k=1}^{r} \alpha_k \sigma(\beta_k^T x)$ be given along with signal-to-noise parameter $\tau > 0$, where $\|\beta_k\| = 1$, $\min_{i \neq j} \|\beta_i \neq \beta_j\| \geq 2\tau$, and define temperature $\epsilon := \frac{\tau^{\alpha d}}{2^{\alpha+1}}$. [+ σ ReLU]

If $\boxed{m \geq \frac{1}{\epsilon^2} \ln\left(\frac{1}{\delta}\right)}$, w/ pr $\geq 1-\delta$, $\exists V \in \mathbb{R}^{m \times d}$ with:

1. $\|V - V_0\|_F \leq \frac{\|\alpha\|_2}{\epsilon}$ and $\max_j \|v_j - v_{0,j}\| \leq \frac{\|\alpha\|_\infty}{\epsilon\sqrt{m}}$.
2. $\forall x$, $\|x\| \leq 1$, $F(x;V) = F_0(x;V)$, and
$$\left|\frac{\epsilon}{\sqrt{m}} F(x;V) - g(x)\right| \leq \epsilon\left(16 d \ln\left(\frac{m}{\delta}\right) + \|\alpha\|_1\right).$$

Lemma. σ ReLU.

1. For any fixed $x \in \mathbb{R}^d$, w/ pr $\geq 1-\delta$,
$$|F(x;V_0)| \leq \|x\| \sqrt{m} \ln\left(\frac{2m}{\delta}\right).$$
2. W/ pr $\geq 1-2\delta$, $\forall x \in \mathbb{R}^d$ with $\|x\| \leq 1$,
$$|F(x;V_0)| \leq 32 d \sqrt{m} \ln\left(m/\delta\right).$$

Overparametrization

"interpolation" & "benign overfitting"

Proof of "noise magnitude" lemma.

① $\forall x$ w/ $\|x\| \le 1$ w/ pr $\ge 1-\delta$ $|F(x; w_0)| \le O(\sqrt{m} \ln \frac{m}{\delta})$ — $g_j \sim \mathcal{P}$

$F(x; w_0) = \sum_j a_j \, \sigma(v_j^T x) = \|x\| \sum_j a_j \, \sigma(g_j) = \|x\| \sum_j \underbrace{a_j \, |g_j| \, \mathbb{1}[g_j \ge 0]}$.

w/ pr $\ge 1-\delta$ $|g_j| \le \left(1 + \sqrt{2 \ln m/\delta}\right)\|x\|$

Define $z_j := a_j \sigma(g_j) \longrightarrow |\sum_j z_j| \le \sqrt{\|x\|^2 m \left(1 + \sqrt{2 \ln 2m/\delta}\right)^2 \ln\left(\frac{2}{\delta}\right)}$

$= O(\sqrt{m} \ln m/\delta)$.

Remark. mostly wrong, should be $(\sqrt{m} \ln 1/\delta)$.

② w/ pr $\ge 1-\delta$, $\forall x$, $\|x\| \le 1$, $|F(x; w_0)| \le O(d \sqrt{m} \ln m/\delta)$.

Pick ε-cover Set $\|x\| \le 1$; i.e., $\forall x$ $\|x\| \le 1$, $\exists x' \in S$, $\|x - x'\| \le \varepsilon$.

By brute force, $|S| \le \left(\frac{2d}{\varepsilon}\right)^d$.

By previous part + union bound, w/ pr $\ge 1-\delta$ $\max_{x \in S} |F(x; w_0)| = O\left(\sqrt{m} \ln \frac{m|S|}{\delta}\right)$

w/ pr $\ge 1-\delta$, $\max_j \|v_j\| \le 1 + \sqrt{2 \ln m/\delta}$,

$\Rightarrow \forall x$, $\|x\| \le 1$, pick $z \in S$ s.t. $\|x - z\| \le \varepsilon$, whereby

$|F(x; w_0)| \le \underbrace{|F(z; w_0)|}_{\le O(\sqrt{m} \ln m/\delta)} + \underbrace{|F(x; w_0) - F(z; w_0)|}$

$\le \sum_j a_j \, |\sigma(v_j^T x) - \sigma(v_j^T z)|$

$\le \sum_j a_j \|v_j\| \cdot \|x - z\|$.

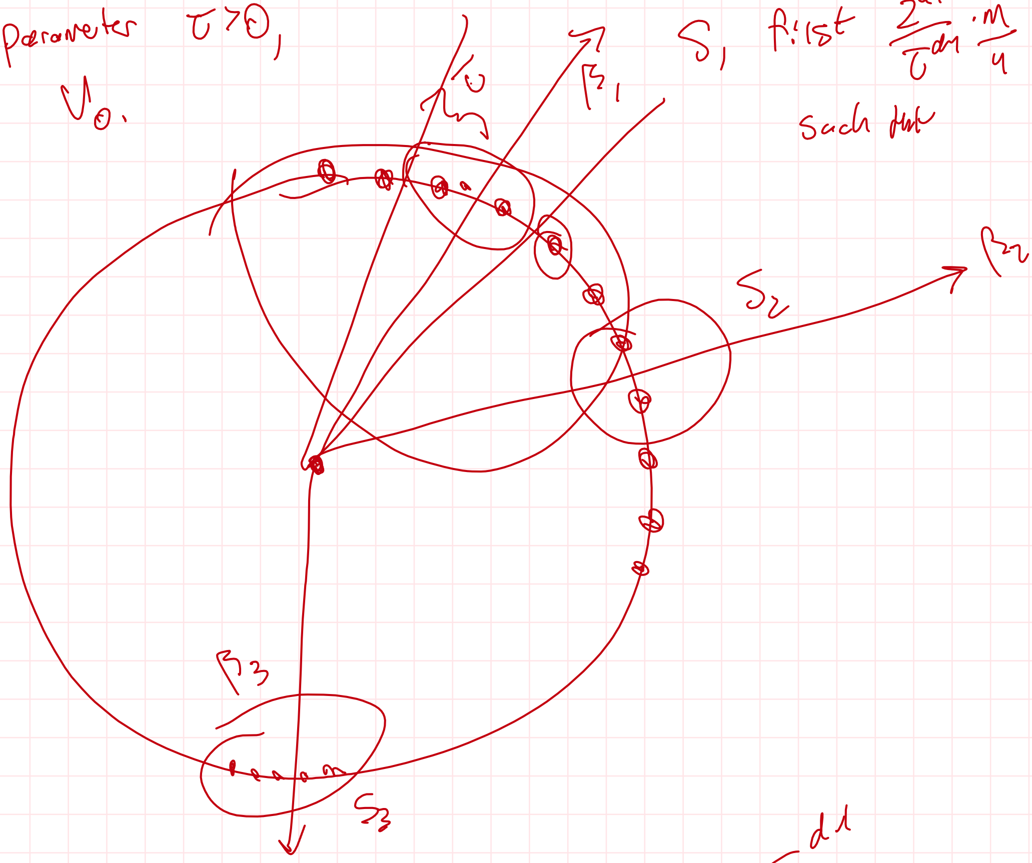
$\le |a|_\infty \, m \cdot (1 + \sqrt{2 \ln m/\delta}) \cdot \varepsilon$. //

Proof of theorem.

Given $g(x) = \sum_{k=1}^{r} \alpha_k \, \sigma(\beta_k^T x)$, $\|\beta_k\| = 1$,

noise parameter $\tau > 0$,

random V_0.



$$\# |S_i| = m \cdot \Pr[v_1 \in S_1] \gtrsim m \cdot \frac{\tau^{d-1}}{2^{d+1}}.$$

$$\forall_j \quad v_j := \begin{cases} v_{0,j} & \text{if } j \notin S_i \ \forall i \\ v_{0,j} + \dfrac{v_{0,j}}{\|v_{0,j}\|} \underbrace{\boxed{\dfrac{2^{d+1} \, |\alpha_i|}{\tau^{d-1} \cdot \tau \cdot \sqrt{m}}}}_{c|\alpha_i|} & \text{if } j \in S_i \ \& \ \operatorname{sgn}(a_i) = \operatorname{sgn}(\alpha_i) \end{cases}$$

Note:

* $\|\cdot\|_F$ & $\|\cdot\|_{2,\infty}$ conditions satisfied

* $F(x; w_0) = F_0(x; w_0)$

$$a_j \sigma(v_j^T x) = a_j \sigma\Big(\underbrace{\Big(1 + \frac{c \cdot |\alpha_i|}{\|v_{0,j}\|}\Big)}_{} v_{0,j}^T x\Big)$$

$$= a_j \Big(1 + \frac{c\,|\alpha_i|}{\|v_{0,j}\|}\Big) \sigma(v_{0,j}^T x).$$

* $\Big| F(x; w_0) - \underbrace{c_0}_{} \, g(x) \Big|$ — cardinality of $S_1 \dots S_n$

$$\le \underbrace{\boxed{\Big| \sum_{k=1}^{r} c_0 \, \alpha_k \, \sigma(\beta_k^T x) - \sum_{j \in S_i} a_j \big(\sigma(v_j^T x) - \sigma(v_{0,j}^T x)\big) \Big|}}_{\text{also small}} + \underbrace{\Big| F(x; V_0) \Big|}_{= O(d \sqrt{m} \ln m/\delta)}$$

also small by construction + Lipschitz.