Lecture 7: Signal-to-noise in wide networks

* HW1?
* HW2 with easy code?

Plan for next few lectures

\[ F(x; U_0) + \langle \nabla F(x; U_0), U - U_0 \rangle \]

\[ = \sum_j a_j (\sigma(x_i U_{0j}) + \sigma'(x_i U_{0j}) x_j (U_{j0} - U_{0j})) \]

\[ = \frac{1}{2} \|U - U_0\|^2 \cdot \frac{D}{2} \| \alpha \| \cdot \|U\| \]

ReLU

\[ |F(x; W) - [F(x; V) + \langle \nabla F(x; V), W - V \rangle]| \leq 5 m^{\frac{1}{3}} (B^{\frac{4}{3}} + B^{\frac{2}{3}} \ln(m/r)^{\frac{1}{4}}) \]

* Signal-to-noise phenomenon

\( F_0 \) is a universal approximator

\( \Rightarrow \) Implies scaling

\( \frac{e}{\tilde{m}} F_0 \)

\( \Rightarrow \) Then can take limits

\( \lim_{m \to \infty} \frac{e}{\tilde{m}} F_0 \overset{a.s.}{\to} F_\infty \)

Krends
In the limit, \[ ||W|| = 5M \]

**Signal property**, given \( \frac{1}{\sqrt{n}} \), prove \( E(\hat{W}, V_0) \in 16 \ln(\frac{C}{T}) \)

Note \( E(\hat{V}, V_0) = 2, e + c \cdot d ||V_0||^2 \)

**Theorem** (signal-sieve phenomenon), i.e., \( \hat{W} \)

Let \( \text{return} \) with \( g(x) = \sum_{\omega} \sigma_0(\phi_\omega(x)) \) given with \( ||W|| = 1 \).

Let \( \text{sieve parameter} \) \( C \) be given with \( \frac{1}{\sqrt{n}} > \frac{1}{C} \).

with \( \frac{1}{\sqrt{n}} > \frac{1}{C} \), we have \( \sigma \in \Omega(1/||V_0||) \).

with \( \alpha = \gamma(1/||V_0||) \)

such that \( f(x) = \sum_{\omega} \xi_\omega \).

Proved: \( ||V|| \leq \sqrt{2}L \), with \( ||V|| = ||V_0|| = 0 \).

**Remark**: Implies unsupervised learning initialization.

Given \( \ln(t) \) (with \( \in R \) a constant, can change \( \sigma(x) \) = \( \sigma(\phi(x)) \)

and \( \sigma \) in \( C \).

then \( \sigma \) is known and \( \sigma \) is fixed with \( ||V|| = 0 \).

and \( \frac{1}{\sqrt{n}} \ln(t) - \frac{1}{\sqrt{n}} \)

with \( \frac{1}{\sqrt{n}} \ln(t) \) and \( \alpha = \gamma(1/||V_0||) \).

For certain \( \alpha \), it is known how to prove this result with \( \alpha \) replaced by \( 0 \) in \( \alpha \).

**Example**: A special case of deep learning learning heuristics.

Define \( \sigma = \frac{1}{\sqrt{n}} \ln(t) \cdot \sigma(x) \)

Given \( (\phi_{(x)}, \alpha_{(x)}) \), define \( \alpha_{(x)} = \alpha_{(x)} + \frac{1}{\sqrt{n}} \ln(t) \cdot \sigma(x) \)

where \( T \) = \( T \cdot \sigma(x) \).

(\( \alpha_{(x)} \)) \in \( \sigma(x) \).

Let \( \sigma(x) \) = \( \sigma(x) \).

Theorem: Under regularity conditions on \( \ln(t) \), \( T \) = \( T \), \( \sigma(x) \) = \( \sigma(x) \).

Let \( \sigma(x) = \sigma(x) \).

when \( \sigma(x) = \sigma(x) \).

Suppose \( T \) is large.

Need \( \sigma(x) \), \( \sigma(x) \) is large.

Thus, \( \sigma(x) \) is large.

Proof:

\[
\frac{1}{\sqrt{n}} \ln(t) \cdot \sum_{\omega} \sigma_0(\phi_\omega(x)) + \frac{1}{\sqrt{n}} \ln(t) \cdot \sum_{\omega} \sigma_0(\phi_\omega(x))
\]

\[\rightarrow \text{RMS w.r.t. SLLN.}\]
Office hours

(ReLU theorem from last time.)

Want \( S \subseteq \mathbb{R} \) s.t. \( v_j \in S \), \( \langle v_j, x \rangle \geq 0 \) \( \iff \langle w_j, x \rangle \geq 0 \)

Define \( S = \bigcup_{j \in \mathbb{Z}, m} : |v_{o,j}^T| \leq \tau \|x\| \) or \( |w_j - v_{o,j}^T\| \geq \tau \) or \( \|w_j - v_{o,j}^T\| \geq \tau \)

Consider \( j \notin S \Rightarrow \exists \theta : |v_{o,j}^T| > \tau \|x\| \) and \( \|w_j - v_{o,j}^T\| \leq \tau \) or \( \|w_j - v_{o,j}^T\| \leq \tau \)

Suppose \( v_{o,j}^T x > \tau \|x\| \)

\( v \) - done it as shown \( v_{o,j}^T x \geq 0 \). \( v_{o,j}^T x = \langle v_j - v_{o,j}^T + v_{o,j}^T, x \rangle \)

\( w \) - done it as shown \( w_{o,j} x \geq 0 \).

Otherwise \( v_{o,j}^T x < -\tau \|x\| \)

\( w \) - done it \( v_j^T x \leq 0 \).

\( w \) - done it \( w_{o,j}^T x \leq 0 \).