Uniform Convergence of Interpolators: Gaussian Width, Norm Bounds and Benign Overfitting. (2021, Koehler et al.)

이종진

Seoul National University

ga0408@snu.ac.kr

Jan 17, 2021

Contribution(1)

- Prove a generic uniform convergence guarantee on the generalization error of interpolators.
  - In high-dimensional linear regression with Gaussian data
  - With an arbitrary (Compact) hypothesis class
  - With the class's Gaussian width (and Gaussian radius)
- Norm based generalization bound
  - -> Used to analyze benign overfitting (minimal norm interpolators)

Contribution(2)

- Applying generic bound to Euclidean norm balls
  - -> Recover consistency result for Barlett et al. (2020) [1]
  - -> Confirms a prediction of Zhou et al. (2020)[2] for near-minimal-norm interpolators (in the case of Gaussian data)
- Applying generic bound to ℓ_1-norm
  - -> A novel consistency result for minimum ℓ_1- norm interpolators

[1] "Failures of model-dependent generalization bounds for least-nrom interpolation"

[2] "On Uniform Convergence and Low Norm Interpolation Learning"

Main results; summary

1. A Generic uniform convergence guarantee

$$\sup_{w \in \mathcal{K}, \hat{L}(w)=0} L(w) \leq \frac{1+\beta}{n} \left[W\left(\Sigma_2^{1/2}\mathcal{K}\right) + \left(\text{rad}\left(\Sigma_2^{1/2}\mathcal{K}\right) + \|w^*\|_{\Sigma_2}\right) \sqrt{2\log\left(\frac{32}{\delta}\right)} \right]^2$$

2. Compact set; General norm ball

$$\sup_{\|w\| \leq B, \hat{L}(w)=0} L(w) \leq (1+\gamma) \frac{\left(B \cdot \mathbb{E}\left\|\Sigma_2^{1/2} H\right\|_*\right)^2}{n}$$

3. General norm bound with minimum interpolator $\hat{w}$

$$\|\hat{w}\| \leq \|w^*\| + (1+\epsilon)^{1/2}\sigma \frac{\sqrt{n}}{\mathbb{E}\left\|\Sigma_2^{1/2} H\right\|_*}$$

4. Benign overfitting with general norm (bound $\to \sigma^2$ as $n \to \infty$)

$$L(\hat{w}) \leq (1+\gamma)(1+\epsilon)\left(\sigma + \|w^*\| \frac{\mathbb{E}\left\|\Sigma_2^{1/2} H\right\|_*}{\sqrt{n}}\right)^2$$

Data Model

- Data Model:

$$Y = Xw^* + \xi, \quad X_i \overset{iid}{\sim} N(0, \Sigma), \quad \xi \sim N\left(0, \sigma^2 I_n\right)$$

  where $X \in \mathbb{R}^{n \times d}$(i.i.d. , $d >> n$), w^* is arbitrary, and ξ is Gaussian and $X \perp \xi$.

- Papers results hold for
  - X_i should be Gaussian
  - ξ can be relaxed to be Sub-Gaussian

Minimum Norm Interpolator

- The population loss

$$L(w) = \mathop{\mathbb{E}}_{(x,y)} (y - \langle w, x \rangle)^2 = \sigma^2 + \|w - w^*\|_\Sigma^2$$

- For an arbitrary norm $\|\cdot\|$, the minimal norm interpolator is defined by

$$\hat{w} = \text{argmin}_{\hat{L}(w)=0} \|w\|$$

Gaussian Width and Gaussian radius

- The Gaussian width of a set $S \subset \mathbb{R}^d$:

$$W(S) := \mathop{\mathbb{E}}_{H \sim N(0, I_d)} \sup_{s \in S} |\langle s, H \rangle|$$

- The radius of a set $S \subset \mathbb{R}^d$:

$$\mathrm{rad}(S) := \sup_{s \in S} \|s\|_2$$

- For $\mathcal{K} = \{w : \|w\|_2 \leq B\}$,
  - $W\left(\Sigma^{1/2}\mathcal{K}\right) = B \cdot \mathop{\mathbb{E}}_{H \sim N(0, I_d)} \left\|\Sigma^{1/2} H\right\|_2$
  - $\mathrm{rad}\left(\Sigma^{1/2}\mathcal{K}\right) = \sup_{\|w\| \leq B} \|w\|_\Sigma$

Generic Uniform Convergece Guanratee

Theorem 1, Main generalization bound

$\exists C_1, C_1 \leq 66$ *such that the following is true.*

Let $\mathcal{K}$ *be an arbitrary compact set, take any covariance splitting* $\Sigma = \Sigma_1 \oplus \Sigma_2$, *fix* $\delta \leq 1/4$ *and let* $\beta = C_1 \left(\sqrt{\frac{\log(1/\delta)}{n}} + \sqrt{\frac{\operatorname{rank}(\Sigma_1)}{n}} \right)$.

If n is large enough that $\beta \leq 1$, *then the following holds with probability at least* $1 - \delta$ *:*

$$\sup_{w \in \mathcal{K}, \hat{L}(w)=0} L(w) \leq \frac{1+\beta}{n} \left[W\left(\Sigma_2^{1/2}\mathcal{K}\right) + \left(\operatorname{rad}\left(\Sigma_2^{1/2}\mathcal{K}\right) + \|w^*\|_{\Sigma_2}\right)\sqrt{2\log\left(\frac{32}{\delta}\right)} \right]^2 .$$

Effective ranks

- The dual norm of a norm $\|\cdot\|$ on $\mathbb{R}^d$

$$\|u\|_* := \max_{\|v\|=1} \langle v, u \rangle$$

And the set of all its sub-gradients with respect to u is

$$\partial \|u\|_* = \{v : \|v\| = 1, \langle v, u \rangle = \|u\|_*\}$$

- The effective $\|\cdot\|$-ranks of a covariance matrix Σ are given as follows

$$r_{\|\cdot\|}(\Sigma) = \left(\frac{\mathbb{E} \left\| \Sigma^{1/2} H \right\|_*}{\sup_{\|w\| \leq 1} \|w\|_\Sigma} \right)^2 \quad \text{and} \quad R_{\|\cdot\|}(\Sigma) = \left(\frac{\mathbb{E} \left\| \Sigma^{1/2} H \right\|_*}{\mathbb{E} \|v^*\|_\Sigma} \right)^2$$

where $H \sim N(0, I_d)$ and $v^* = \arg\min_{v \in \partial \left\| \Sigma^{1/2} H \right\|_*} \|v\|_\Sigma$

General norm

Corollary 3

$\exists C_1, C_1 \leq 66$ *such that the following is true.*

Take any covariance splitting $\Sigma = \Sigma_1 \oplus \Sigma_2$, *let* $\|\cdot\|$ *be an arbitrary norm, fix* $\delta \leq 1/4$ *and let* $\gamma = C_1 \left(\sqrt{\frac{\log(1/\delta)}{r_{\|\cdot\|}(\Sigma_2)}} + \sqrt{\frac{\log(1/\delta)}{n}} + \sqrt{\frac{\text{rank}(\Sigma_1)}{n}} \right)$. *If* $B \geq \|w^*\|$ *and n is large enough that* $\gamma \leq 1$, *then the following holds with probability at least* $1 - \delta$ *:*

$$\sup_{\|w\| \leq B, \hat{L}(w) = 0} L(w) \leq (1 + \gamma) \frac{\left(B \cdot \mathbb{E} \left\| \Sigma_2^{1/2} H \right\|_* \right)^2}{n}$$

General norm

Theorem 4

$\exists C_2, C_2 \leq 64$ such that the following is true.

Take any covariance split $\Sigma = \Sigma_1 \oplus \Sigma_2$, let $\|\cdot\|$ be an arbitrary norm, and fix $\delta \leq 1/4$. Denote P as ℓ_2 orthogonal projection matrix onto the space spanned by Σ_2 and let $H \sim N(0, I_d)$, and let $v^* = \arg\min_{v \in \partial \left\| \Sigma_2^{1/2} H \right\|_*} \|v\|_{\Sigma_2}$.
Suppose that there exist $\epsilon_1, \epsilon_2 \geq 0$ such that with probability at least $1 - \delta/4$

$$\|v^*\|_{\Sigma_2} \leq (1 + \epsilon_1)\, \mathbb{E}\, \|v^*\|_{\Sigma_2} \quad \text{and} \quad \|Pv^*\|^2 \leq 1 + \epsilon_2$$

let $\epsilon = C_2 \left(\sqrt{\frac{\log(1/\delta)}{r_{\|\cdot\|}(\Sigma_2)}} + \sqrt{\frac{\log(1/\delta)}{n}} + (1 + \epsilon_1)^2 \frac{n}{R_{\|\cdot\|}(\Sigma_2)} + \epsilon_2 \right)$.
Then if n and the effective ranks are large enough that $\epsilon \leq 1$, with probability at least $1 - \delta$, it holds that

$$\|\hat{w}\| \leq \|w^*\| + (1 + \epsilon)^{1/2} \sigma \frac{\sqrt{n}}{\mathbb{E} \left\| \Sigma_2^{1/2} H \right\|_*}$$

Benign overfitting with general norm

Theorem 5

Fix any $\delta \leq 1/2$. let $\|\cdot\|$ be an arbitrary norm and pick a covariance split $\Sigma = \Sigma_1 \oplus \Sigma_2$. Suppose that n and the effective ranks are sufficiently large such that $\gamma, \epsilon \leq 1$ with the same choice of γ and ϵ as in Corollary 3 and Theorem 4. Then, with probability at least $1-\delta$,

$$L(\hat{w}) \leq (1+\gamma)(1+\epsilon)\left(\sigma + \|w^*\| \frac{\mathbb{E}\left\|\Sigma_2^{1/2} H\right\|_*}{\sqrt{n}}\right)^2 .$$

Sufficient conditions

- If there exists a sequence of covariance splits $\Sigma = \Sigma_1 \oplus \Sigma_2$ such that

$$\frac{\operatorname{rank}(\Sigma_1)}{n} \to 0, \quad \frac{\|w^*\| \, \mathbb{E} \left\| \Sigma_2^{1/2} H \right\|_*}{\sqrt{n}} \to 0, \quad \frac{1}{r_{\|\cdot\|}(\Sigma_2)} \to 0, \quad \frac{n}{R_{\|\cdot\|}(\Sigma_2)} \to 0$$

then, $L(\hat{w})$ converges in probability to σ^2 as $n \to \infty$

The end

Appendix; Euclidean Norm Ball

- $\mathcal{K} = \{w \in \mathbb{R}^d : \|w\|_2 \leq B\}$

Corollary 1, Proof of the speculative bound ($\star$) for Gaussian data

Fix any $\delta \leq 1/4$. *with* $B \geq \|w^*\|_2$ *and* $n \gtrsim \log(1/\delta)$, *for some* $\gamma \lesssim \sqrt[4]{\log(1/\delta)/n}$, *it holds with probability at least* $1 - \delta$ *that*

$$\sup_{\|w\|_2 \leq B, \hat{L}(w)=0} L(w) \leq (1+\gamma)\frac{B^2 \operatorname{Tr}(\Sigma)}{n}$$

Appendix; Euclidean Norm Ball

Theorem 2, Euclidean norm bound; special case of Theorem 4

Fix any $\delta \leq 1/4$*. Under the model assumptions in (1) with any choice of covariance splitting* $\Sigma = \Sigma_1 \oplus \Sigma_2$*, there exists some* $\epsilon \lesssim \sqrt{\frac{\log(1/\delta)}{r(\Sigma_2)}} + \sqrt{\frac{\log(1/\delta)}{n}} + \frac{n \log(1/\delta)}{R(\Sigma_2)}$ *such that the following is true. If n and the effective ranks are such that* $\epsilon \leq 1$ *and* $R(\Sigma_2) \gtrsim \log(1/\delta)^2$*, then with probability at least* $1 - \delta$*, it holds that*

$$\|\hat{w}\|_2 \leq \|w^*\|_2 + (1+\epsilon)^{1/2} \sigma \sqrt{\frac{n}{\operatorname{Tr}(\Sigma_2)}}$$

Corollary 2

There exists an absolute constant $C_1 \leq 66$ *such that the following is true. Pick any split* $\Sigma = \Sigma_1 \oplus \Sigma_2$, *fix* $\delta \leq 1/4$, *and let* $\gamma = C_1 \left(\sqrt{\frac{\log(1/\delta)}{r(\Sigma_2)}} + \sqrt{\frac{\log(1/\delta)}{n}} + \sqrt{\frac{\operatorname{rank}(\Sigma_1)}{n}} \right)$. *If* $B \geq \|w^*\|_2$ *and* n *is large enough that* $\gamma \leq 1$, *the following holds with probability at least* $1 - \delta$ *:*

$$\sup_{\|w\|_2 \leq B, \hat{L}(w)=0} L(w) \leq (1+\gamma) \frac{B^2 \operatorname{Tr}(\Sigma_2)}{n}.$$

Appendix; Euclidean Norm Ball

Theorem 3, Benign overfitting

Theorem 3 (Benign overfitting). Fix any $\delta \le 1/2$*. With any covariance splitting* $\Sigma = \Sigma_1 \oplus \Sigma_2$*, let* γ *and* ϵ *be as defined in Corollary 2 and Theorem 2. Suppose that* n *and the effective ranks are such that* $R(\Sigma_2) \gtrsim \log(1/\delta)^2$ *and* $\gamma, \epsilon \le 1$*. Then, with probability at least* $1 - \delta$*,*

$$L(\hat{w}) \le (1+\gamma)(1+\epsilon)\left(\sigma + \|w^*\|_2 \sqrt{\frac{\mathrm{Tr}(\Sigma_2)}{n}}\right)^2 .$$

Appendix; Euclidean Norm Ball

- Sufficient conditions for consistency of $\hat{w}$. As $n \to \infty$, $L(\hat{w})$ converges in probability to σ^2 if there exists a sequence of covariance splits $\Sigma = \Sigma_1 \oplus \Sigma_2$ such that

$$\frac{\operatorname{rank}(\Sigma_1)}{n} \to 0, \quad \|w^*\|_2 \sqrt{\frac{\operatorname{Tr}(\Sigma_2)}{n}} \to 0, \quad \frac{n}{R(\Sigma_2)} \to 0$$

Appendix; ℓ_1 Norm Balls for Basis Pursuit

- Sufficient conditions for consistency of $\hat{w}_{BP}$. As $n \to \infty$, $L(\hat{w})$ converges to σ^2 in probability if there exists a sequence of covariance splits $\Sigma = \Sigma_1 \oplus \Sigma_2$ such that Σ_2 is diagonal and

$$\frac{\operatorname{rank}(\Sigma_1)}{n} \to 0, \quad \frac{\|w^*\|_1 \, \mathbb{E}\left\|\Sigma_2^{1/2} H\right\|_\infty}{\sqrt{n}} \to 0, \quad \frac{n}{r_1(\Sigma_2)} \to 0$$