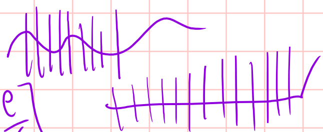
Lecture 4: Infinite-width apx over $\mathbb{R}^d$ via Fourier.

Announcements.

* HW1 out.
* Room switch

1. short questions
   - (A) } apx setup
   - (B)
   - (C)
   - (d)
2. 

   $\lceil \frac{e}{\varepsilon} \rceil$

   adaptive complexity

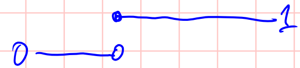
3. NTK ←
4. univariate apx
   not a polynomial

   $\boxed{\text{Leshno et al. '93}}$

$$\mathbb{1}[P] = \begin{cases} 1 & P \text{ true} \\ 0 & P \text{ false} \end{cases}$$

e.g., $\lceil \frac{e}{\varepsilon} \rceil$ univariate apx,

$$\sum_i c_i \underbrace{\mathbb{1}[x - b_i \geq 0]}$$



lec 4 infinite-width apx over $\mathbb{R}^d$ — $a^T\sigma(Ux+b) \to \sum_j a_j \exp(v_j^T x + b_j)$

$= \sum_j \underbrace{a_j e^{b_j}} \exp(v_j^T x)$

3-layer

lec 5-8 near initialization/overparametrization (NTK)

lec 9-? apx

lec 12 opt

Infinite-width

* popular again
* representation with equalities; maybe can be less loose?

Last time: $g(x) - g(0) = \int_0^1 g'(b)\,\mathbb{1}[x \ge b]\,db$

where "complexity"/"weight mass" was $\left(\int_0^1 |g'(b)|\,db\right)^2$

[HW1 P2 gets roughly same complexity but as a direct finite-width construction]

Today infinite width over $\mathbb{R}^d$ via Fourier.

Goal: $g(x) - g(0) = \iint \underbrace{g(w,b)}\,dw\,db$ based on Fourier transform

where $\iint |g(w,b)|\,dw\,db \le \boxed{\int \|\nabla \widetilde{g}(w)\|\,dw}$

Remark (adaptive complexity).

* Barron '93 has many estimates for $\int \|\widetilde{\nabla} g\|$; flattish gaussians get poly(d); radial functions in general seem to get $\int \|\widetilde{\nabla g}\| \approx \exp(d)$.
* Homework uses $\int |g'|$ ("bounded variation); can abstract into norms of higher order partials; e.g., Sobolev Besov spaces [reference in notes.]
* Maybe > 2 layers or other representation
* For fixed target functions can still be hard to get good estimates; e.g., for sparse parity.
* Should algorithm be ignored? E.g., seems GD (a) stays close to initialization, (b) prefers low norm solutions.

Lemma (Fourier transforms (Folland)). Assume $\int |g| < \infty$

$$\tilde{g}(\omega) = \int \exp(-2\pi i \omega^T x)\, g(x)\, dx$$

(1) (Inversion.) If $\int |\tilde{g}| < \infty$ then

$$g(x) = \int \exp(2\pi i \omega^T x)\, \tilde{g}(\omega)\, d\omega.$$

[complex exponential activation, infinite width network.]

(2) (Derivative.) $2\pi \|\omega\| \cdot |\tilde{g}(\omega)| = \|\widetilde{\nabla g}\|$.

(3) (Euler formula.)

$$\exp(i\omega) = \cos(\omega) + i \sin(\omega).$$

(4) (Polar form) given g, $\exists \theta$, $|\theta| \le 1$

$$\tilde{g}(\omega) = |\tilde{g}(\omega)| \cdot \exp(2\pi i\, \theta(\omega))$$

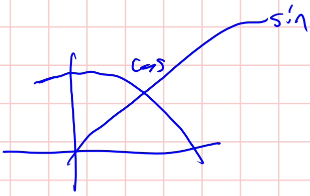
Start from inversion formula: $g(x) = \int \exp(2\pi i \omega^T x)\, \tilde{g}(\omega)\, d\omega$

(1) Remove complex numbers.

polar part of $\tilde{g}$

$$\begin{aligned} g(x) &= \mathrm{Re}[g(x)] = \mathrm{Re}\left[\int \exp(2\pi i \omega^T x)\, \tilde{g}(\omega)\, d\omega\right] \\ &= \int \mathrm{Re}\left[\exp(2\pi i \omega^T x)\, \tilde{g}(\omega)\right] d\omega = \int \mathrm{Re}\left[\exp(2\pi i \omega^T x)\, \exp(2\pi i\, \theta(\omega))\, |\tilde{g}(\omega)|\right] d\omega \\ &= \int \mathrm{Re}\left[\exp(2\pi i (\omega^T x + \theta(\omega)))\right] |\tilde{g}(\omega)|\, d\omega \\ &= \int \mathrm{Re}\left[\cos(2\pi(\omega^T x + \theta(\omega))) + i \cdot \sin(2\pi(\omega^T x + \theta(\omega)))\right] |\tilde{g}(\omega)|\, d\omega. \\ &= \int \cos(2\pi(\omega^T x + \theta(\omega)))\, |\tilde{g}(\omega)|\, d\omega. \end{aligned}$$

(2) Introduce thresholds; just like univariate apx, $\|x\| \le 1$



$$\begin{aligned} &\cos(2\pi(\omega^T x + \theta(\omega))) - \cos(2\pi\, \theta(\omega)) \\ &= -2\pi \int_0^{\omega^T x} \sin(2\pi(b + \theta(\omega)))\, db \\ &= -2\pi \int_0^{\|\omega\|} \sin(2\pi(b + \theta(\omega)))\, \mathbb{1}[\omega^T x \ge b]\, db \\ &\quad + 2\pi \int_{-\|\omega\|}^{0} \sin(2\pi(-b + \theta(-\omega)))\, \mathbb{1}[\omega^T x \ge b]\, db. \end{aligned}$$

Theorem. If $\int |g| < \infty$ & $\int |\tilde{g}| < \infty$, then

$$g(x) - g(0) = \iint g(\omega, b)\, \mathbb{1}[\omega^T x - b \ge 0]\, d\omega\, db$$

where $g(b,\omega) = +2\pi\, \mathbb{1}[b \in [0, \|\omega\|]]\, |\tilde{g}(\omega)| \left[-\sin(2\pi(+b + \theta(\omega))) + \sin(2\pi(-b + \theta(-\omega)))\right]$

and $\iint |g(b,\omega)|\, d\omega\, db \le 2 \int \|\widetilde{\nabla g}(\omega)\|\, d\omega.$

Proof

$$\begin{aligned} \iint |g(b,\omega)|\, db\, d\omega &\le \int \int_0^{\|\omega\|} 2\pi\, |\tilde{g}(\omega)| \cdot 2\, db\, d\omega = 2\int 2\pi \cdot \|\omega\| \cdot |\tilde{g}(\omega)|\, d\omega \\ &= 2 \int \|\widetilde{\nabla g}(\omega)\|\, d\omega. \end{aligned}$$

//