Topics in Applied Mathematics: Infinitely Large Neural Networks, 3341.751
E. Ryu
Spring 2022

Homework 2
Due 5pm, Friday, April 1, 2022

Problem 1: Show that the quantitative approximation guarantee can be improved to

$$\|f_\theta - f_\star\|_{L^2(B)}^2 \le \frac{4Q^2B^2}{N}.$$

Hint. Start by slightly sharpening the Maurice lemma. (The proof of the Maurice lemma given in the recorded lecture had a minor error. Refer to the updated proof.)

Problem 2: Let $\Omega \subseteq \mathbb{R}^d$ be compact and $\sigma \in \mathcal{C}(\mathbb{R})$ non-polynomial. Show that the set of 3-layer neural networks of the form

$$f_\theta(x) = A^{(3)}\sigma(A^{(2)}\sigma(A^{(1)}x + b^{(1)}) + b^{(2)}) + b^{(3)},$$

where σ applies elementwise, $N_1, N_2 \in \mathbb{N}$, $A^{(1)} \in \mathbb{R}^{N_1\times d}$, $b^{(1)} \in \mathbb{R}^{N_1}$, $A^{(2)} \in \mathbb{R}^{N_2\times N_1}$, $b^{(2)} \in \mathbb{R}^{N_2}$, $A^{(3)} \in \mathbb{R}^{1\times N_2}$, and $b^{(3)} \in \mathbb{R}^1$ is dense in $(\mathcal{C}(\Omega), \|\cdot\|_\infty)$.

Problem 3: Let $\sigma\colon \mathbb{R} \to \mathbb{R}$ be a continuous function such that there is a $R \in (0,\infty)$ such that

$$\begin{aligned} \sigma(r) &= 0, \qquad \forall\, r \le -R \\ \sigma(r) &= 1, \qquad \forall\, r \ge R. \end{aligned}$$

Let

$$A = \begin{bmatrix} a_1^\intercal \\ a_2^\intercal \\ \vdots \\ a_m^\intercal \end{bmatrix} \in \mathbb{R}^{m\times d}, \qquad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \in \mathbb{R}^m,$$

and assume the polytope $P = \{x \,|\, Ax \le b\} \subset \mathbb{R}^d$ is compact. Show that 3-layer neural networks of the form

$$f_\theta(x) = A^{(3)}\sigma(A^{(2)}\sigma(A^{(1)}x + b^{(1)}) + b^{(2)}) + b^{(3)},$$

where σ applies elementwise, $N_1, N_2 \in \mathbb{N}$, $A^{(1)} \in \mathbb{R}^{N_1\times d}$, $b^{(1)} \in \mathbb{R}^{N_1}$, $A^{(2)} \in \mathbb{R}^{N_2\times N_1}$, $b^{(2)} \in \mathbb{R}^{N_2}$, $A^{(3)} \in \mathbb{R}^{1\times N_2}$, and $b^{(3)} \in \mathbb{R}^1$ can approximate $\mathbf{1}_{\{x \,|\, Ax\le b\}} \in L^p$ for all $p \in [1,\infty)$.

Clarification. We are considering the L^p space with respect to the Lebesgue measure.
Hint. Define

$$d(x,P) = \begin{cases} 0 & \text{if } Ax \le b \\ \max_{i=1,\dots,m} a_k^\intercal x - b & \text{otherwise.} \end{cases}$$

Show that (i) $d(x,P) > 0$ if and only if $x \notin P$ and (ii) there is an $M \in (0,\infty)$ such that $d(x,P) \ge 1$ if $\|x\| \ge M$.

Problem 4: Let $\mathcal{X} = (-1, 1)$. Show that $K\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ defined as

$$K(x, x') = \frac{1}{1 - xx'}$$

is PDK.

Problem 5: Let $\mathcal{Z}$ be some set and let

$$\mathcal{X} = \{A \subseteq \mathcal{Z} \mid |A| < \infty\}.$$

Show that $K\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ defined as

$$K(A, A') = 2^{|A \cap A'|}$$

is a PDK.

Clarification. $|A|$ denotes the cardinality of the set A.
Hint. Consider

$$\sum_{S \in A \cup A'} \mathbf{1}_{S \subseteq A}(A) \mathbf{1}_{S \subseteq A}(A').$$

Problem 6: Let $\mathcal{Z}$ be a topological space, let μ be a nonnegative finite Borel measure on $\mathcal{Z}$, and let

$$\mathcal{X} = \{A \subseteq \mathcal{Z} \mid A \text{ is Borel measurable}\}.$$

Show that $K\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ defined as

$$K(A, B) = \mu(A \cap B)$$

is a PDK.

Hint. Consider the feature map $\phi(A) = \mathbf{1}_A \in L^2(\mu)$.

Problem 7: *Simple RKHS facts.* Let $\mathcal{H}$ be an RKHS of functions defined on $\mathcal{X}$ with RK $K\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. Show the following.

(i) $K(x, x) \geq 0$ for all $x \in \mathcal{X}$.

(ii) If $f_k \to f_\infty$ in $\mathcal{H}$, then $f_k(x) \to f_\infty(x)$ for all $x \in \mathcal{X}$.

(iii) Define $d_K\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ as

$$d_K(x, x') = \|K(x, \cdot) - K(x', \cdot)\|_{\mathcal{H}}.$$

Then d_K is a pseudometric on $\mathcal{X}$. If K further is strictly positive definite, then d_K is a metric on $\mathcal{X}$.

(iv) If $K(x, x) = 0$, then $K(x, x') = 0$ for all $x' \in \mathcal{X}$.

(v) The *normalized kernel* $\tilde{K}\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ defined as

$$\tilde{K}(x, x') = \begin{cases} \frac{K(x,x')}{\sqrt{K(x,x)K(x',x')}} & \text{if } K(x, x)K(x', x') > 0 \\ 0 & \text{otherwise} \end{cases}$$

is a PDK.