Foundations of Deep Learning **Fall 2024**

Exercise 2: Approximation Theory

Lecturer: Aurelien Lucchi

Problem 1 (Weierstrass theorem):

In this exercise, we seek to derive a formal proof of the Weierstrass theorem discussed in the lecture. Recall that the theorem can be stated as follows:

Theorem 1 (Weierstrass)**.** *Given a continuous function $f(x)$ on $a \leq x \leq b$ and an arbitrary positive constant $\epsilon > 0$, it is possible to construct an approximating polynomial $P(x)$ such that*

$$|f(x) - P(x)| \leq \epsilon, \ \forall a \leq x \leq b \tag{1}$$

Without loss of generality, we assume $a = 0, \ b = 1$ and extend f on the whole $\mathbb{R}$ by setting $f(x) = 0, \ \forall x \notin (0, 1]$, such that f is continuous on $(-\infty, 1]$. (Why can we do this?)

Let $P_n(x)$ be a polynomial of degree $2n$ such that

$$P_n(x) = \frac{1}{J_n} \int_0^1 f(t)[1 - (t - x)^2]^n dt, \tag{2}$$

where $J_n := \int_{-1}^1 (1 - u^2)^n du$ is a constant .

a) Show that

$$f(x) = \frac{1}{J_n} \int_{-1}^1 f(x)(1 - u^2)^n du \tag{3}$$

b) Show that

$$P_n(x) - f(x) = \frac{1}{J_n} \int_{-1}^1 [f(x + u) - f(x)](1 - u^2)^n du \tag{4}$$

where $x \in [0, 1]$. The problem is now to show that this expression approaches zero as $n \to \infty$.

c) Let $\epsilon > 0$. Use the following facts freely:

- since $f(x)$ is continuous on $[-1, 1]$, there exists a $\delta > 0$ such that $|f(x + u) - f(x)| \leq \frac{\epsilon}{2}$ for each x, u with $|u| < \delta$ and $x, x + u \in [-1, 1]$;
- there exist a positive constant $M > 0$, such that $|f(x)| \leq M, \ \forall x \in [-1, 1]$.

Show that

$$|f(x + u) - f(x)| \leq \frac{\epsilon}{2} + 2M\frac{u^2}{\delta^2}, \ \forall x \in [-1, 1]. \tag{5}$$

Hint: Think of the case distinction where $|u| \geq \delta$, i.e. $1 \leq \frac{u^2}{\delta^2}$; and where $|u| < \delta$.

d) Using integration by part, show that

$$J'_n := \int_{-1}^1 u^2(1 - u^2)^n du = \frac{J_{n+1}}{2(n + 1)}$$

and also show that

$$J_n > J_{n+1}, \ \forall n \in \mathbb{N}.$$

e) Finally, re-using the answers to the previous sub-problems, prove that

$$|f(x) - P_n(x)| \leq \epsilon, \ \forall x \in [0, 1] \tag{6}$$

for sufficiently large n.

Problem 2 (Stone-Weierstrass theorem):
The following is a generalization of Weierstrass theorem.

Theorem 2 (Stone-Weierstrass, see Theorem 2.2. in Telgarsky (2021))**.** *Let function class $\mathcal{F}$ of real-valued functions defined on $[0,1]^d$ be given as follows:*

i) Each $f \in \mathcal{F}$ is continuous.

ii) For every $\mathbf{x} \in [0,1]^d$, there exists $f \in \mathcal{F}$ with $f(\mathbf{x}) \neq 0$.

iii) For every $\mathbf{x} \neq \mathbf{x}' \in [0,1]^d$, there exists $f \in \mathcal{F}$ with $f(\mathbf{x}) \neq f(\mathbf{x}')$ ($\mathcal{F}$ separates points).

iv) $\mathcal{F}$ is closed under multiplication and vector space operations, i.e. $\mathcal{F}$ is an algebra.

Then $\mathcal{F}$ is a universal approximator: for every continuous $g : [0,1]^d \to \mathbb{R}$ and $\epsilon > 0$, there exists $f \in \mathcal{F}$ with

$$|f(\mathbf{x}) - g(\mathbf{x})| \leq \epsilon, \ \forall \mathbf{x} \in [0,1]^d.$$

a) Consider unbounded width networks with one hidden layer:

$$\mathcal{F}_{\sigma,d,m} := \mathcal{F}_{d,m} := \left\{x \mapsto a^\top \sigma(Wx+b) : a \in \mathbb{R}^m, W \in \mathbb{R}^{m \times d}, b \in \mathbb{R}^m\right\}. \tag{7}$$

$$\mathcal{F}_{\sigma,d} := \mathcal{F}_d := \bigcup_{m \geq 0} \mathcal{F}_{\sigma,d,m}. \tag{8}$$

Using Theorem 2, show that $\mathcal{F}_{\cos,d}$ is universal, where $\cos : \mathbb{R} \to \mathbb{R}$ is the cosine function.

b) Show that Theorem 2 does not hold if

i) condition (i) in Theorem 2 does not hold;
ii) condition (ii) in Theorem 2 does not hold;
iii) condition (iii) in Theorem 2 does not hold;
iv) condition (iv) in Theorem 2 does not hold.

References

Matus Telgarsky. Deep learning theory lecture notes. `https://mjt.cs.illinois.edu/dlt/`, 2021. Version: 2021-10-27 v0.0-e7150f2d (alpha).