Lecture 6, Friday, April 19, 2024

- Finish the proof of a theorem: σ is universal $\Leftrightarrow$ σ is not a polynomial.
- Univ. Approximation Thms for L^p-approximations.

<u>Theorem 2</u> (Leshno et al. 1993) Let $\sigma: \mathbb{R} \to \mathbb{R}$ be Lebesgue measurable, bounded on bounded sets. Assume that $m(\overline{D(\sigma)}) = 0$. Let $d \in \mathbb{N}$ and $K \subseteq \mathbb{R}^d$ compact. Then for any $f \in C(K)$, there exist $\Phi_n \in \mathcal{N}_{\sigma,d} := \mathcal{NN}(\sigma, L=2, N_0=d, N_L=1)$ $(n=1,2,\cdots)$ such that $\|\Phi_n - f\|_{L^\infty(K)} \to 0$ if and only if σ is not a polynomial m-a.e. on $\mathbb{R}$.

We have proved several lemmas.

<u>Lemma 1</u> Let $\sigma: \mathbb{R} \to \mathbb{R}$ be Lebesgue-measurable. Assume that σ is a polynomial m-a.e. on $\mathbb{R}$. Then for any $d \in \mathbb{N}$, any compact subset $K \subseteq \mathbb{R}^d$, and any $p \in [1, \infty]$,

$$\mathcal{N}_{\sigma,d} := \mathcal{NN}(\sigma, L=2, N_0=d, N_L=1) \subseteq L^p(K),$$

$\mathcal{N}_{\sigma,d}$ is not dense in $L^p(K)$ (w.r.t. the $L^p(K)$-norm). Moreover, $C(K) \nsubseteq \overline{\mathcal{N}_{\sigma,d}}$ w.r.t. to the $L^\infty(K)$-norm.

<u>Lemma 2</u> (Reduction to one-dimensional approximations). Let $\sigma \in L^\infty_{loc}(\mathbb{R})$. Suppose for any compact $H \subseteq \mathbb{R}$, $C(H) \subseteq \overline{\mathcal{N}_\sigma}$ w.r.t. the $L^\infty(H)$-norm. ($\mathcal{N}_\sigma := \mathcal{N}_{\sigma,1}$). Then, for any $d \in \mathbb{N}$ and any $K \subseteq \mathbb{R}^d$: compact. $C(K) \subseteq \overline{\mathcal{N}_{\sigma,d}}$ w.r.t. the $L^\infty(K)$-norm.

<u>Lemma 3</u> If $\sigma \in C^\infty(\mathbb{R})$ and σ is not a polynomial. Then $\mathcal{N}_\sigma$ is dense in $C(K)$ for any compact $K \subseteq \mathbb{R}$.

<u>Lemma 4</u> Let $\sigma: \mathbb{R} \to \mathbb{R}$ be Lebesgue measurable and bounded on bounded sets. Assume $m(\overline{D(\sigma)}) = 0$. Let $\varphi \in C_c^\infty(\mathbb{R})$. Then for any $K \subseteq \mathbb{R}$: compact, $\sigma * \varphi \in \overline{\mathcal{N}_\sigma}$ w.r.t. the $L^\infty(K)$-norm.

<u>Lemma 5</u> Let $\sigma: \mathbb{R} \to \mathbb{R}$ be Lebesgue measurable and bounded on bounded sets. Assume $m(\overline{D(\sigma)}) = 0$. Let $\varphi \in C_c^\infty(\mathbb{R})$. Assume $\sigma * \varphi$ is not a polynomial. Then $C(K) \subseteq \overline{\mathcal{N}_\sigma}$ w.r.t. the $L^\infty(K)$-norm for any compact $K \subseteq \mathbb{R}$.

Lemma 6 Let $\sigma: \mathbb{R} \to \mathbb{R}$ be a measurable function and $\sigma \in L^1_{loc}(\mathbb{R})$. If for any $\varphi \in C_c^\infty(\mathbb{R})$, $\sigma * \varphi$ is a polynomial, then σ is a polynomial m-a.e. $\mathbb{R}$.

Proof of Theorem 2 If for any $f \in C(K)$, $\exists\, \Phi_n \in \mathcal{N}_\sigma$ $(n = 1, 2, \cdots)$ s.t. $\|\Phi_n - f\|_{L^\infty(K)} \to 0$, then by Lemma 1, σ is not a polynomial m-a.e. $\mathbb{R}$.

Assume σ is not a polynomial m-a.e. $\mathbb{R}$. By Lemma 2, we need only to show that for any $K \subseteq \mathbb{R}$ compact, $C(K) \subseteq \overline{\mathcal{N}_\sigma}$ w.r.t. $L^\infty(K)$-norm. By Lemma 6, $\exists\, \varphi \in C_c^\infty(\mathbb{R})$ such that $\sigma * \varphi$ is not a polynomial. By Lemma 5, $C(K) \subseteq \overline{\mathcal{N}_\sigma}$ w.r.t. $L^\infty(K)$-norm. QED

Universal Approximations w.r.t. L^p-norms

Consider a (positive) Borel measure μ on $\mathbb{R}^d$. We ask if $\mathcal{N}_{\sigma,d}$ can approximate $L^p(\mu)$-functions, locally, for $1 \le p < \infty$. Specifically, $\forall f \in L^p(\mu)$, are there $\Phi_n \in \mathcal{N}_{\sigma,d}$ s.t.

$$\|\Phi_n - f\|_{L^p(\mu)} = \left(\int_{\mathbb{R}^d} |\Phi_n - f|^p d\mu \right)^{1/p} \to 0 \ ?$$

Integration on $\mathbb{R}^d$ is often too much, requiring more properties for μ. So, often we consider $\int_A$ instead of $\int_{\mathbb{R}^d}$, for some bounded Borel set A.

Notation

$\mathcal{M}_c(\mathbb{R}^d) = \{$all compactly supported (positive) Radon measures on $\mathbb{R}^d\}$.

Remarks

- A Radon measure on $\mathbb{R}^d$ is a (positive) Borel measure on $\mathbb{R}^d$ that is finite on any compact subset of $\mathbb{R}^d$.
- Examples
  - compactly supported finite Borel measures, in particular, compactly supported Borel probability measures.
  - The Lebesgue measure m $(= m_d)$ on $\mathbb{R}^d$ is not compactly supported. But for any bounded Borel set $A \subseteq \mathbb{R}^d$, the restriction $m \llcorner A$ (defined by $(m \llcorner A)(B) = m(A \cap B)$ $\forall B$)

is compactly supported: $m \llcorner A \in \mathcal{M}_c(\mathbb{R}^d)$.

— If $u \in L^1(\mathbb{R}^d) = L^1(\mathbb{R}^d, m)$, $u \geq 0$, then $d\mu = u\,dm$ defines a finite (positive) Borel measure on $\mathbb{R}^d$, which is absolutely continuous w.r.t. m. Note

$$\mu(B) = \int_B f\,d\mu \quad \forall B \subseteq \mathbb{R}^d, \text{ Borel}.$$

If $\operatorname{supp}(u)$ is compact then $\mu \in \mathcal{M}_c(\mathbb{R}^d)$.

— Let $y_1, \cdots, y_n \in \mathbb{R}^d$, distinct, and $a_1, \cdots, a_n \in [0, \infty)$. Define $\mu = \sum_{j=1}^n a_j \delta_{y_j}$; Then $\mu \in \mathcal{M}_c(\mathbb{R}^d)$, $\operatorname{supp}(\mu) = \{y_1, \cdots, y_n\}$.

$$\mu(B) = \sum_{j=1}^n a_j \delta_{y_j}(B) \quad \forall B \subseteq \mathbb{R}^d, \text{ Borel}.$$

Here, δ_y is the Dirac mass at y: $\delta_y(B) = 1$ if $y \in B$, $\delta_y(B) = 0$ if $y \notin B$. So, $\mu(\mathbb{R}^d) = \sum_{j=1}^n a_j$.

Let μ be a (positive) Radon measure on $\mathbb{R}^d$. Let $1 \leq p \leq \infty$. The space $L^p(\mu)$ is defined by

$$L^p(\mu) = \{u : \mathbb{R}^d \to \mathbb{R} : u \text{ is Borel measurable, and } \|u\|_p < \infty\}.$$

Here,

$$\|u\|_p = \begin{cases} \left(\int_{\mathbb{R}^d} |u(x)|^p \, d\mu(x) \right)^{1/p} & \text{if } 1 \leq p < \infty, \\ \operatorname*{ess\,sup}_{x \in \mathbb{R}^d} |u(x)| & \text{if } p = \infty. \end{cases}$$

Same remarks

⊙ Two functions $u, v \in L^p(\mu)$ are the same means that $u(x) = v(x)$ μ-a.e. $x \in \mathbb{R}^d$ (μ-a.e. = almost everywhere w.r.t. μ).

If $u \in L^\infty(\mu)$ then $\|u\|_\infty$ is the smallest number s.t. $|u(x)| \leq \|u\|_\infty$ μ-a.e. $\mathbb{R}^d$.

⊙ $L^p(\mu)$ is a Banach space. If $1 \leq p < \infty$ then it is separable. If $p = 2$ then $L^2(\mu)$ is a Hilbert space with the inner product

$$\langle u, v \rangle = \int_{\mathbb{R}^d} u(x) v(x) \, d\mu(x) \quad \forall u, v \in L^2(\mu).$$

If $1 \leq p_1 \leq p_2 \leq \infty$ then $L^{p_2}(\mu) \subseteq L^{p_1}(\mu)$.

⊙ For $1 \leq p < \infty$, $f_n \to f$ in $L^p(\mu)$ means that

$$\int_{\mathbb{R}^d} |f_n - f|^p \, d\mu \to 0 \quad \text{as } n \to \infty.$$

For $p=\infty$, $f_n \to f$ in $L^\infty(\mu)$ means that, $\exists A_0 \subseteq \mathbb{R}^d$, Borel $\mu(A_0)=0$, such that $f_n(x) \to f(x)$ uniformly for $x \in \mathbb{R}^d \setminus A_0$.

⊙ For any p: $1 \le p \le \infty$. Simple functions $\sum_{j=1}^{N} \alpha_j \chi_{E_j}$; $E_j \subseteq \mathbb{R}^d$, Borel, $\alpha_j \in \mathbb{R}$, are dense in $L^p(A)$.

Any $L^p (1 \le p < \infty)$ function can be approximated by smooth and compactly supported functions on $\mathbb{R}^d$.

⊙ The dual space $[L^p(\mu)]^*$ for $1 \le p < \infty$ is $L^q(\mu)$, where $q = \frac{p}{p-1}$ if $p > 1$ and $q = \infty$ if $p = 1$ $\left(\frac{1}{p} + \frac{1}{q} = 1\right)$. This means that for any $F \in [L^p(\mu)]^*$ (i.e., F is a bounded linear functional on $L^p(\mu)$), $\exists!\ g \in L^q(\mu)$ s.t.

$$F(f) = \int_{\mathbb{R}^d} f(x) g(x)\, d\mu(x) \quad \forall f \in L^p(\mu).$$

Moreover, $\|F\| = \|g\|_q$.

In general, $[L^\infty(\mu)]^* \supsetneqq L^1(\mu)$. Rather, $[L^\infty(\mu)]^*$ = the space of all bounded, additive, signed measures that are absolutely continuous w.r.t. μ.

<u>Questions</u> Can a three-layer NN approximate L^p functions? What are the conditions on σ for such approximations?

Recall $\mathcal{N}_{\sigma,d} = \mathcal{NN}(\sigma, L=2, N_0=d, N_L=1)$ consists of functions

$$\Phi(x) = \sum_{j=1}^{N} \alpha_j \sigma(w_j \cdot x + \theta_j) + \beta \quad (x \in \mathbb{R}^d),$$

where $\alpha_j, \theta_j, \beta \in \mathbb{R}$ and $w_j \in \mathbb{R}^d$. To approximate L^p functions by such functions, we naturally require that $\sigma \in L^p(\mathbb{R}^1)$. So, the "universality" of σ may depend on $d \in \mathbb{N}$, μ, and p. But if we consider compactly supported $\mu \in \mathcal{M}(\mathbb{R}^d)$, then such dependence may not be important.

<u>Theorem</u> Let $\sigma \in C(\mathbb{R})$ and assume it is not a polynomial. Then for any $d \in \mathbb{N}$, any $\mu \in \mathcal{M}_c(\mathbb{R}^d)$, and any p with $1 \le p < \infty$, $\mathcal{N}_{\sigma,d}$ is dense in $L^p(\mu)$.

<u>Proof</u> Let $f \in L^p(\mu)$ and $\varepsilon > 0$. There exists $\varphi \in C_c(\mathbb{R}^d)$ s.t. $\|\varphi - f\|_p < \frac{\varepsilon}{2}$.

We know that σ is universal w.r.t. uniform approximations. Thus, $\exists\, \Phi \in \mathcal{N}_{\sigma,d}$ s.t. $\|\Phi - \varphi\|_C < \alpha\varepsilon$ for $\alpha = (\frac{1}{2})[\mu(K)]^{-\frac{1}{p}}$ with $K = \text{supp}(\mu)$. We have then $\|\Phi - \varphi\|_{L^p(\mu)} < \alpha\varepsilon\,[\mu(A)]^{1/p} = \varepsilon/2$, and $\|\Phi - f\|_{L^p(\mu)} \le \varepsilon$. QED

Definition Let $\sigma: \mathbb{R} \to \mathbb{R}$ be Borel measurable. Assume σ is bounded on any bounded subset of $\mathbb{R}$. We say σ is L^p-universal, if $\mathcal{N}_{\sigma,d}$ is dense in $L^p(\mu)$ for any $d \in \mathbb{N}$, any $\mu \in \mathcal{M}_c(\mathbb{R}^d)$, any $p \in [1, \infty)$.

Remark The previous thm indicates that any continuous, non polynomial activation function σ is L^p-universal. We should relax the continuity assumption.

Recall: Given $\sigma: \mathbb{R} \to \mathbb{R}$, Borel measurable, denote

$$D(\sigma) = \{x \in \mathbb{R} : \sigma \text{ is discontinuous at } x\}.$$

If $m(\overline{D(\sigma)}) = 0$ then for any finite $[a,b]$ and any $\delta > 0$, there is a union of finitely many open intervals, $U \subseteq [a,b]$, such that $m(U) < \delta$ and σ is continuous on $[a,b] \setminus U$.

Here is the main thm on the L^p-universal approximation.

Theorem (Leshno et al. 1993) Let $\sigma: \mathbb{R} \to \mathbb{R}$ be Borel measurable and bounded on bounded subsets of $\mathbb{R}$. Assume $m(\overline{D(\sigma)}) = 0$. Then, σ is L^p-universal $\iff$ σ is not a polynomial (m-a.e. $\mathbb{R}$).

Remark Very large class of σ — practically, all σ, are L^p-universal.

Proof of Thm. Same reason as before that a polynomial σ will not work. So, only prove: σ is not a polynomial (a.e.) $\Longrightarrow$ σ is L^p-universal.

Fix $d \in \mathbb{N}$, $1 \le p < \infty$, $\mu \in \mathcal{M}_c(\mathbb{R}^d)$, and $f \in L^p(\mu)$. Denote $K = \text{supp}(\mu)$. Let $\varepsilon > 0$. There exists $\varphi \in C_c(\mathbb{R}^d)$ s.t.

$$\|\varphi - f\|_{L^p(\mu)} < \varepsilon/2.$$

By Theorem 2, with φ replacing that f in Theorem 2, there exists $\Phi \in \mathcal{N}_{\sigma,d}$ s.t.

$$\|\Phi - \varphi\|_{L^\infty(K)} < \varepsilon/2 \; \alpha, \qquad \alpha = \mu(K)^{-1/p}.$$

Thus,

$$\|\Phi - \varphi\|_{L^p(\mu)} = \Big(\int_K |\Phi - \varphi|^p d\mu\Big)^{1/p} < \varepsilon/2 \, \alpha \cdot \mu(K)^{1/p} = \varepsilon/2.$$

Hence, $\|\Phi - f\|_{L^p(\mu)} < \varepsilon$. <u>QED</u>