Lecture 5, Wednesday, April 17, 2024

Today: ⊙ Review: A general framework for proving the universal approximation w.r.t. to uniform norm

⊙ Two cases: $-\ \sigma = \text{ReLU}$

$-\ \sigma\ (\neq \text{const})$ is bounded

⊙ Most general case: σ is universal $\Longleftrightarrow$ σ is not a polynomial.

Next time:

universal approximations w.r.t. L^p and $W^{k,p}$ norms.

<u>Review and discussions</u>

⊙ Given $\sigma \in C(\mathbb{R})$, an activation function. To prove $\mathcal{NN}(\sigma, 2, d, 1)$ is dense in $C(K)$ for any compact $K \subseteq \mathbb{R}^d$ and any $d \in \mathbb{N}$, it suffices to prove that $\forall \mu \in \mathcal{M}(\mathbb{R}^d)$, compactly supported,

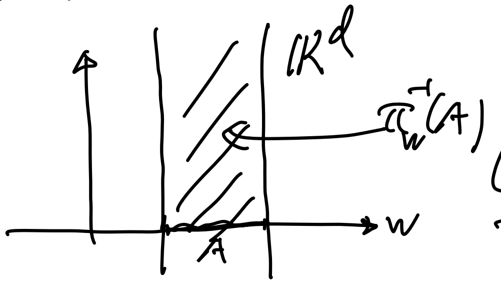
$$\int_{\mathbb{R}^d} \sigma(w\cdot x + \theta)\, d\mu(x) = 0 \quad \forall w \in \mathbb{R}^d\ \forall \theta \in \mathbb{R} \Longrightarrow \mu = 0. \qquad (*)$$

Note: the dimension d and the compact subset K are rather arbitrary.

⊙ An observation. Reduction from d-dimensional to one-dimensional. If $\mu \in \mathcal{M}_c(\mathbb{R}^d)$ and $w \in \mathbb{R}^d$ then

$$\int_{\mathbb{R}^d} g(w\cdot x)\, d\mu(x) = \int_{\mathbb{R}} g(t)\, d\mu_w(t)$$

for any one-variable Borel measurable $g: \mathbb{R} \to \mathbb{R}$. Here, $\mu_w = \mu \circ \pi_w^{-1} = \pi_{w\#}\mu$. $\pi_w: \mathbb{R}^d \to \mathbb{R}$: $\pi_w(x) = w\cdot x \quad \forall x \in \mathbb{R}^d$.



If $w = 0$ then $\pi_0(x) = 0\ \forall x \in \mathbb{R}^d$. So, if $A \subseteq \mathbb{R}$, Borel, $0 \notin A$, then $\pi_0^{-1}(A) = \phi$ (the empty set). Hence $\mu_0(A) = \mu(\pi_0^{-1}(A)) = 0$. Thus $\operatorname{supp}(\mu_0) \subseteq \{0\} \subseteq \mathbb{R}^d$.

If $\mu \in \mathcal{M}_c(\mathbb{R}^d)$ satisfies

$$\int_{\mathbb{R}^d} \sigma(w\cdot x + \theta)\, d\mu(x) = 0 \quad \forall w \in \mathbb{R}^d, \forall \theta \in \mathbb{R}.$$

then

$$\int_{\mathbb{R}^d} \sigma(\lambda w\cdot x + \theta)\, d\mu(x) = 0 \quad \forall w \in \mathbb{R}^d, \forall \theta, \lambda \in \mathbb{R}.$$

Hence

$$\boxed{\int_{\mathbb{R}} \sigma(\lambda t+\theta)\, d\mu_w(t)=0 \quad \forall \lambda, \theta \in \mathbb{R}. \ \forall w \in \mathbb{R}^d}$$

It's one-dimensional!

Note: for $w=0$, choose $\lambda=0$. we get $\sigma(\theta)\mu_0(\{0\})=0$. since μ_0 is concentrated at $t=0$. This is true for all θ. Thus, if $\sigma \not\equiv$ const. then $\mu_0=0$.

We only consider nonconstant σ's.

Definition Let $\sigma \in C(\mathbb{R})$ be nonconst.

(1) σ is universal w.r.t. the uniform norm if $NN(\sigma, 2, d, 1)$ is dense in $C(K)$ for any compact $K \subseteq \mathbb{R}^d$ and for any $d \in \mathbb{N}$.

(2) σ is discriminatory w.r.t. the uniform norm if for any compactly supported $\nu \in \mathcal{M}(\mathbb{R})$,

$$\int_{\mathbb{R}} \sigma(\lambda t+\theta)\, d\nu(t)=0 \quad \forall \lambda, \theta \in \mathbb{R} \implies \nu=0. \tag{**}$$

Theorem If $\sigma \in C(\mathbb{R})$ is nonconstant then it is universal w.r.t. the uniform norm $\iff$ it is discriminatory.

Proof "$\Rightarrow$" Let $\nu \in \mathcal{M}(\mathbb{R})$ be compactly supported. Assume that

$$\int_{\mathbb{R}} \sigma(\lambda t+\theta)\, d\nu(t)=0 \quad \forall \lambda, \theta \in \mathbb{R}. \tag{***}$$

We show that $\nu=0$.

Assume $\nu \neq 0$. Let $K=\mathrm{supp}(\nu) \subseteq \mathbb{R}^1$. $K(\neq \phi)$ is compact. Define $F \in [C(K)]^*$ by

$$F(g)=\int_K g\, d\nu \quad \forall g \in C(K). \tag{****}$$

Then, by Riesz's representation Thm, $F \leftrightarrow \mu$. In particular $f \neq 0$. Thus, $\exists f \in C(K)$ such that $F(f) \neq 0$. On the other hand, any function in $NN(\sigma, 2, d=1, 1)$ is of the form

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

The constant β is "absorbed" into $\sigma(\lambda x+\theta)$ for $\lambda=0$ and $\sigma(\theta) \neq 0$ such θ exists! By (***) and (****), we have $F(\Phi)=0$. Thus,

$F=0$ on the closure of $NN(\sigma,2,1,1)$. Thus, f is not in this closure. So, σ is not universal.

"$\Longleftarrow$" Suppose σ is discriminatory. Let $d\in\mathbb{N}$ and $K\subseteq\mathbb{R}^d$ compact. We prove that $NN(\sigma,2,d,1)$ is dense in $C(K)$. It suffices to show that if $\mu\in\mathcal{M}(\mathbb{R}^d)$, compactly supported, and

$$\int_{\mathbb{R}^d}\sigma(w\cdot x+\theta)\,d\mu(x)=0 \quad \forall w\in\mathbb{R}^d \ \forall\theta\in\mathbb{R}, \qquad (*****)$$

then $\mu=0$. (In fact, suppose this is proved. If $\exists f\in C(K)$ that is not in the closure of $NN(\sigma,2,d,1)$ (in $C(K)$), then by the Hahn-Banach thm and Riesz's representation thm, $\exists\tilde{\mu}\in\mathcal{M}(K)$, $\tilde{\mu}\neq0$, satisfying $(*****)$ with μ replaced by $\tilde{\mu}$. If we extend $\tilde{\mu}$ to $\mathbb{R}^d$ trivially, $\tilde{\mu}\in\mathcal{M}(\mathbb{R}^d)$, compactly supported. Then we would have $\tilde{\mu}=0$, a contradiction.)

As remarked above, $(*****)$ implies that

$$\int_{\mathbb{R}}\sigma(\lambda t+\theta)\,d\mu_w(t)=0 \quad \forall w\in\mathbb{R}^d \ \forall\lambda,\theta\in\mathbb{R}.$$

where $\mu_w=\mu\circ\pi_w^{-1}$, $\pi_w:\mathbb{R}^d\to\mathbb{R}$, $\pi_d(x)=w\cdot x$ $\forall x\in\mathbb{R}^d$. Clearly, each $\mu_w\in\mathcal{M}(\mathbb{R})$ is compactly supported. By the assumption, σ is discriminatory. $\mu_w=0$ $\forall w\in\mathbb{R}^d$. Thus, the Fourier transform of μ at w is

$$\hat{\mu}(w)=\int_{\mathbb{R}^d}e^{-2\pi i\,w\cdot x}\,d\mu(x)=\int_{\mathbb{R}}e^{-2\pi i t}\,d\mu_w(t)=0.$$

Hence $\hat{\mu}\equiv0$ and $\mu=0$. <u>QED</u>

⊙ Let's try to understand better the meaning of "σ is discriminatory". Let $\nu\in\mathcal{M}(\mathbb{R})$, compactly supported. Assume ν has a compactly supported density $u:\mathbb{R}\to\mathbb{R}$, say, u is smooth. Suppose σ is discriminatory. Then

$$\int_{\mathbb{R}}\sigma(\lambda t+\theta)\,u(t)\,dt=0 \quad \forall\lambda,\theta\in\mathbb{R}$$

whould imply that $u\equiv0$.

Compare with: $\displaystyle\int_{\mathbb{R}}p(x)\,u(t)\,dt=0 \ \forall p:\text{polynomial} \Longrightarrow u\equiv0.$

Why the class of functions $\sum_j \alpha_j \sigma(\lambda_j t + \theta_j)$ $(\alpha_j, \lambda_j, \theta_j \in \mathbb{R})$ is so "rich / abundant" that they can annihilate any ν?

Example $\sigma(t) = \sin t$. $\sigma(kt) = \sin kt$ $(\forall k \in \mathbb{Z})$. $\sigma(\frac{\pi}{2} - kt) = \cos kt$.

Theorem (Afzman 1983) If $\sigma \in C_0(\mathbb{R})$, nonnegative, $\sigma \not\equiv 0$. Then linear combinations of $\sigma(\lambda x + \theta)$, $\lambda \in \mathbb{R}$, $\theta \in \mathbb{R}$, are dense in $C_0(\mathbb{R})$.

QED

Very similar proof leads to:

Hornik (1991) If $\sigma \in C(\mathbb{R})$ is bounded and non constant, then it is discriminatory.

In fact, Hornik (1991) proved more: If $\sigma: \mathbb{R} \to \mathbb{R}$ is measurable, bounded, and non constant, and if $\nu \in \mathcal{M}(\mathbb{R})$ (not necessary compactly supported) satisfies

$$\int_{\mathbb{R}} \sigma(\lambda t + \theta)\, d\nu(t) = 0 \quad \forall \lambda, \theta \in \mathbb{R},$$

then $\nu = 0$.

(skip it in class)

Sketch of proof Let $g(t) = e^{-t^2}$. (Note: $\hat{g}(t) \neq 0$ $\forall t \in \mathbb{R}$.) By Fubini's Thm, $\forall \lambda, \theta \in \mathbb{R}$:

$$0 = \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \sigma(\lambda t - \theta + \lambda s)\, d\nu(t) \right] g(s)\, ds$$

$$= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \sigma(\lambda(t+s) - \theta)\, g(s)\, ds \right] d\nu(t)$$

$$\overset{t+s=s'}{=\!=} \int_{\mathbb{R}} \int_{\mathbb{R}} \sigma(\lambda s' - \theta)\, g(s' - t)\, ds'\, d\nu(t)$$

$$\overset{s' \to s}{=\!=} \int_{\mathbb{R}} \sigma(\lambda s - \theta) \overbrace{\left[\int_{\mathbb{R}} g(s-t)\, d\nu(t) \right]}^{= (g * \nu)(s)} ds$$

$$= \int_{\mathbb{R}} \sigma(\lambda s - \theta)\, h(s)\, ds,$$

where $h(s) = (g * \nu)(s)$ $(s \in \mathbb{R})$, $h \in L^p(\mathbb{R})$ $(1 \le p < \infty)$, and h is absolutely continuous.

We want to show that $h \equiv 0$. This implies

$$\hat{h} = \widehat{g * \nu} = \hat{g}\,\hat{\nu} = 0$$

But $\hat{g}(\xi) \neq 0$ $(\forall \xi \in \mathbb{R})$. Hence $\hat{\nu}(\xi) = 0$ $(\forall \xi \in \mathbb{R})$, and $\nu = 0$.

Now, $\forall \alpha \in \mathbb{R}, \alpha \neq 0, \forall \gamma \in \mathbb{R}$: $\int_{\mathbb{R}} \sigma(\frac{s}{\alpha} - \frac{\gamma}{\alpha}) h(s) ds = 0$.

Change variable: $\frac{s}{\alpha} - \frac{\gamma}{\alpha} = t$.

$$\boxed{\int_{\mathbb{R}} \sigma(t)\, h(\alpha t + \gamma)\, dt = 0. \quad \forall \alpha \neq 0, \forall \gamma.}$$

[Denote $h_\alpha(t) = h(\alpha t)$ ($\forall t \in \mathbb{R}$). Then, $\forall u \in L^1(\mathbb{R})$,

$$\int_{\mathbb{R}} \sigma(t) (u * h_\alpha)(t)\, dt = 0.$$

Thus, $\forall u \in L^1(\mathbb{R})$, $\int_{\mathbb{R}} \sigma(t)\, (u * h_\alpha)(t)\, dt = 0$.]

Denote by A the subspace of $L^1(\mathbb{R})$ spanned by functions $h_\alpha = h_\alpha(t)$, $\alpha \neq 0$. This is translation-invariant. Its $L^1(\mathbb{R})$-closure $\overline{A}$ is a translation-invariant subspace of $L^1(\mathbb{R})$. Then $\overline{A}$ is an ideal w.r.t. to the convolution. (see Ch7. of Rudin's Fourier Analysis on Groups.)

Note that $\hat{\gamma}(0) = 0$ since $\sigma \not\equiv$ const. Hence $\hat{h}(0) = 0$. If $\exists \xi_0 \neq 0$ s.t. $\hat{h}(\xi_0) = 0$ then $\hat{h}_\alpha(\xi) = \frac{1}{\alpha} \hat{h}(\xi/\alpha)$ ($\forall \alpha \neq 0$) implies that $\hat{h} \equiv 0$ and hence $h \equiv 0$, we are done. Now, assume $\hat{h}(\xi) = 0 \Leftrightarrow \xi = 0$. This means that $\hat{h} \not\equiv 0$. Moreover, (note: $u_n \to u$ in $L^1(\mathbb{R}^d) \Rightarrow \hat{u}_n \to \hat{u}$ pointwise.)

$$Z(\overline{A}) = \bigcap_{u \in \overline{A}} Z(u) = \bigcap_{u \in A} Z(u) = \{0\},$$

where for $u \in L^1(\mathbb{R})$, $Z(u) = \{\xi \in \mathbb{R} : \hat{u}(\xi) = 0\}$. Now, by Theorem 7.2.4 of Rudin's Fourier Analysis on Groups (1962 edition), the following is true: if $u \in L^1(\mathbb{R})$ and $\hat{u}(0) = 0$ then $u \in \overline{A}$.

Now, by the above boxed formula,

$$\int_{\mathbb{R}} \sigma(t)\, u(t)\, dt = 0 \quad \forall u \in \overline{A}.$$

In particular, $\forall u \in L^1(\mathbb{R})$ with $\hat{u}(0) = \int_{\mathbb{R}} u(t)\, dt = 0$,

$$\int_{\mathbb{R}} \sigma(t)\, u(t)\, dt = 0.$$

Let $-\infty < a < b < \infty$. Let $u \in C_c(\mathbb{R})$ with $\mathrm{supp}(u) \subseteq [a,b]$. Denote $\overline{\sigma}_{a,b} = \frac{1}{b-a} \int_a^b \sigma(t)\, dt$ and $\overline{u}_{a,b} = \frac{1}{b-a} \int_a^b u(t)\, dt$. Then $U = \chi_{[a,b]}(u - \overline{u}_{a,b}) \in L^1(\mathbb{R})$ with $\int_{\mathbb{R}} U(t)\, dt = 0$. Hence, we have

$\int_{\mathbb{R}} \sigma(t) U(t) dt = 0$. i.e., $\int_a^b \sigma(t)(u(t) - \bar{u}_{a,b}) dt = 0$. Thus,

$$\int_a^b (\sigma(t) - \bar{\sigma}_{a,b}) u(t) dt = \int_a^b (\sigma(t) - \bar{\sigma}_{a,b})(u(t) - \bar{u}_{a,b}) dt$$
$$= \int_a^b \sigma(t)(u(t) - \bar{u}_{a,b}) dt - \int_a^b \bar{\sigma}_{a,b}(u(t) - \bar{u}_{a,b}) dt = 0.$$

Thus, $\sigma(t) \equiv \bar{\sigma}_{a,b}$ $\forall t \in [a,b]$. Since a,b are arbitrary, $\sigma(t) =$ a const. on $\mathbb{R}$, a contradiction. Thus, $\hat{h} \equiv 0$ and $h \equiv 0$. QED

About the ReLU: ReLU is universal.

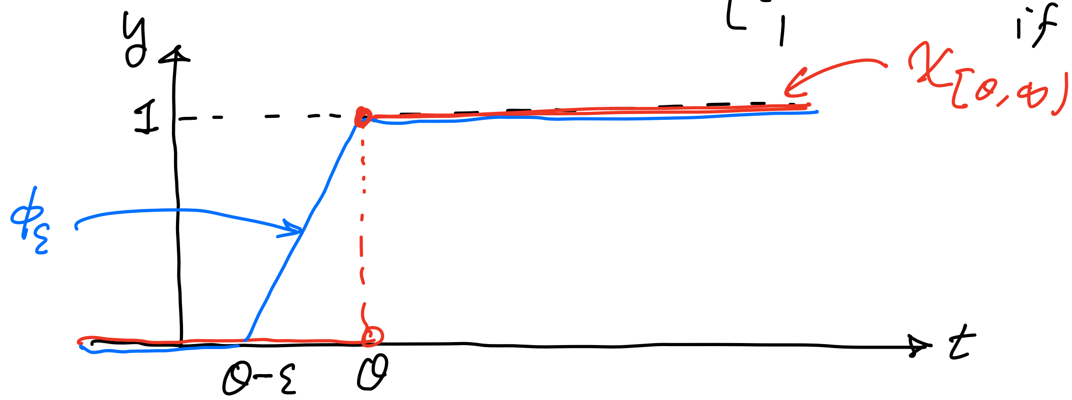
Proof Let $\nu \in M(\mathbb{R})$ with compact support. Suppose

$$\int_{\mathbb{R}} \sigma(\lambda t + \theta) d\nu(t) = 0 \quad \forall \lambda, \theta \in \mathbb{R}. \qquad (*)$$

We prove that $\nu = 0$. It suffices to show that for any $\theta \in \mathbb{R}$, $\int_{\mathbb{R}} \chi_{[\theta,\infty)}(t) d\nu(t) = 0$.

Define for $\varepsilon > 0$

$$\phi_\varepsilon(t) = \frac{1}{\varepsilon}\sigma(t - (\theta - \varepsilon)) - \frac{1}{\varepsilon}\sigma(t - \theta) = \begin{cases} 0 & \text{if } t < \theta - \varepsilon, \\ \frac{1}{\varepsilon}(t - \theta + \varepsilon) & \text{if } \theta - \varepsilon \le t < \theta, \\ 1 & \text{if } t \ge \theta. \end{cases}$$



Then, $\forall t \in \mathbb{R}$: $\phi_\varepsilon(t) \longrightarrow \chi_{[\theta,\infty)}(t)$ as $\varepsilon \to 0^+$ (check the cases: $t > \theta$, $t = \theta$, $t < \theta$. For $t \ge \theta$, $\phi_\varepsilon(t) = \chi_{[\theta,\infty)}(t)$ $\forall \varepsilon > 0$.).

By the Lebesgue Dominated Convergence Theorem, applied to $|\nu|$, and the assumption $(*)$,

$$0 = \int_{\mathbb{R}} \phi_\varepsilon(t) d\nu(t) \longrightarrow \int_{\mathbb{R}} \chi_{[\theta,\infty)}(t) d\nu(t).$$

Thus for any $\theta_1 < \theta_2$, $\int_{\mathbb{R}} \chi_{[\theta_1,\theta_2]}(t) d\nu(t)$. By approximations,

$$\int_{\mathbb{R}} \cos \xi t \, d\nu(t) = 0, \qquad \int_{\mathbb{R}} \sin \xi t \, d\nu(t) = 0 \qquad \forall \xi \in \mathbb{R}.$$

Hence $\hat{\nu} \equiv 0$ and $\nu = 0$. QED

We now present the most general result.

<u>Theorem 1 (Leshno et al. 1993)</u> Let $\sigma \in C(\mathbb{R})$, $\sigma \not\equiv$ const. Then σ is universal $\iff$ σ is not a polynomial.

As in Leshno et al (1993), we prove an improved version, with less assumptions on σ. Let $\sigma : \mathbb{R} \to \mathbb{R}$ be Lebesgue measurable and bounded on bounded sets (i.e., $\sigma \in L^\infty_{loc}(\mathbb{R})$). Denote

$$D(\sigma) = \{x \in \mathbb{R} : \sigma \text{ is discontinuous at } x\}.$$
$$m(A) = \text{the Lebesgue measure of } A.$$

<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 $\mathbb{R}$.

<u>Remark</u> ⊙ If $\sigma \notin C(\mathbb{R})$, $\Phi \in \mathcal{N}_{\sigma,d}$ may not be continuous. But, $\Phi \in L^\infty(K)$. We can approximate a continuous function w.r.t. L^∞-norm. But an L^∞-function cannot in general be approximated by continuous functions.

⊙ Clearly Thm 2 implies Thm 1, as $\mathcal{N}_{\sigma,d} \subseteq C(K)$ if $\sigma \in C(\mathbb{R})$.

We prove Thm 2 by first proving some 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 <u>not</u> dense in $L^p(K)$ (w.r.t. the $L^p(K)$-norm). Moreover, $C(K) \not\subseteq \overline{\mathcal{N}_{\sigma,d}}$ w.r.t. to the $L^\infty(K)$-norm.

<u>Proof</u> By the assumption, $\exists A_0 \subseteq \mathbb{R}$ s.t. $m(A_0) = 0$ and there exists a polynomial $\tilde{\sigma} : \mathbb{R} \to \mathbb{R}$ s.t. $\sigma(s) = \tilde{\sigma}(s)$ $\forall s \in \mathbb{R} \setminus A_0$. Note that, if $w \in \mathbb{R}^d$ and $\theta \in \mathbb{R}$, then the set $\{x \in \mathbb{R}^d : w \cdot x + \theta \in A_0\}$ has the d-dimensional Lebesgue measure 0, as $\{x \in \mathbb{R}^d : w \cdot x \in A_0\} = \pi_w^{-1}(A_0)$ has the zero d-dim. Lebesgue measure. Thus $x \mapsto \sigma(w \cdot x + \theta)$ and $x \mapsto \tilde{\sigma}(w \cdot x + \theta)$ are the same in $L^p(K)$ for $1 \le p \le \infty$. Thus, $\mathcal{N}_{\sigma,d} = \mathcal{N}_{\tilde{\sigma},d}$ is a subspace

of $L^p(K)$ $(1 \le p \le \infty)$.

Suppose $\deg(\tilde{\sigma}) = n \ge 0$. Then $N_{\tilde{\sigma},d}$ is a subclass of polynomials of $x \in \mathbb{R}^d$ with total deg. $\le n$. Hence, the closure of $N_{\tilde{\sigma},d} = N_{\sigma,d}$ in any $L^p(K)$ $(1 \le p \le \infty)$ is a finitely dimensional space, and hence $N_{\sigma,d}$ is not dense in $L^p(K)$, and $C(K) \not\subseteq \overline{N_{\sigma,d}}$, as all $L^p(K)$ and $C(K)$ are infinite-dimensional spaces. QED

With this lemma, we need only to consider a measurable $\sigma : \mathbb{R} \to \mathbb{R}$, bounded on bounded sets, that is not a polynomial m-a.e. $\mathbb{R}$.

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

Proof First note that $N_{\sigma,d} \subseteq L^\infty(A)$ for any $A \subseteq \mathbb{R}^d$ that is bounded and Lebesgue measurable. Fix $d \in \mathbb{N}$ and $K \subseteq \mathbb{R}^d$: compact.

Denote by $\mathcal{A}$ the class of functions on $\mathbb{R}^d$ that are linear combinations of $\cos(v \cdot x)$, $\sin(w \cdot x)$, $v, w \in \mathbb{R}^d$. By the Stone-Weierestrass Thm, $\mathcal{A}$ (restricted to $K \subseteq \mathbb{R}^d$, compact) is dense in $C(K)$ w.r.t. to the $C(K)$-norm

Let $f \in C(K)$, $\varepsilon > 0$. Then, there exists $g_i \in C(\mathbb{R})$ $w_i \in \mathbb{R}^d$ $(i = 1, \cdots, n)$ s.t.

$$\max_{x \in K} \left| f(x) - \sum_{i=1}^{n} g_i(w_i \cdot x) \right| < \varepsilon/2.$$

Note that all $w_i \cdot x \in [a,b]$ $(i = 1, \cdots, n,\ x \in K)$ for some $-\infty < a < b < \infty$. But, by the assumption, $C([a,b]) \subseteq \overline{N_\sigma}$ w.r.t. $L^\infty([a,b])$-norm. Thus, $\exists \phi_i \in N_\sigma$ $(1 \le i \le n)$, s.t.

$$\| \phi_i - g_i \|_{L^\infty([a,b])} < \varepsilon/2n, \quad i = 1, \cdots, n.$$

In particular, $\exists A_0 \subseteq \mathbb{R}$, $m(A_0) = 0$, s.t.

$$|\phi_i(t) - g_i(t)| < \varepsilon/n \quad \forall t \in K \setminus A_0, \quad i = 1, 2, \cdots, n.$$

Since $\varphi_i \in N_\sigma$,

$$\varphi_i(t) = \sum_{j=1}^{N_i} \alpha_j^{(i)} \sigma(\beta_j^{(i)} t + \theta_j^{(i)}), \text{ all } \alpha_j^{(i)}, \beta_j^{(i)}, \theta_j^{(i)} \in \mathbb{R}.$$

Define

$$\Phi_i(x) = \sum_j \alpha_j^{(i)} \sigma(\beta_j^{(i)} w_i \cdot x + \theta_j^{(i)}), \quad x \in \mathbb{R}^d, \quad i = 1, \cdots, n.$$

Then, each $\Phi_i \in \mathcal{N}_{\sigma,d}$. If $x \in K$, $w_i \cdot x \notin A_0$, then with $t = w_i \cdot x$,

$$|\Phi_i(x) - g_i(w_i \cdot x)| = |\varphi_i(t) - g_i(t)| < \varepsilon/2n,$$

Since the set $\{x \in K : w_i \cdot x \in A_0\}$ has the d-dimensional Lebesgue measure 0, we have

$$\|\Phi_i - g_i(w_i \cdot \bullet)\|_{L^\infty(K)} < \varepsilon/2n. \quad i = 1, \dots, n.$$

Hence, for $\Phi := \sum_i \Phi_i \in \mathcal{N}_{\sigma,d}$, we have

$$\|\Phi - f\|_{L^\infty(K)} \le \|\Phi - \sum_i g_i(w_i \cdot \cdot)\|_{L^\infty(K)} + \|f - \sum_i g_i(w_i \cdot \cdot)\|_{L^\infty(K)} < \varepsilon.$$

Thus, $\overline{\mathcal{N}_{\sigma,d}} \supseteq C(K)$ w.r.t. the $L^\infty(K)$-norm. <u>QED</u>

<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>Proof</u> Let $K \subseteq [a,b]$ for some $a, b \in \mathbb{R}$ with $-\infty < a < b < \infty$. $\forall w, \theta \in \mathbb{R}$. Let $h \in (0,1)$. Define $\sigma_{w,\theta}(x) = \sigma(w \cdot x + \theta)$ $(x \in \mathbb{R})$. Then $\sigma_{w,\theta} \in \mathcal{N}_\sigma$.
Define

$$\Delta_h \sigma_{w,\theta}(x) = \frac{1}{h}\left[\sigma_{w+h,\theta}(x) - \sigma_{w,\theta}(x)\right]$$
$$= \frac{1}{h}\left[\sigma((w+h)x + \theta) - \sigma(wx+\theta)\right] \quad (x \in \mathbb{R})$$

Each $\Delta_h \sigma_{w,\theta} \in \mathcal{N}_\sigma$. Let $x \in [a,b]$. Then.

$$\Delta_h \sigma_{w,\theta}(x) - \frac{d}{d\alpha}\left[\sigma(\alpha \cdot x + \theta)\right]\Big|_{\alpha = w}$$
$$= \frac{1}{h}\left[\sigma((w+h)x+\theta) - \sigma(wx+\theta) - h\frac{d}{d\alpha}\sigma(\alpha x + \theta)\Big|_{\alpha=w}\right]$$
$$= \frac{1}{h}\,\frac{1}{2}\,\frac{d^2}{d\alpha^2}\left[\sigma(\alpha x + \theta)\right]\Big|_{\alpha = \xi} \quad \text{for some } \xi = \xi(x,\theta) \in [w, w+h]$$

Thus, since $\{\alpha x + \theta \mid \alpha \in [w, w+1], x \in [a,b]\}$ is bounded in $\mathbb{R}$,

$$\max_{a \le x \le b} \left|\Delta_h \sigma_{w,\theta}(x) - \frac{d}{d\alpha}\left[\sigma(\alpha \cdot x + \theta)\right]\Big|_{\alpha=w}\right| \to 0 \quad \text{as } h \to 0.$$

Hence $\frac{d}{dw}\sigma(w \cdot x + \theta)$ $\left(= \frac{d}{d\alpha}[\sigma(\alpha x + \theta)]\big|_{\alpha = w}\right.$, a $C([a,b])$-function of x $\big)$ is in $\overline{\mathcal{N}_\sigma}$. By induction $\frac{d^k}{dw^k}\sigma(wx+\theta)$ $\left(= \frac{d^k}{d\alpha^k}[\sigma(\alpha x + \theta)]\big|_{\alpha=w}\right.$, a $C([a,b])$-function of x $\big)$ is in $\overline{\mathcal{N}_\sigma}$ for any integer $k \ge 1$. $\left(\overline{\mathcal{N}_\sigma} = \overline{\mathcal{N}_\sigma}^{C([a,b])}\right)$

For each $k \ge 1$, $\exists\, \theta_k \in \mathbb{R}$ s.t. $\sigma^{(k)}(\theta_k) \ne 0$, since σ is not a polynomial.
Since

$$\frac{d^k}{d\alpha^k}\left[\sigma(\alpha x + \theta)\right]\Big|_{\alpha = w} = x^k \sigma^{(k)}(wx+\theta).$$

and with $w = 0$ and $\theta = \theta_k$, the function of x,

$$\frac{d^k}{d\alpha^k}\left[\sigma(\alpha x+\theta_k)\right]\Big|_{\alpha=0} = x^k\sigma^{(k)}(\theta_k),$$

as a function in $C([a,b])$, is in $\overline{N_\sigma}$. Thus, $\overline{N_\sigma}$ contains all polynomials of deg. ≥ 1. But constant functions are in N_σ. Thus, $\overline{N_\sigma}$ contains all polynomials, and hence $\overline{N_\sigma}=C([a,b])$ by the Weierstrass Thm.

QED

Lemma 4 Let $\sigma:\mathbb{R}\to\mathbb{R}$ be Lebesgue measurable and bounded on bounded sets. Assume $m(D(\sigma))=0$. Let $\varphi\in C_c^\infty(\mathbb{R})$. Then for any $K\subseteq\mathbb{R}^1$, compact, $\sigma*\varphi\in\overline{N_\sigma}$ w.r.t. the $L^\infty(K)$-norm.

Proof Note that $\sigma*\varphi\in C^\infty(\mathbb{R})$. Let $\varepsilon>0$. Assume $\mathrm{supp}(\varphi)\subseteq[-\alpha,\alpha]$ for some $\alpha>0$. Let $\delta>0$ be s.t.

$$4\delta\|\varphi\|_{L^\infty}\|\sigma\|_{L^\infty([-2\alpha,2\alpha])}\leq\varepsilon. \qquad (*)$$

For this $\delta>0$, there are $r=r(\delta)$ open intervals in $[-2\alpha,2\alpha]$ their union is U, s.t. $m(U)<\delta$ and σ is uniformly continuous on $[-2\alpha,2\alpha]\setminus U$.

Now, let $N\in\mathbb{N}$ satisfy

(1) $\delta N\geq\alpha r(\delta)$;

(2) $|\sigma(s)-\sigma(t)|\leq\dfrac{\varepsilon}{\|\varphi\|_{L^1(\mathbb{R})}}$ if $s,t\in[-2\alpha,2\alpha]\setminus U$, $|s-t|\leq\dfrac{\alpha}{N}$.

Let $h=\alpha/N$ and $y_k=-\alpha+kh$ $(k=0,1,\dots,2N)$, and $I_k=[y_{k-1},y_k]$.

Since $\sigma\in L^\infty_{loc}(\mathbb{R})$, $\exists B_0\subseteq[-\alpha,\alpha]$ with $m(B_0)=0$ s.t.

$$|\sigma(x)|\leq\|\sigma\|_{L^\infty([-2\alpha,2\alpha])}\quad\forall x\in[-\alpha,\alpha]\setminus B_0. \qquad (**)$$

Let $x\in[-\alpha,\alpha]\setminus B_0$. Then $x-y\in[-2\alpha,2\alpha]$ if $y\in[-\alpha,\alpha]$. We have

$$\left|\int_{\mathbb{R}}\sigma(x-y)\varphi(y)\,dy-\sum_{k=1}^{2N}\sigma(x-y_k)\int_{I_k}\varphi(y)\,dy\right|$$

$$=\left|\sum_{k=1}^{2N}\int_{I_k}\left[\sigma(x-y)-\sigma(x-y_k)\right]\varphi(y)\,dy\right| \qquad \left(\begin{array}{l}\text{since } \mathrm{supp}(\varphi)\subseteq[-\alpha,\alpha]\\ \text{and } \bigcup_k I_k=[-\alpha,\alpha].\end{array}\right)$$

$$\leq\sum_{k=1}^{2N}\int_{I_k}|\sigma(x-y)-\sigma(x-y_k)||\varphi(y)|\,dy. \qquad (\star)$$

If $(x-I_k)\cap U=\phi$ then by (3),

$$\int_{I_k}|\sigma(x-y)-\sigma(x-y_k)||\varphi(y)|\,dy\leq\frac{\varepsilon}{\|\varphi\|_{L^1(\mathbb{R})}}\int_{I_k}|\varphi(y)|\,dy.$$

Thus,

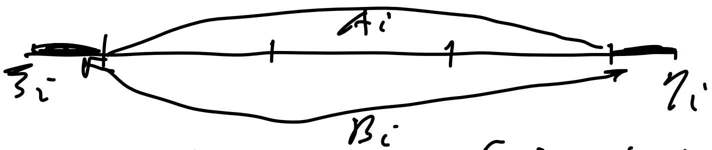
$$\sum_{k:\,(x-I_k)\cap U=\phi}\int_{I_k}|\sigma(x-y)-\sigma(x-y_k)||\varphi(y)|\,dy\leq\varepsilon. \qquad (\star\star)$$

Consider now $(x-I_k)\cap U\neq\phi$. We have

$$\sum_{k:\,(x-I_k)\cap U\neq\phi}\int_{I_k}|\sigma(x-y)-\sigma(x-y_k)|\,|\varphi(y)|\,dy$$

$$\le 2\|\sigma\|_{L^\infty([-2\alpha,2\alpha])}\,\|\varphi\|_{L^\infty}\cdot\Big(\sum_{k:\,(x-I_k)\cap U\neq\phi}1\Big)h. \qquad (***)$$

Let $U=\bigcup_i(\xi_i,\eta_i)\subseteq[-2\alpha,2\alpha]$, the union is a disjoint, finite union. For each i, $(\xi_i,\eta_i)=A_i\cup B_i$, disjoint, where A_i is the union of some I_k's such that $(x-I_k)\cap U\subseteq(\xi_i,\eta_i)$, and $B_i=(\xi_i,\eta_i)\setminus A_i$ has measure $m(B_i)\le 2h$. Thus, since $m(U)<\delta$ and U is the union of $r(\delta)$ open



intervals, we have $m(\bigcup_i A_i)\le m(U)\le\delta$ (since $\bigcup_i A_i\subseteq U$), and $m(\bigcup_i B_i)\le 2h\,r(\delta)$. Consequently, by (1) (note: $h=\frac{\alpha}{N}$), and $(*)$,

$$\Big(\sum_{k:\,(x-I_k)\cap U\neq\phi}1\Big)h\le\delta+2h\,r(\delta)\le 2\delta.$$

This and $(***)$ imply that

$$\sum_{k:\,(x-I_k)\cap U\neq\phi}\int_{I_k}|\sigma(x-y)-\sigma(x-y_k)|\,|\varphi(y)|\,dy\le\varepsilon. \qquad (\star\star\star)$$

Now, $(\star)-(\star\star\star)$ imply

$$\Big|\int_{\mathbb{R}}\sigma(x-y)\varphi(y)\,dy-\sum_{k=1}^{2N}\sigma(x-y_k)\int_{I_k}\varphi(y)\,dy\Big|\le 2\varepsilon,\quad \text{a.e. } x\in[-\alpha,\alpha]$$

Since each $\sigma(x-y_k)$ as a function of x is in N_σ and all $\int_{I_k}\varphi(y)dy$ are constants, $\sum_k\sigma(x-y_k)\int_{I_k}(y)dy$ is a function in N_σ. Thus $\sigma*\varphi\in\overline{N_\sigma}$ with respect to the $L^\infty(K)$-norm. <u>QED</u>

<u>Remark</u> If $\sigma\in C(\mathbb{R})$ then the proof can be simplied. Here is the proof.

Fix a compact $K\subseteq\mathbb{R}$. It suffices to show that for any $\nu\in M(\mathbb{R})$, compactly supported,

$$\int_{\mathbb{R}}\sigma*\varphi\,d\nu=0 \quad\text{provided}\quad \int_{\mathbb{R}}\sigma(wx+\theta)\,d\nu(x)=0\quad\forall w,\theta\in\mathbb{R}.$$

Fix such a measure ν. We have by Fubini's Thm that

$$\int_{\mathbb{R}}(\sigma*\varphi)(x)\,d\nu(x)=\int_{\mathbb{R}}\Big[\int_{\mathbb{R}}\sigma(x-s)\varphi(s)\,ds\Big]d\nu(x)$$

$$=\int_{\mathbb{R}}\Big[\int_{\mathbb{R}}\sigma(x-s)\,d\nu(x)\Big]\varphi(s)\,ds$$

$$=0.$$

<u>QED</u>

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

Remark. Since σ may not be continuous, $N_\sigma \subseteq L^\infty(K)$ but N_σ may not be a subset of $C(K)$.

Proof Since $\sigma * \varphi \in C^\infty(\mathbb{R})$ is not a polynomial, by Lemma 3, $C(K) = \overline{N_{\sigma*\varphi}}$ w.r.t. $C(K)$-norm for any compact $K \subseteq \mathbb{R}$. Moreover, by Lemma 4, $\sigma * \varphi \in \overline{N_\sigma}$ w.r.t. $L^\infty(K)$-norm for any compact $K \subseteq \mathbb{R}$. Now, fix a compact $K \subseteq \mathbb{R}$. We show that $C(K) \subseteq \overline{N_\sigma}$ with resp. to $L^\infty(K)$-norm.

$\forall f \in C(K)$. $\forall \varepsilon > 0$. $\exists \Phi \in N_{\sigma*\varphi}$ s.t. $\|\Phi - f\|_{C(K)} \le \varepsilon$. Let $\Phi(x) = \sum_{j=1}^N \alpha_j (\sigma * \varphi)(w_j x + \theta_j) + \beta_0$ ($\alpha_j, w_j, \theta_j, \beta_0 \in \mathbb{R}$). For each j, $\sigma * \varphi \in \overline{N_\sigma}$ w.r.t. $L^\infty(K_{w_j,\theta_j})$-norm, where $K_{w_j,\theta_j} = w_j K + \theta_j$ (compact). Thus, $\exists \psi_j \in N_\sigma$ s.t. $\|\alpha_j \sigma * \varphi - \psi_j\|_{L^\infty(K_{w_j,\theta_j})} \le \varepsilon / N$. Since the function $\tilde{\psi}_j : x \longmapsto \psi_j(w_j x + \theta_j)$ is in N_σ, $\tilde{\Psi} := \sum_j \tilde{\psi}_j + \beta_0 \in N_\sigma$, and

$$\|\tilde{\Psi} - \Phi\|_{L^\infty(K)} \le \sum_j \|\tilde{\psi}_j - \alpha_j (\sigma * \varphi)\|_{L^\infty(K)} = \sum_j \|\psi_j - \alpha \sigma * \varphi\|_{L^\infty(K_{w_j,\theta_j})} \le \varepsilon.$$

Finally, $\|\tilde{\Psi} - f\|_{L^\infty(K)} \le 2\varepsilon$. QED.

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 We show that there exists $l \in \mathbb{N}$ s.t. degree $(\sigma * \varphi) \le l$ for all $\varphi \in C_c^\infty(\mathbb{R})$. If this is proved, then

$$\frac{d^{l+1}}{dx^{l+1}}(\sigma * \varphi)(0) = 0, \quad \text{i.e.,} \int_K \sigma(-s) \varphi^{(l+1)}(s)\, ds = 0 \quad \forall \varphi \in C_c^\infty(\mathbb{R}).$$

Thus, the distributional derivatives $\sigma^{(l+1)}(-s) = 0$ $\forall s \in \mathbb{R}$. Hence, σ is a polynomial of degree $\le l$, m-a.e. $\mathbb{R}$.

Let $-\infty < a < b < \infty$. Denote $C_c^\infty([a,b]) = \{\varphi \in C_c^\infty(\mathbb{R}) : \operatorname{supp} \varphi \subseteq [a,b]\}$.

Define
$$\rho(\varphi_1, \varphi_2) = \sum_{n=0}^\infty \frac{1}{2^n} \frac{\|\varphi_1 - \varphi_2\|_n}{1 + \|\varphi_1 - \varphi_2\|_n} \quad \forall \varphi_1, \varphi_2 \in C_c^\infty([a,b]),$$

where $\|\varphi\|_n = \sum_{j=0}^{n} \|\varphi^{(j)}\|_{C([a,b])}$. Then $(C_c^\infty([a,b]), \rho)$ is a complete metric space (a Fréchet space).

By assumption, $\sigma * \varphi$ is a polynomial if $\varphi \in C_c^\infty([a,b])$.

Define $V_k = \{\varphi \in C_c^\infty([a,b]) : \deg(\sigma * \varphi) \le k\}$. Then $V_0 \subseteq V_1 \subseteq \cdots$

All V_k's are closed subspaces of $C_c^\infty([a,b])$, and

$$C_c^\infty([a,b]) = \bigcup_{k=0}^{\infty} V_k .$$

(Closedness: $\deg(\sigma * \varphi_j) \le k$ $(j=1,\cdots)$, $\rho(\varphi_j, \varphi) \to 0$ $(j \to \infty)$ then

$$0 = \frac{d^{k+1}}{dx^{k+1}} (\sigma * \varphi_j)(x) = \int_{[a,b]} \sigma(s)\, \varphi_j^{(k+1)}(x-s)\, dx$$

$$\to \int_{[a,b]} \sigma(s)\, \varphi^{(k+1)}(x-s)\, ds = \frac{d^{k+1}}{dx^{k+1}} (\sigma * \varphi)(x) \qquad \forall x \in [a,b].)$$

By Baire's Category Thm, $\exists\, l \ge 0$ such that V_l contains an open ball of $C_c^\infty([a,b])$. But V_l is a subspace. Thus, $C_c^\infty([a,b]) = V_l$.

We show that for all $\varphi \in C_c^\infty(\mathbb{R})$, $\deg(\sigma * \varphi) \le l$. By translation, l depends on $b-a$ only. [If $\varphi \in C_c^\infty(\mathbb{R})$, $\operatorname{supp} \varphi \subseteq [\alpha, \beta]$, $\beta - \alpha \le b-a$. Define $\theta = a - \alpha$, $\varphi_\theta(x) = \varphi(x+\theta)$. Then $\varphi_\theta \in C_c^\infty(\mathbb{R})$ with $\operatorname{supp} \varphi_\theta \subseteq [a,b]$. $(\sigma * \varphi_\theta)(x) = (\sigma * \varphi)(x+\theta)$ $\forall x \in \mathbb{R}$. $\deg(\sigma * \varphi) = \deg(\sigma * \varphi_\theta) \le l$.] Fix $a=0$, $b=1$. $\forall \varphi \in C_c^\infty(\mathbb{R})$, $\exists N \in \mathbb{N}$ s.t. $\operatorname{supp}(\varphi) \subseteq [-N, N]$. By partition of unity, $\exists\, \varphi_i \in C_c^\infty(\mathbb{R})$ with $\operatorname{supp}(\varphi_i) \subseteq [-N+i-1, -N+i]$, $i = 1, \cdots, 2N$, and $\varphi = \sum_{i=1}^{2N} \varphi_i$. So, each $\sigma * \varphi_i$ is a polynomial of degree $\le l$. Hence, degree$(\sigma * \varphi) \le l$.

QED