Lecture 4, Friday, April 12, 2024.

- A univ. approx. thm (UAT) and its proof.
- A corollary. $\forall L \geq 2$.
- Remarks on the result / proof.

A Universal Approximation Theorem (UAT)

Theorem (Cybenko 1989, Funahashi 1989, and Hornik, Stinchcombe, and White 1989) Let $\sigma \in C(\mathbb{R})$ be such that $\sigma(-\infty) = 0$ and $\sigma(+\infty) = 1$. Then $NN(\sigma, 2, d, 1)$ is dense in $C(K)$ for any compact $K \subseteq \mathbb{R}^d$ and any $d \in \mathbb{N}$.

Proof Fix $d \in \mathbb{N}$ and $K \subseteq \mathbb{R}^d$ (compact). Assume $\exists f \in C(K) \setminus \overline{N_\sigma}$, where $N_\sigma = NN(\sigma, 2, d, 1)$. We show that there is a contradiction.

Step 1 By the Hahn-Banach Theorem, $\exists \varphi \in C(K)^*$ s.t. $\varphi(f) \neq 0$ and $\varphi(\Phi) = 0 \ \forall \Phi \in \overline{N_\sigma}$.

Step 2 By Riesz's Theorem, $\exists \mu \in \mathcal{M}(K)$ (the set of signed, finite Radon measures on K) s.t.

$$\varphi(g) = \int_K g \, d\mu \quad \forall g \in C(K).$$

In particular,

$$0 = \varphi(\Phi) = \int_K \Phi \, d\mu \quad \forall \Phi \in N_\sigma. \qquad (*)$$

Step 3 This leads to $\mu = 0$ by the lemma below.

Hence, $\varphi = 0$ Contradicting $\varphi(f) \neq 0$. QED

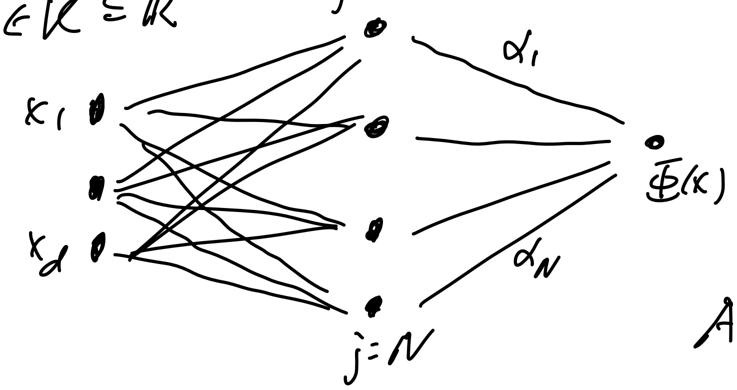
Remarks

- $\sigma(-\infty) = 0$, $\sigma(\infty) = 1$ can be replaced by $\sigma(-\infty) = a$, $\sigma(+\infty) = b$, $a \neq b$.
- Any $\Phi \in N_\sigma = NN(\sigma, 2, d, 1)$ can be expressed as

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

where $w_j \in \mathbb{R}^d$, $\alpha_j, \theta_j, \beta \in \mathbb{R}$, $j = 1, \dots, N$, $N = W(\Phi)$. Note that we assume $\sigma \not\equiv$ const. So, by setting all $w_j \equiv 0$, we can absorb β. i.e., $NN(\sigma, 2, d, 1) = \{$all Φ in $(*)$ with $\beta = 0\}$.

$x \in K \subseteq \mathbb{R}^d$



$$\Phi(x) = A_2(\sigma(A_1 x + b_1)) + b_2$$

$$A_1 = \begin{bmatrix} w_1^T \\ \vdots \\ w_N^T \end{bmatrix}_{N\times d} \qquad b_1 = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_N \end{bmatrix}.$$

$$A_2 = [\alpha_1 \cdots \alpha_N]_{1\times N} \qquad b_2 = \beta \in \mathbb{R}^1$$

$L=2$, $N_0 = d$, $N_L = 1$, $W = N$.

(*) Now, (*) is equivalent to $\mu(K) = 0$ and

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

If $d\mu(x) = \rho(x)dx$ for some $\rho \in C(K)$ then

$$\int_K \sigma(w^T \cdot x + \theta)\, \rho(x) dx = 0 \quad \forall w \in \mathbb{R}^d, \forall \theta \in \mathbb{R}.$$

We want to show then $\rho(x) \equiv 0$. $(x \in K)$. Analogue of

$$\int_0^1 x^n u(x) dx = 0 \ (\forall n = 0, 1, 2, \cdots),\ u \in C([0,1]) \Longrightarrow u(x) \equiv 0.\ (0 \le x \le 1).$$

Notation: $\forall d \in \mathbb{N}$

$$\mathrm{Aff}(d) = \{\text{affine maps } \mathbb{R}^d \to \mathbb{R}\}$$
$$= \{a : \mathbb{R}^d \to \mathbb{R} : a(x) = w \cdot x + \theta,\ w \in \mathbb{R}^d, \theta \in \mathbb{R}\}$$

<u>Definition</u> (discrimination; modified from Cybenko 1989)

A bounded and Lebesgue measurable function $\sigma : \mathbb{R} \to \mathbb{R}$ is discrimatory if for any $d \in \mathbb{N}$ and any $\mu \in \mathcal{M}_c(\mathbb{R}^d)$

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

Cybenko's def. is "local" w.r.t. a $K \subseteq \mathbb{R}^d$ (compact):

$\mu \in \mathcal{M}(K)$: $\int_K \sigma(w^T x + \theta)\, d\mu(x) = 0\ \forall w, \theta \Longrightarrow \mu = 0$.

Are the local version and global one equivalent?

<u>Lemma</u> Let $\sigma : \mathbb{R} \to \mathbb{R}$ be measurable, bounded, and satisfy $\sigma(-\infty) = 0$ and $\sigma(+\infty) = 1$ Then σ is discrimatory.

(Note: see Lecture 5 for some unified view of proof.)

Proof Fix $d \in \mathbb{N}$. Fix $\mu \in \mathcal{M}_c(\mathbb{R}^d)$. Denote $K = \operatorname{supp} \mu$, which is a compact subset of $\mathbb{R}^d$. Assume

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

We show that $\mu = 0$.

$\forall w \in \mathbb{R}^d$, $\forall \theta \in \mathbb{R}$, $\forall \alpha \in \mathbb{R}$. $\forall \lambda \in \mathbb{R}$, define

$$\sigma_\lambda(x) = \sigma(\lambda(w^T x + \theta) + \alpha) \quad \forall x \in \mathbb{R}^d.$$

We have

$$\lim_{\lambda \to +\infty} \sigma_\lambda(x) = \gamma(x) = \begin{cases} 1 & \text{if } w^T x + \theta > 0, \\ 0 & \text{if } w^T x + \theta < 0. \\ \sigma(\alpha) & \text{if } w^T x + \theta = 0. \end{cases}$$

Since $\int_K \sigma_\lambda(x)\, d\mu(x) = 0$, and σ_λ is bounded, by the Lebesgue Dominated Convergence Thm.

$$0 = \lim_{\lambda \to +\infty} \int_K \sigma_\lambda(x)\, d\mu(x) = \int_K \gamma(x)\, d\mu(x)$$

$$= \sigma(\alpha)\, \mu(\{x \in K : w^T x + \theta = 0\}) + \mu(\{x \in K : w^T x + \theta > 0\}).$$

Let $\alpha \to +\infty$. We get [So, $\sigma(+\infty) = 1$ can be replaced by $\sigma(+\infty) = b \neq 0$]

$$\mu(K \cap H^+_{w,\theta}) = 0 \quad \forall w \in \mathbb{R}^d, \forall \theta \in \mathbb{R},$$

where $H^+_{w,\theta} = \{x \in \mathbb{R}^d : w^T x + \theta \geq 0\}$.

We now show that $\mu = 0$ in $\mathcal{M}(K)$. Fix $w \in \mathbb{R}^d$. $\forall h \in L^\infty(\mathbb{R})$. Define

$$F_w(h) = \int_{\mathbb{R}^d} h(w^T x)\, d\mu(x) = \int_K h(w^T x)\, d\mu(x).$$

Then $F_w \in (L^\infty(\mathbb{R}))^*$. For $\theta \in \mathbb{R}$, set $h = \chi_{[\theta, \infty)}$. Then

$$F_w(\chi_{[\theta,\infty)}) = \int_K \chi_{[\theta,\infty)}(w^T x)\, d\mu(x) = \mu(K \cap H^+_{w,\theta}) = 0.$$

If $\theta_1 < \theta_2$ then

$$F_w(\chi_{[\theta_1, \theta_2)}) = F_w(\chi_{[\theta_1,\infty)}) - F_w(\chi_{[\theta_2,\infty)}) = 0.$$

Hence $F(h) = 0$ if h is a step function.

Now, for $h(u) = \cos(2\pi u)$ $(u \in \mathbb{R})$, there exist step functions $h_k \in L^\infty(\mathbb{R})$ such that $h_k \to h$ uniformly on $\mathbb{R}$ as $k \to \infty$. Thus.

$$F_w(h) = \int_K \cos(2\pi w^T x)\, d\mu(x) = \lim_{k\to\infty} F_w(h_k) = 0.$$

Similarly,

$$\int_K \sin(2\pi w^T x)\, d\mu(x) = 0.$$

Thus,

$$\int_{\mathbb{R}^d} \cos(2\pi w^T x)\, d\mu(x) = \int_{\mathbb{R}^d} \sin(2\pi w^T x)\, d\mu(x) = 0 \quad \forall w \in \mathbb{R}^d.$$

Hence, the Fourier transform of μ, defined by

$$\hat{\mu}(w) = \int_{\mathbb{R}^d} e^{-i 2\pi w^T x}\, d\mu(x) \quad (\forall w \in \mathbb{R}^d),$$

satisfies $\hat{\mu}(w) = 0$ $\forall w \in \mathbb{R}^d$. Hence, $\mu = 0$. <u>QED</u>

<u>Add some details.</u> $\mu \in \mathcal{M}(\mathbb{R}^d)$, $\hat{\mu} = 0 \Rightarrow \mu = 0$.

<u>Pf</u> $\forall g \in C_c(\mathbb{R}^d, \mathbb{C})$. $\hat{g}(\xi) = \int_{\mathbb{R}^d} g(x) e^{-2\pi i \xi \cdot x}\, dx$, $\forall \xi \in \mathbb{R}^d$. $\hat{g} \in BC(\mathbb{R}^d)$ and decays rapidly ($\hat{g} \in \mathcal{S}$). Moreover,

$$g(x) = \int_{\mathbb{R}^d} \hat{g}(\xi) e^{2\pi i \xi \cdot x}\, d\xi \quad \forall x \in \mathbb{R}^d.$$

Now, by Fubini's Thm with the fact that $\|\mu\| < \infty$ and the definition of $\hat{\mu}$, we get

$$\int_{\mathbb{R}^d} g(x)\, d\mu(x) = \int_{\mathbb{R}^d} \left[\int_{\mathbb{R}^d} \hat{g}(\xi) e^{2\pi i \xi \cdot x}\, d\xi \right] d\mu(x)$$

$$= \int_{\mathbb{R}^d} \hat{g}(\xi) \left[\int_{\mathbb{R}^d} e^{2\pi i \xi \cdot x}\, d\mu(x) \right] d\xi$$

$$= \int_{\mathbb{R}^d} \hat{g}(\xi)\, \hat{\mu}(-\xi)\, d\xi$$

$$= 0.$$

But $C_c(\mathbb{R}^d, \mathbb{C})$ is dense in $C_0(\mathbb{R}^d, \mathbb{C})$. Thus,

$$\int_{\mathbb{R}^d} g\, d\mu = 0 \quad \forall g \in C_0(\mathbb{R}^d, \mathbb{C}).$$

Finally, $C_0(\mathbb{R}^d, \mathbb{C})^* \cong \mathcal{M}(\mathbb{R}^d)$ (isomorphic, isometric). Hence $\mu = 0$.

What about $\sigma = \mathrm{ReLU}$?
(see next lecture for a simplified proof, similar though.)

<u>Theorem</u> $\mathcal{NN}(\mathrm{ReLU}, 2, d, 1)$ is dense in $C(K)$ for any compact $K \subseteq \mathbb{R}^d$.

<u>Proof</u> By the proof of the above theorem, we need only to show the following: If $\mu \in \mathcal{M}(K)$ and ($\sigma = \mathrm{ReLU}$)

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

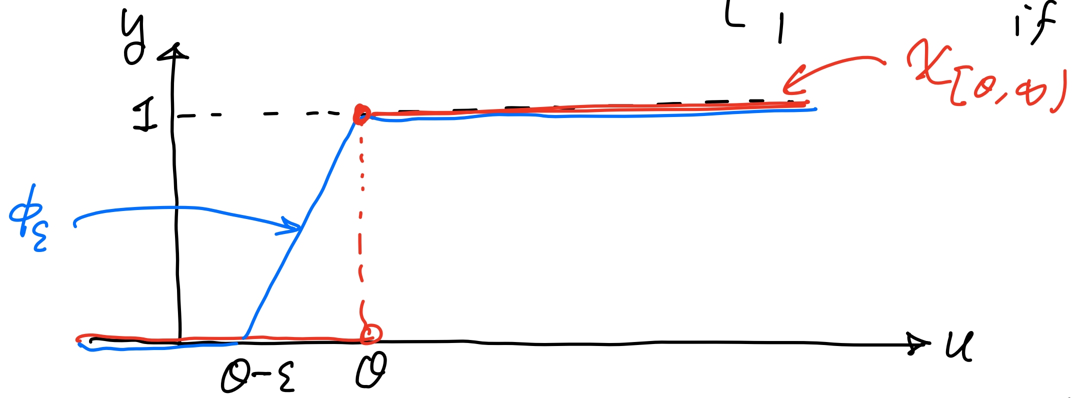
then, for any $w \in \mathbb{R}^d$ and $\theta \in \mathbb{R}$.

$$\int_K \chi_{[\theta,\infty)}(w\cdot x)\, d\mu(x) = 0. \qquad (**)$$

This is done by approximating $\chi_{[\theta,\infty} = \chi_{[\theta,\infty)}(u)$ $(u \in \mathbb{R})$ by $\phi_\varepsilon = \phi_\varepsilon(u)$ $(u \in \mathbb{R})$ using σ, so that $\Phi_\varepsilon(x) := \phi_\varepsilon(w\cdot x) \in \mathcal{NN}(\sigma, 2, d, 1)$.

Let $0 < \varepsilon < 1$. Define

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



Then, $\forall u \in \mathbb{R}$: $\phi_\varepsilon(u) \longrightarrow \chi_{[\theta,\infty)}(u)$ as $\varepsilon \to 0^+$ (check the cases: $u > \theta$, $u = \theta$, $u < \theta$. For $u \ge \theta$, $\phi_\varepsilon(u) = \chi_{[\theta,\infty)}(u)$ $\forall \varepsilon > 0$.). Let $w \in \mathbb{R}^d$. Define $\Phi_{w,\varepsilon}(x) = \phi_\varepsilon(w\cdot x)$ $(x \in \mathbb{R}^d)$. Then, $\Phi_{w,\varepsilon} \in \mathcal{NN}(\sigma, 2, d, 1)$ and

$$\Phi_{w,\varepsilon}(x) \longrightarrow \chi_{[\theta,\infty)}(w\cdot x), \quad \forall x \in \mathbb{R}^d \quad \left(\begin{matrix}\forall x, \text{ not} \\ \mu\text{-a.e. } x\end{matrix}\right)$$

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

$$0 = \int_K \Phi_{w,\varepsilon}(x)\, d\mu(x) \longrightarrow \int_K \chi_{[\theta,\infty)}(x)\, d\mu(x).$$

Thus, $(**)$ is true.

<u>QED</u>

Questions

- Other proofs / ideas?
- More general σ?
- If $NN(\sigma, 2, d, 1)$ is dense in $C(K)$, what about $NN(\sigma, L, d, 1)$ with $L \geq 3$?

First, we give a positive answer to the last question. Note that this would be trivial if

$$NN(\sigma, L, d, 1) \subseteq NN(\sigma, L+1, d, 1).$$

However, we don't know if this is true in general. Still, we have the following:

Theorem Let $\sigma \in C(\mathbb{R})$ and $K \subseteq \mathbb{R}^d$ be compact. Assume that $NN(\sigma, 2, d, 1)$ is dense in $C(K)$. Then

(1) $NN(\sigma, L, d, 1)$ is dense in $C(K)$ for any $L \geq 3$.

(2) Let $d' \in \mathbb{N}$ with $d' \geq 2$. Then $NN(\sigma, L, d, d')$ is dense in $C(K, \mathbb{R}^{d'})$.

Proof (1) Let $f \in C(K)$ and $\varepsilon > 0$. By the theorem, $\exists\, \Phi \in NN(\sigma, L=2, d, 1)$ s.t.

$$\max_{x \in K} |\Phi(x) - f(x)| < \varepsilon/2. \qquad (A)$$

Let $-\infty < a < b < \infty$ be such that $K \subseteq [a, b]^d$. Denote $Q = [a-1, b+1]^d$. Then $\Phi \in C(Q)$. Hence, Φ is uniformly continuous on Q. Thus, there exists $\delta \in (0, 1)$ such that

$$|\Phi(y) - \Phi(y')| < \varepsilon/2 \quad \forall y, y' \in Q, \text{ s.t. } \|y - y'\| < \delta. \qquad (B)$$

($\|\cdot\|$ is the Euclidean norm of $\mathbb{R}^d$.)

By the theorem again, $\exists\, \phi_j \in NN(\sigma, L=2, d, 1)$, $j = 1, \cdots, d$, s.t.

$$\max_{x \in Q} |\phi_j(x) - x_j| < \delta/\sqrt{d}, \quad j = 1, \cdots, d.$$

Now, by the parallelization with shared inputs we can define $\phi \in NN(\sigma, L=2, d, d)$ by $\phi(x) = \begin{bmatrix} \phi_1(x) \\ \vdots \\ \phi_d(x) \end{bmatrix}$, $\forall x \in Q$

We have

$$\max_{x \in Q} \| \phi(x) - x \| < \delta .$$

This also implies that $\phi(x) \in Q$ if $x \in K$, as $a \le x_j \le b$ $(j=1,\cdots,d)$, and

$$|\phi_j(x) - x_j| \le \| \phi(x) - x \| < \delta < 1 \quad j=1,\cdots,d.$$

Thus by (B),

$$\left| \Phi(\phi(x)) - \Phi(x) \right| < \varepsilon/2 \quad \forall x \in K. \qquad (C)$$

Finally, by the concatenation/composition, we define $\widetilde{\Phi} = \Phi \circ \phi$

$\widetilde{\Phi} \in NN(\sigma, L+1, d, 1) = NN(\sigma, 3, d, 1)$.

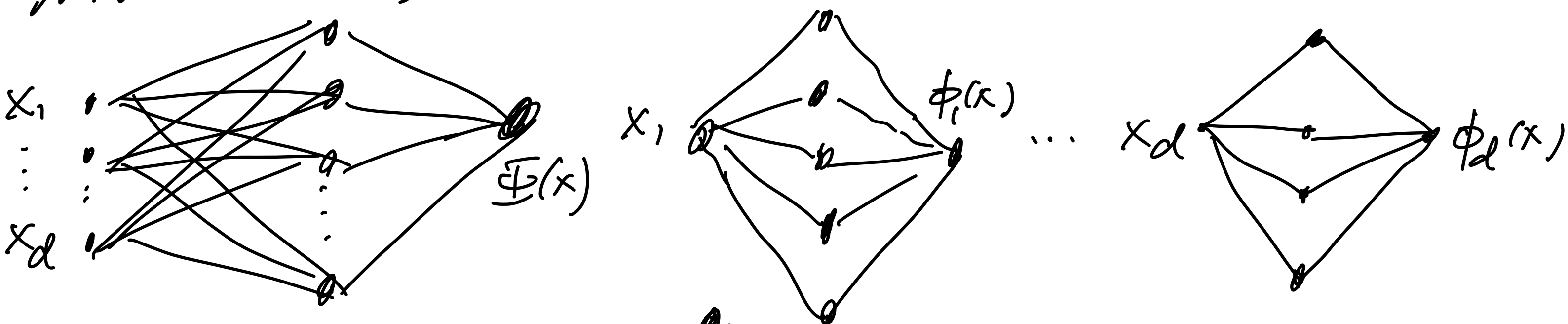
$$\mathbb{R}^d \ni x \xrightarrow{\ \phi\ } \phi(x) \in \mathbb{R}^d \xrightarrow{\ \Phi\ } \Phi(\phi(x)) \in \mathbb{R}.$$

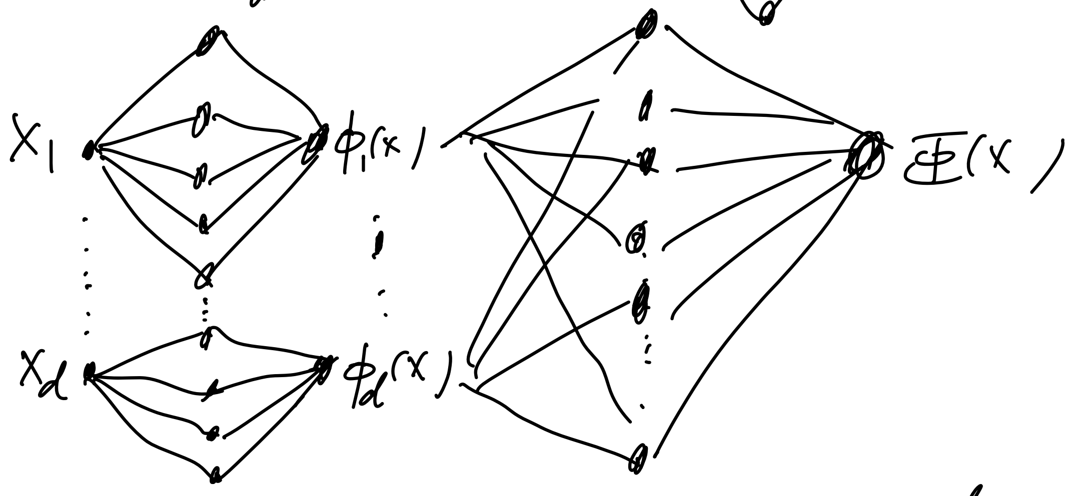
For any $x \in K$, we have by (A) and (C) that

$$\left| \widetilde{\Phi}(x) - f(x) \right| \le \left| \Phi(\phi(x)) - \Phi(x) \right| + \left| \Phi(x) - f(x) \right| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Thus, $\overline{NN(\sigma, 3, d, 1)} = C(K)$.

In general, if $\overline{NN(\sigma, L, d, 1)} = C(K)$, $L \ge 3$, then the same argument by adding the approximation of the identity map by $NN(\sigma, L=2, d, d)$, will lead to $\overline{NN(\sigma, L+1, d, 1)} = C(K)$.





$$\widetilde{\Phi}(x) = \Phi(\phi(x)), \quad \forall x \in \mathbb{R}^d.$$

(2) This follows from the theorem and the construction of an NN by parallelization with shared inputs. QED

Now, discuss the first issue. A formula by Irie and Miyake (1988), Funahashi's proof.

$$f(x) \approx \sum_j \alpha_j \sigma(w_j \cdot x - \theta_j), \text{ weighted sum.}$$

$$f(x) = \int_{\mathbb{R}^d} \int_{\mathbb{R}} \sigma(w \cdot x - \theta)\, \alpha(w,\theta)\, dw\, d\theta \quad \text{— expected.}$$

$\alpha = \alpha(w,\theta)$ — distribution of weights, dep. on w, θ.

What is $\alpha(w,\theta)$. It should dep. on σ, f. The Fourier transform of f is

$$\hat{f}(w) = \int_{\mathbb{R}^d} f(x) e^{-2\pi i w \cdot x} dx$$

$$f(x) = \check{\hat{f}}(x) = \int_{\mathbb{R}^d} \hat{f}(w) e^{2\pi i w \cdot x} dw$$

So,

$$\hat{f}(w) \underline{\underline{e^{2\pi i w \cdot x}}} = \int_{\mathbb{R}} \sigma(w \cdot x - \theta)\, \alpha(w,\theta)\, d\theta.$$

What about $\hat{\sigma}(\alpha)$?

$$\hat{\sigma}(\xi) = \int_{\mathbb{R}} \sigma(s) e^{-2\pi i s \xi} ds$$

$$= \int_{\mathbb{R}} \sigma(w \cdot x - \theta) e^{-2\pi i (w \cdot x - \theta)\xi} d\theta$$

$$= \underline{\underline{e^{-2\pi i w \cdot x}}} \int_{\mathbb{R}} \sigma(w \cdot x - \theta) e^{2\pi i \theta \xi} d\theta$$

$$\frac{\hat{f}(w)}{\hat{\sigma}(\xi)} \int_{\mathbb{R}} \sigma(w \cdot x - \theta) e^{2\pi i \theta \xi} d\theta = \int_{\mathbb{R}} \sigma(w \cdot x - \theta)\, \alpha(w,\theta)\, d\theta.$$

Choose ξ = some number, e.g. $\xi = 1$. or $\xi = \xi_0$ fixed.

$$\alpha(w,\theta) = \frac{\hat{f}(w)}{\hat{\sigma}(1)} e^{2\pi i \theta},$$

Then

$$\boxed{f(x) = \int_{\mathbb{R}^d} \int_{\mathbb{R}} \sigma(w\cdot x-\theta)\, \frac{\hat{f}(w)}{\hat{\sigma}(\xi_0)}\, e^{2\pi i \theta \xi_0}\, dw\, d\theta.} \quad (*)$$

This is called the Irie-Miyake integral formula.

<u>Theorem (Irie-Miyake, 1988)</u> Let $\sigma \in L^1(\mathbb{R})$ and $f \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$. Let $\hat{\sigma}$ and $\hat{f}$ denot the Fourier tranfoms of σ and f, respectively. Assume $\hat{\sigma}(\xi_0) \neq 0$ for some $\xi_0 \in \mathbb{R}$, then the Irie-Miyake integral formula $(*)$ holds true. <u>QED</u>

Funahashi used this formula to prove that $NN(\sigma, 2, d, 1)$ is dense in $C(K)$.

<u>Theorem</u> Let $\sigma \in C(\mathbb{R})$ be monotonically increasing, bounded, and nonconstant. Then for any $K \subseteq C(\mathbb{R}^d)$ compact, $NN(\sigma, 2, d, 1)$ is dense in $C(K)$.

<u>Sketch of Funahashi's proof</u>

- Without loss of generality, assume $f \in C_c^\infty(\mathbb{R}^d)$.
- Define for $\alpha \in \mathbb{R}$, $\delta > 0$
$$\psi(x) = \sigma\left(\frac{x}{\delta}+\alpha\right) - \sigma\left(\frac{x}{\delta}-\alpha\right) \quad (x \in \mathbb{R}).$$

Then $\psi \in L^1(\mathbb{R})$ and $\psi(1) \neq 0$ for some $\delta > 0$, $\alpha \in \mathbb{R}$.

- Replacing σ by ψ in the <u>Irie-Miyake's integral</u> formula to get (using ψ, not σ.)
$$f(x) = \int_{\mathbb{R}^d} dw \int_{\mathbb{R}} dw_0\, [\cdots]$$
- Approximate $f(x)$ by $\int_{[-A,A]^d} dw \int_{\mathbb{R}} dw_0\, [\cdots]$ for $A \gg 1$.
- Riemann sum approximates the integral
$$\int_{[-A,A]^d} dw \int_{[-A',A']} dw_0\, [\cdots]$$

uniformly for $x \in K$.

⊙ Riemann sums are in $NN(\sigma, 2, d, 1)$. QED

The proof by Hornik, Stinchcombe, and White (1989) need σ to be increasing.

⊙ By the Stone-Weierstrass Thm,

$\mathcal{A}_d = \{$ linear combinations of finite products of $\sigma \circ A$ of affine functions $A: \mathbb{R}^d \to \mathbb{R} \}$

is dense in $C(K)$.

⊙ If σ is a special cos-squashing function σ_c, then $C_d = \{$ linear combinations of $\sigma_c \circ A$ of all affine functions $A: \mathbb{R}^d \to \mathbb{R}\}$ is dense in $C(K)$

⊙ The cosine function can be approximated by linear combinations of $\sigma \circ A$ with $A: \mathbb{R} \to \mathbb{R}$ affine. (See Gallant & White (1988) — Fourier net.) QED

The 4th approach: Proof by Leshno et al. (1993) of most general case. (next lecture.)

Finally, how general a σ can be?

Hornik 1991: with a Fourier analysis argument, relaxed the conditions for σ to be measurable, bounded, and non constant, along the line of Cybenko's (next lecture).

The most general result is the following:

Theorem (Leshno et al. 1993) Assume $\sigma: \mathbb{R} \to \mathbb{R}$ satisfies the Lebesgue measure of $\{x \in \mathbb{R}: \sigma$ is discont. at $x\}$ is 0.

⊙ If σ is not a polynomial (a.e. $\mathbb{R}$) then for any compact $K \subseteq \mathbb{R}^d$, $NN(\sigma, 2, d, 1)$ is dense in $C(K)$.

⊙ Conversely, if $NN(\sigma, 2, d, 1)$ is dense in $C(K)$ for any compact $K \subseteq \mathbb{R}^d$, then σ is not a polynomial.

(To be discussed next time.)