

Kähler and Tropical Geometry for the Universal Approximation Theorem

João Camarneiro

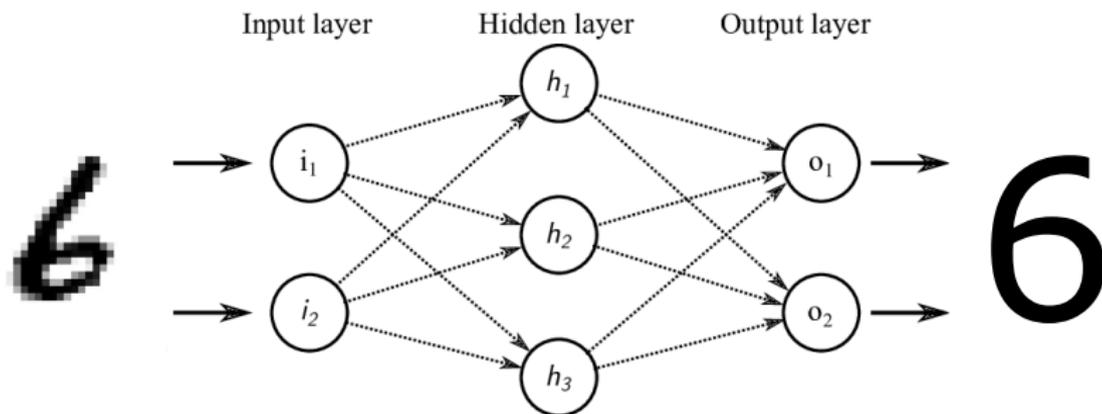
Tutor: José Mourão

Encontro Nacional *Novos Talentos em Matemática*

28 de julho de 2021

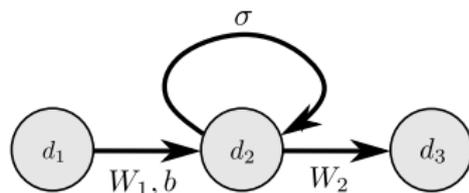
Redes neuronais artificiais

- Camada de *input*, camada(s) oculta(s), camada de *output*
- Pesos das arestas
- Enviesamentos (*bias*) dos neurónios
- Função de ativação



Aproximador de funções

Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ uma função desconhecida que pretendemos aproximar.

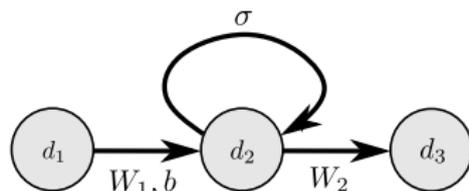


$$f_{W_1, W_2, b}(x) = W_2 \sigma(W_1 x + b)$$

onde W_1, W_2 são aplicações lineares.

Aproximador de funções

Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ uma função desconhecida que pretendemos aproximar.



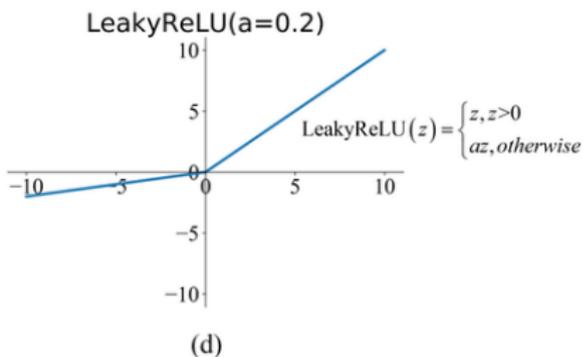
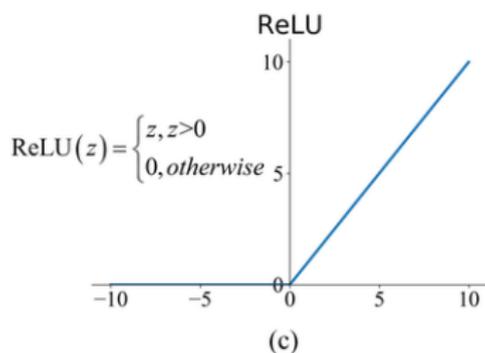
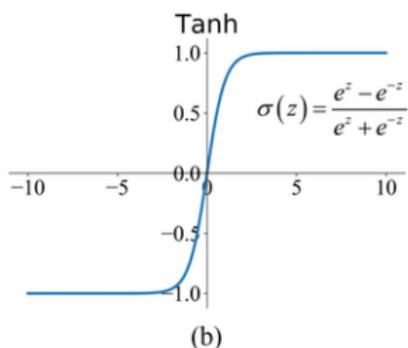
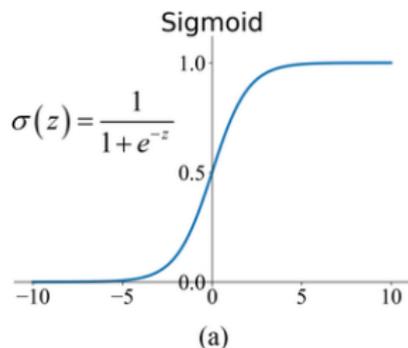
$$f_{W_1, W_2, b}(x) = W_2 \sigma(W_1 x + b)$$

onde W_1, W_2 são aplicações lineares.

Teorema de Aproximação Universal para redes neuronais artificiais

Seja $K \subseteq \mathbb{R}^{d_1}$ um compacto e $f : K \rightarrow \mathbb{R}^{d_3}$ uma função contínua. Então, qualquer que seja $\varepsilon > 0$, existem $d_2 \in \mathbb{Z}^+$, $W_1 \in \text{Mat}_{d_2 \times d_1}$, $W_2 \in \text{Mat}_{d_3 \times d_2}$ e $b \in \mathbb{R}^{d_2}$ tais que $\|f - f_{W_1, W_2, b}\|_{L^2(K)} < \varepsilon$.

Funções de ativação usuais



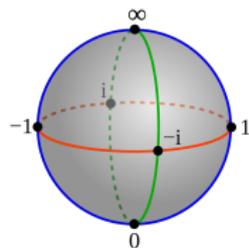
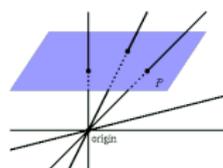
Espaços projetivos complexos

Uma relação de equivalência em $\mathbb{C}^{n+1} \setminus \{0\}$:

$$z \sim w \iff \text{existe } \lambda \in \mathbb{C}^* \text{ tal que } z = \lambda w$$

Então, o **espaço projetivo complexo** de dimensão n é definido como o conjunto das classes de equivalência de \sim .

$$\mathbb{C}P^n = \frac{\mathbb{C}^{n+1} \setminus \{0\}}{\sim}$$



Exemplo

- $[2 : 1] = [-6 : -3] \in \mathbb{C}P^1$
- $[i : 0 : -2] = [1 : 0 : 2i] \in \mathbb{C}P^2$

- Um ponto $[Z_0 : Z_1 : \cdots : Z_n] \in \mathbb{C}\mathbb{P}^n$ é uma classe de equivalência com múltiplos representantes.

- Um ponto $[Z_0 : Z_1 : \cdots : Z_n] \in \mathbb{C}\mathbb{P}^n$ é uma classe de equivalência com múltiplos representantes.
- Em $U_0 = \{[Z_0 : Z_1 : \cdots : Z_n] \in \mathbb{C}\mathbb{P}^n : Z_0 \neq 0\}$ estão bem definidas as coordenadas locais $z : U_0 \subseteq \mathbb{C}\mathbb{P}^n \rightarrow \mathbb{C}^n$,

$$z_j = \frac{Z_j}{Z_0}, \quad U_0 \cong \mathbb{C}^n.$$

- Um ponto $[Z_0 : Z_1 : \cdots : Z_n] \in \mathbb{CP}^n$ é uma classe de equivalência com múltiplos representantes.
- Em $U_0 = \{[Z_0 : Z_1 : \cdots : Z_n] \in \mathbb{CP}^n : Z_0 \neq 0\}$ estão bem definidas as coordenadas locais $z : U_0 \subseteq \mathbb{CP}^n \rightarrow \mathbb{C}^n$,

$$z_j = \frac{Z_j}{Z_0}, \quad U_0 \cong \mathbb{C}^n.$$

- Podemos pensar nestes pontos através das suas coordenadas $z = (z_1, \dots, z_n) \in \mathbb{C}^n$.
- \mathbb{CP}^n é uma variedade de dimensão $2n$.

- Seja $U \subseteq \mathbb{C}\mathbb{P}^n$ o aberto onde $Z_j \neq 0$ para todo o $j = 0, 1, \dots, n$.

- Seja $U \subseteq \mathbb{C}\mathbb{P}^n$ o aberto onde $Z_j \neq 0$ para todo o $j = 0, 1, \dots, n$.
- Em U , temos $z_j = \frac{Z_j}{Z_0} \neq 0$. Podemos reescrever em *coordenadas polares*: $z_j = r_j e^{i\theta_j}$.

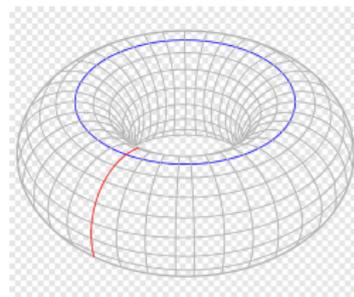
- Seja $U \subseteq \mathbb{C}\mathbb{P}^n$ o aberto onde $Z_j \neq 0$ para todo o $j = 0, 1, \dots, n$.
- Em U , temos $z_j = \frac{Z_j}{Z_0} \neq 0$. Podemos reescrever em *coordenadas polares*: $z_j = r_j e^{i\theta_j}$.
- $r_j > 0$ e portanto $r_j = e^{y_j}$ com $y_j \in \mathbb{R}$. Logo:

$$z_j = e^{y_j + i\theta_j} = e^{w_j}$$

- Vai ser conveniente trabalhar com a *coordenada logarítmica* $w = (y, \theta)$ e a sua parte real y .

Toros

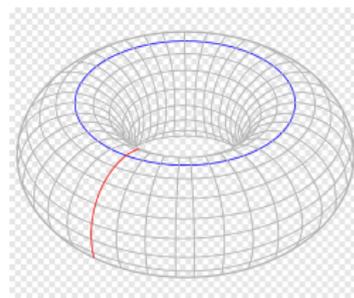
- Definimos o **toro** de dimensão n como $\mathbb{T}^n = (S^1)^n$.



Toros

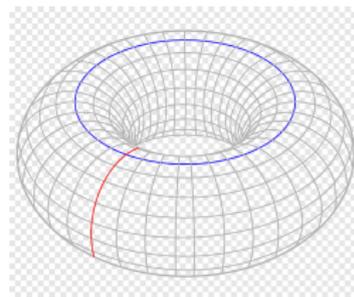
- Definimos o **toro** de dimensão n como $\mathbb{T}^n = (S^1)^n$.

$$S^1 = \{e^{it} : t \in \mathbb{R}\}$$



Toros

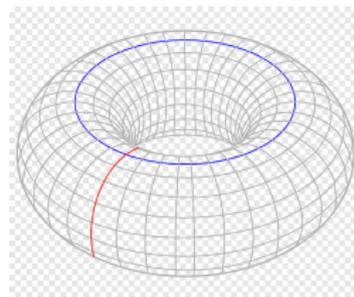
- Definimos o **toro** de dimensão n como $\mathbb{T}^n = (S^1)^n$.



$$S^1 = \{e^{it} : t \in \mathbb{R}\} \quad \{e^{i(t+is)} : t, s \in \mathbb{R}\}$$

Toros

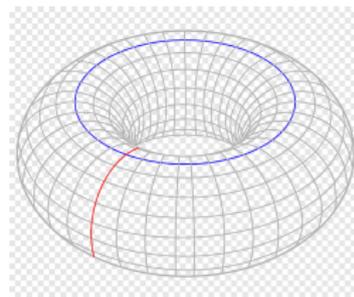
- Definimos o **toro** de dimensão n como $\mathbb{T}^n = (S^1)^n$.



$$S^1 = \{e^{it} : t \in \mathbb{R}\} \quad \{e^{i(t+is)} : t, s \in \mathbb{R}\} = \mathbb{C}^*$$

Toros

- Definimos o **toro** de dimensão n como $\mathbb{T}^n = (S^1)^n$.



$$S^1 = \{e^{it} : t \in \mathbb{R}\} \quad \{e^{i(t+is)} : t, s \in \mathbb{R}\} = \mathbb{C}^*$$

- Podemos considerar a **complexificação** do toro: $\mathbb{T}_{\mathbb{C}}^n = (\mathbb{C}^*)^n$.

Ação de \mathbb{T}^n em $\mathbb{C}\mathbb{P}^n$

Consideramos uma ação do toro \mathbb{T}^n em $\mathbb{C}\mathbb{P}^n$:

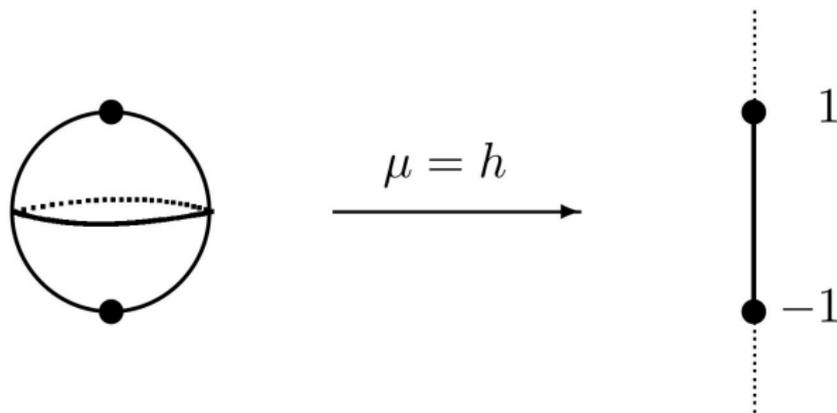
$$(e^{it_1}, \dots, e^{it_n}) \cdot [Z_0 : Z_1 : \dots : Z_n] = [Z_0 : e^{it_1} Z_1 : \dots : e^{it_n} Z_n]$$

Ação de \mathbb{T}^n em $\mathbb{C}\mathbb{P}^n$

Consideramos uma ação do toro \mathbb{T}^n em $\mathbb{C}\mathbb{P}^n$:

$$(e^{it_1}, \dots, e^{it_n}) \cdot [Z_0 : Z_1 : \dots : Z_n] = [Z_0 : e^{it_1} Z_1 : \dots : e^{it_n} Z_n]$$

Caso $n = 1$: rotações na esfera de Riemann; as órbitas são os paralelos.



Ação de \mathbb{T}^n em $\mathbb{C}\mathbb{P}^n$

A ação estende a uma ação do toro complexo $(\mathbb{C}^*)^n$:

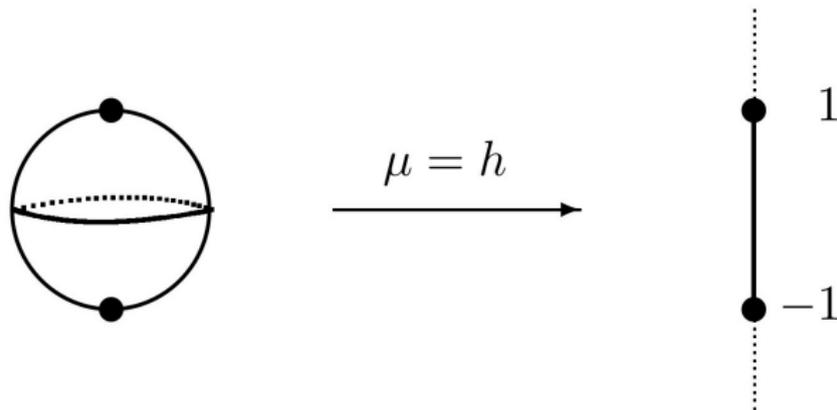
$$(w_1, \dots, w_n) \cdot [Z_0 : Z_1 : \dots : Z_n] = [Z_0 : w_1 Z_1 : \dots : w_n Z_n]$$

Ação de \mathbb{T}^n em $\mathbb{C}P^n$

A ação estende a uma ação do toro complexo $(\mathbb{C}^*)^n$:

$$(w_1, \dots, w_n) \cdot [Z_0 : Z_1 : \dots : Z_n] = [Z_0 : w_1 Z_1 : \dots : w_n Z_n]$$

Caso $n = 1$: órbita densa + polos



O caso do $\mathbb{C}\mathbb{P}^2$

$$(e^{it_1}, e^{it_2}) \cdot [Z_0 : Z_1 : Z_2] = [Z_0 : e^{it_1} Z_1 : e^{it_2} Z_2]$$

O caso do \mathbb{CP}^2

$$(e^{it_1}, e^{it_2}) \cdot [Z_0 : Z_1 : Z_2] = [Z_0 : e^{it_1} Z_1 : e^{it_2} Z_2]$$

Podemos tomar como aplicação momento para esta ação o seguinte $\mu : \mathbb{CP}^2 \rightarrow \mathbb{R}^2$:

$$\mu([Z_0 : Z_1 : Z_2]) = \left(\frac{|Z_1|^2}{|Z_0|^2 + |Z_1|^2 + |Z_2|^2}, \frac{|Z_2|^2}{|Z_0|^2 + |Z_1|^2 + |Z_2|^2} \right)$$

O caso do $\mathbb{C}\mathbb{P}^2$

$$(e^{it_1}, e^{it_2}) \cdot [Z_0 : Z_1 : Z_2] = [Z_0 : e^{it_1} Z_1 : e^{it_2} Z_2]$$

Podemos tomar como aplicação momento para esta ação o seguinte $\mu : \mathbb{C}\mathbb{P}^2 \rightarrow \mathbb{R}^2$:

$$\mu([Z_0 : Z_1 : Z_2]) = \left(\frac{|Z_1|^2}{|Z_0|^2 + |Z_1|^2 + |Z_2|^2}, \frac{|Z_2|^2}{|Z_0|^2 + |Z_1|^2 + |Z_2|^2} \right)$$

Nas coordenadas logarítmicas (y, θ) :

$$\mu(y_1, y_2, \theta_1, \theta_2) = \left(\frac{e^{2y_1}}{1 + e^{2y_1} + e^{2y_2}}, \frac{e^{2y_2}}{1 + e^{2y_1} + e^{2y_2}} \right)$$

O caso do \mathbb{CP}^2

$$(e^{it_1}, e^{it_2}) \cdot [Z_0 : Z_1 : Z_2] = [Z_0 : e^{it_1} Z_1 : e^{it_2} Z_2]$$

Podemos tomar como aplicação momento para esta ação o seguinte $\mu : \mathbb{CP}^2 \rightarrow \mathbb{R}^2$:

$$\mu([Z_0 : Z_1 : Z_2]) = \left(\frac{|Z_1|^2}{|Z_0|^2 + |Z_1|^2 + |Z_2|^2}, \frac{|Z_2|^2}{|Z_0|^2 + |Z_1|^2 + |Z_2|^2} \right)$$

Nas coordenadas logarítmicas (y, θ) :

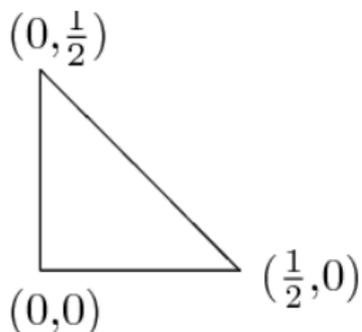
$$\mu(y_1, y_2, \theta_1, \theta_2) = \left(\frac{e^{2y_1}}{1 + e^{2y_1} + e^{2y_2}}, \frac{e^{2y_2}}{1 + e^{2y_1} + e^{2y_2}} \right)$$

Observação

No caso $n = 1$, fica $\mu(y) = \frac{e^{2y}}{1 + e^{2y}} = \frac{1}{1 + e^{-2y}}$, a função de ativação sigmoide.

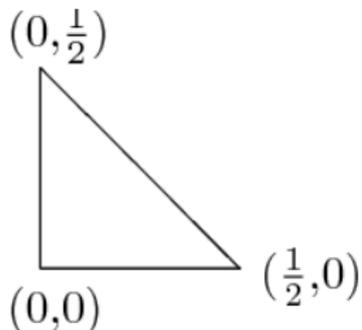
Aplicação momento e convexidade

Imagem da
aplicação momento
de $\mathbb{C}\mathbb{P}^2$ (em \mathbb{R}^2):



Aplicação momento e convexidade

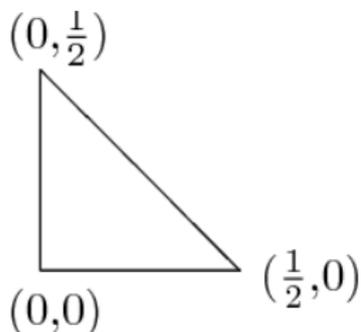
Imagem da
aplicação momento
de $\mathbb{C}P^2$ (em \mathbb{R}^2):



- É um polítopo convexo: o **polítopo momento**, P .

Aplicação momento e convexidade

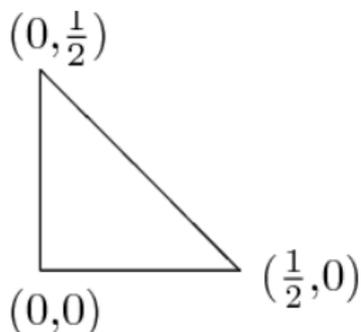
Imagem da
aplicação momento
de $\mathbb{C}P^2$ (em \mathbb{R}^2):



- É um polítopo convexo: o **polítopo momento**, P .
- Cada ponto no polítopo corresponde a uma órbita da ação de \mathbb{T}^2 .

Aplicação momento e convexidade

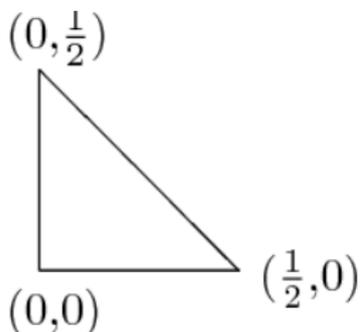
Imagem da
aplicação momento
de $\mathbb{C}P^2$ (em \mathbb{R}^2):



- É um polígono convexo: o **polígono momento**, P .
- Cada ponto no polígono corresponde a uma órbita da ação de \mathbb{T}^2 .
- Interior do polígono: órbita densa da ação de $\mathbb{T}_{\mathbb{C}}^2$

Aplicação momento e convexidade

Imagem da
 aplicação momento
 de $\mathbb{C}\mathbb{P}^2$ (em \mathbb{R}^2):



- É um polígono convexo: o **polígono momento**, P .
- Cada ponto no polígono corresponde a uma órbita da ação de \mathbb{T}^2 .
- Interior do polígono: órbita densa da ação de $\mathbb{T}^2_{\mathbb{C}}$
- Coordenadas simpléticas (x, θ) na órbita densa

$$x_1 = \frac{e^{2y_1}}{1 + e^{2y_1} + e^{2y_2}}, \quad x_2 = \frac{e^{2y_2}}{1 + e^{2y_1} + e^{2y_2}}$$

Podemos considerar a função de transição da coordenada y para a coordenada x .

$$\sigma : \mathbb{R}^2 \rightarrow P^0$$
$$\sigma(y_1, y_2) = \left(\frac{e^{2y_1}}{1 + e^{2y_1} + e^{2y_2}}, \frac{e^{2y_2}}{1 + e^{2y_1} + e^{2y_2}} \right)$$

Vamos utilizá-la como função de ativação.

Podemos considerar a função de transição da coordenada y para a coordenada x .

$$\sigma : \mathbb{R}^2 \rightarrow P^0$$
$$\sigma(y_1, y_2) = \left(\frac{e^{2y_1}}{1 + e^{2y_1} + e^{2y_2}}, \frac{e^{2y_2}}{1 + e^{2y_1} + e^{2y_2}} \right)$$

Vamos utilizá-la como função de ativação.

Comparando com a função sigmoide:

$$\tilde{\sigma}(y_1, y_2) = \left(\frac{e^{2y_1}}{1 + e^{2y_1}}, \frac{e^{2y_2}}{1 + e^{2y_2}} \right)$$

Teorema de Aproximação Universal

Seja $K \subseteq \mathbb{R}$ um compacto e $f : K \rightarrow \mathbb{R}$ uma função contínua. Então, qualquer que seja $\varepsilon > 0$, existem $d \in \mathbb{Z}^+$, aplicações lineares $W_1 : \mathbb{R} \rightarrow \mathbb{R}^d$, $W_2 : \mathbb{R}^d \rightarrow \mathbb{R}$ e um vetor $b \in \mathbb{R}^d$ tais que $\|f - f_{W_1, W_2, b}\|_{L^2(K)} < \varepsilon$.

$$f_{W_1, W_2, b}(x) = W_2 \sigma(W_1 x + b)$$

Teorema de Aproximação Universal

Seja $K \subseteq \mathbb{R}$ um compacto e $f : K \rightarrow \mathbb{R}$ uma função contínua. Então, qualquer que seja $\varepsilon > 0$, existem $d \in \mathbb{Z}^+$, aplicações lineares $W_1 : \mathbb{R} \rightarrow \mathbb{R}^d$, $W_2 : \mathbb{R}^d \rightarrow \mathbb{R}$ e um vetor $b \in \mathbb{R}^d$ tais que $\|f - f_{W_1, W_2, b}\|_{L^2(K)} < \varepsilon$.

$$f_{W_1, W_2, b}(x) = W_2 \sigma(W_1 x + b)$$

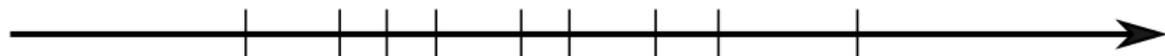
Ideia: Dividir \mathbb{R} em intervalos I_0, I_1, \dots, I_N de modo a que $I_j \cap K$ seja suficientemente pequeno.

Teorema de Aproximação Universal

Seja $K \subseteq \mathbb{R}$ um compacto e $f : K \rightarrow \mathbb{R}$ uma função contínua. Então, qualquer que seja $\varepsilon > 0$, existem $d \in \mathbb{Z}^+$, aplicações lineares $W_1 : \mathbb{R} \rightarrow \mathbb{R}^d$, $W_2 : \mathbb{R}^d \rightarrow \mathbb{R}$ e um vetor $b \in \mathbb{R}^d$ tais que $\|f - f_{W_1, W_2, b}\|_{L^2(K)} < \varepsilon$.

$$f_{W_1, W_2, b}(x) = W_2 \sigma(W_1 x + b)$$

Ideia: Dividir \mathbb{R} em intervalos I_0, I_1, \dots, I_N de modo a que $I_j \cap K$ seja suficientemente pequeno.



Teorema de Aproximação Universal

Seja $K \subseteq \mathbb{R}$ um compacto e $f : K \rightarrow \mathbb{R}$ uma função contínua. Então, qualquer que seja $\varepsilon > 0$, existem $d \in \mathbb{Z}^+$, aplicações lineares $W_1 : \mathbb{R} \rightarrow \mathbb{R}^d$, $W_2 : \mathbb{R}^d \rightarrow \mathbb{R}$ e um vetor $b \in \mathbb{R}^d$ tais que $\|f - f_{W_1, W_2, b}\|_{L^2(K)} < \varepsilon$.

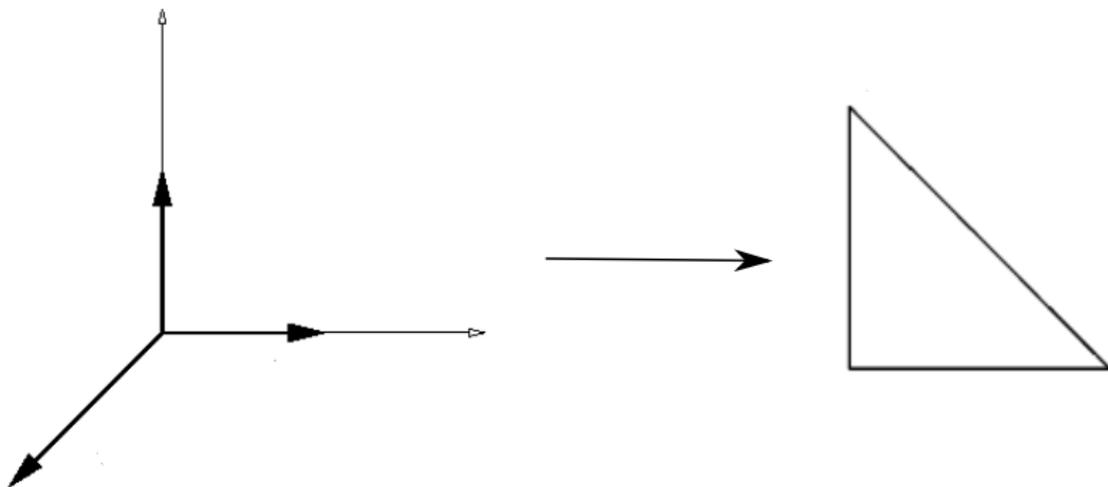
$$f_{W_1, W_2, b}(x) = W_2 \sigma(W_1 x + b)$$

Ideia: Dividir \mathbb{R} em intervalos I_0, I_1, \dots, I_N de modo a que $I_j \cap K$ seja suficientemente pequeno.



Aproximar f por uma função em escada definida nestes intervalos.

Fan de $\mathbb{C}P^n$



Degeneração tropical

- Função de ativação $\sigma : \mathbb{R}^n \rightarrow P^0$
- Degeneração tropical: $\sigma_t(y) = \sigma(ty)$, $\sigma_\infty = \lim_{t \rightarrow +\infty} \sigma_t$

Degeneração tropical

- Função de ativação $\sigma : \mathbb{R}^n \rightarrow P^0$
- Degeneração tropical: $\sigma_t(y) = \sigma(ty)$, $\sigma_\infty = \lim_{t \rightarrow +\infty} \sigma_t$
- No caso $n = 1$: $\sigma(ty) = \frac{e^{2ty}}{1+e^{2ty}}$

$$\sigma_\infty(y) = \begin{cases} 0 & \text{se } y < 0 \\ \frac{1}{2} & \text{se } y = 0 \\ 1 & \text{se } y > 0 \end{cases}$$

Degeneração tropical

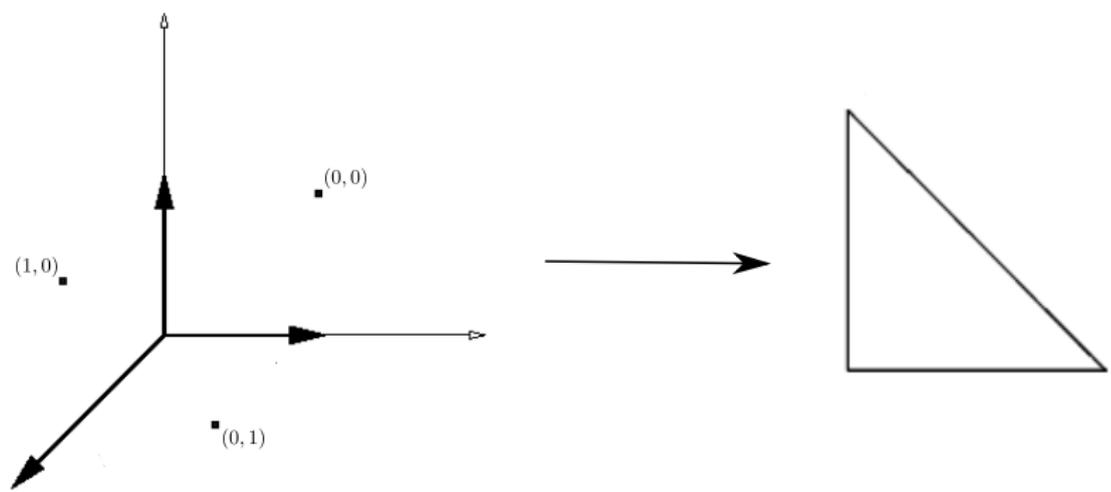
- Função de ativação $\sigma : \mathbb{R}^n \rightarrow P^\circ$
- Degeneração tropical: $\sigma_t(y) = \sigma(ty)$, $\sigma_\infty = \lim_{t \rightarrow +\infty} \sigma_t$
- No caso $n = 1$: $\sigma(ty) = \frac{e^{2ty}}{1+e^{2ty}}$

$$\sigma_\infty(y) = \begin{cases} 0 & \text{se } y < 0 \\ \frac{1}{2} & \text{se } y = 0 \\ 1 & \text{se } y > 0 \end{cases}$$

Lema

A função limite $\sigma_\infty = \lim_{t \rightarrow +\infty} \sigma_t$ é em escada: toma um valor constante no interior de cada cone do fan de $\mathbb{C}P^n$.

Exemplo (caso $n = 2$):



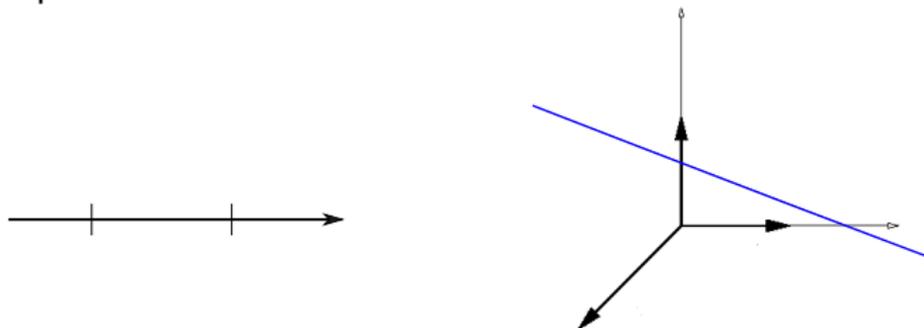
Outro ingrediente

Consideremos uma divisão de \mathbb{R} em intervalos I_0, I_1, \dots, I_N .

Lema

Existe uma aplicação afim $L : \mathbb{R} \rightarrow \mathbb{R}^N$ que envia os intervalos I_j nos cones maximais do fan de \mathbb{CP}^N .

Exemplo:



Retomando a demonstração

- Dividir \mathbb{R} em intervalos I_0, I_1, \dots, I_N de modo a que $I_j \cap K$ seja suficientemente pequeno, aproximar f por uma função em escada $g = \sum_{j=0}^N c_j \mathbb{1}_{I_j}$ definida nestes intervalos.

Retomando a demonstração

- Dividir \mathbb{R} em intervalos I_0, I_1, \dots, I_N de modo a que $I_j \cap K$ seja suficientemente pequeno, aproximar f por uma função em escada $g = \sum_{j=0}^N c_j \mathbb{1}_{I_j}$ definida nestes intervalos.
- Enviar os intervalos para o fan de $\mathbb{C}P^N$ ($L : \mathbb{R} \rightarrow \mathbb{R}^N$)

Retomando a demonstração

- Dividir \mathbb{R} em intervalos I_0, I_1, \dots, I_N de modo a que $I_j \cap K$ seja suficientemente pequeno, aproximar f por uma função em escada $g = \sum_{j=0}^N c_j \mathbb{1}_{I_j}$ definida nestes intervalos.
- Enviar os intervalos para o fan de $\mathbb{C}P^N$ ($L : \mathbb{R} \rightarrow \mathbb{R}^N$)
- Função em escada, $\sigma_\infty : \mathbb{R}^N \rightarrow \mathbb{R}^N$

Retomando a demonstração

- Dividir \mathbb{R} em intervalos I_0, I_1, \dots, I_N de modo a que $I_j \cap K$ seja suficientemente pequeno, aproximar f por uma função em escada $g = \sum_{j=0}^N c_j \mathbb{1}_{I_j}$ definida nestes intervalos.
- Enviar os intervalos para o fan de $\mathbb{C}P^N$ ($L : \mathbb{R} \rightarrow \mathbb{R}^N$)
- Função em escada, $\sigma_\infty : \mathbb{R}^N \rightarrow \mathbb{R}^N$
- Ajustar as constantes: $W : \mathbb{R}^N \rightarrow \mathbb{R}$

Retomando a demonstração

- Dividir \mathbb{R} em intervalos I_0, I_1, \dots, I_N de modo a que $I_j \cap K$ seja suficientemente pequeno, aproximar f por uma função em escada $g = \sum_{j=0}^N c_j \mathbb{1}_{I_j}$ definida nestes intervalos.
- Enviar os intervalos para o fan de $\mathbb{C}P^N$ ($L : \mathbb{R} \rightarrow \mathbb{R}^N$)
- Função em escada, $\sigma_\infty : \mathbb{R}^N \rightarrow \mathbb{R}^N$
- Ajustar as constantes: $W : \mathbb{R}^N \rightarrow \mathbb{R}$
- **Função aproximadora:** $W \circ \sigma \circ (tL)$, com t suficientemente grande.

Referências

- George Jeffreys e Siu-Cheong Lau, *Kähler Geometry of Quiver Varieties and Machine Learning* (2021), arXiv:2101.11487.