

Machine Learning Universal Approximation Theorem vs. toric Kähler geometry

João Camarneiro

Tutor: José Mourão

Oficina Diagonal – Programa *Novos Talentos em Matemática*

6 de março de 2021

Table of Contents

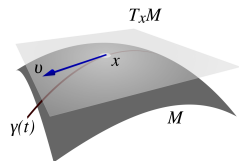
- 1 Algumas noções geométricas
- 2 Variedades tóricas
- 3 Redes neuronais artificiais

Variedades riemannianas

(M, g) , onde g é uma **métrica riemanniana** –
define de forma suave um produto interno g_p
em cada espaço tangente T_pM :

- $g_p : T_pM \times T_pM \rightarrow \mathbb{R}$ *bilinear*
- *Simétrico*: $g_p(u, v) = g_p(v, u)$
- *Não degenerado*: $g_p(u, -) \neq 0$

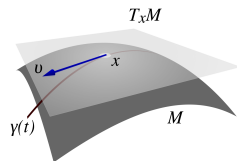
Permite calcular *comprimentos, distâncias, ângulos, áreas, volumes*.



Variedades simpléticas

(M, ω) , onde ω é uma **forma simplética** –
 define de forma suave uma estrutura
 simplética ω_p em cada espaço tangente T_pM :

- $\omega_p : T_pM \times T_pM \rightarrow \mathbb{R}$ *bilinear*
- **Antissimétrico:** $\omega_p(u, v) = -\omega_p(v, u)$
- *Não degenerado:* $\omega_p(u, -) \neq 0$
- *Fechada:* $d\omega = 0$



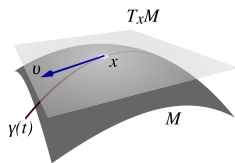
Exemplo típico: \mathbb{R}^{2n} (dimensão par), $\omega = \sum_{j=1}^n dx_j \wedge dx_{n+j}$

Variedades complexas

(M, J) , onde J é uma **estrutura quase complexa integrável** – define de forma suave uma estrutura complexa J_p em cada espaço tangente T_pM :

- $J_p : T_pM \rightarrow T_pM$ linear
- $J_p^2 = -\text{Id}$
- Satisfaz uma *condição de integrabilidade*

Exemplo típico: $\mathbb{R}^{2n} \sim \mathbb{C}^n$ (dimensão real par)



Variedades Kähler

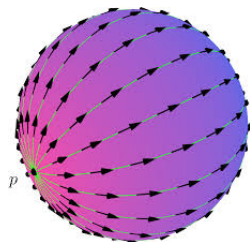
Uma **variedade Kähler** (M, ω, J, g) junta estas três estruturas geométricas de forma *compatível*: quaisquer duas delas determinam a terceira.

Exemplos: \mathbb{C}^n , $(\mathbb{C}\mathbb{P}^n, g_{FS})$

Campos vetoriais e fluxos

Um campo vetorial completo X em M define

$$\text{curvas integrais: } \begin{cases} c'(t) = X_{c(t)} \\ c(0) = p_0 \end{cases} .$$



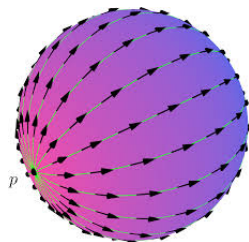
Campos vetoriais e fluxos

Um campo vetorial completo X em M define

$$\text{curvas integrais: } \begin{cases} c'(t) = X_{c(t)} \\ c(0) = p_0 \end{cases} .$$

Podemos definir o *fluxo* gerado por X :

$$\begin{aligned} \varphi : \mathbb{R} &\rightarrow \text{Diff}(M), & t &\mapsto \varphi_t \\ \varphi_t(p) &= c_p(t) \end{aligned}$$



Campos vetoriais e fluxos

Um campo vetorial completo X em M define

$$\text{curvas integrais: } \begin{cases} c'(t) = X_{c(t)} \\ c(0) = p_0 \end{cases} .$$

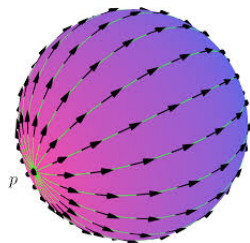
Podemos definir o *fluxo* gerado por X :

$$\begin{aligned} \varphi : \mathbb{R} &\rightarrow \text{Diff}(M), & t &\mapsto \varphi_t \\ \varphi_t(p) &= c_p(t) \end{aligned}$$

O fluxo é uma *ação* de $(\mathbb{R}, +)$ em M :

$$\varphi_t \circ \varphi_s = \varphi_{t+s}$$

Reciprocamente, dado qualquer fluxo existe um campo vetorial que o gera.



Campos vetoriais hamiltonianos

Considere-se uma variedade simplética (M, ω) e uma função $f \in C^\infty(M)$ (*função hamiltoniana*).

Usando a geometria simplética de M , podemos definir o *campo vetorial hamiltoniano* X_f que lhe corresponde:

$$\omega(X_f, -) = df$$

O fluxo de X_f preserva f e ω .

Ação de um toro

Seja $\mathbb{T}^n = (S^1)^n$ o toro de dimensão n . Uma ação de \mathbb{T}^n em M é um homomorfismo de grupos:

$$\begin{aligned}\varphi : \mathbb{T}^n &\rightarrow \text{Diff}(M) \\ \varphi_t \circ \varphi_s &= \varphi_{ts}\end{aligned}$$

Ação de um toro

Seja $\mathbb{T}^n = (S^1)^n$ o toro de dimensão n . Uma ação de \mathbb{T}^n em M é um homomorfismo de grupos:

$$\begin{aligned}\varphi : \mathbb{T}^n &\rightarrow \text{Diff}(M) \\ \varphi_t \circ \varphi_s &= \varphi_{ts}\end{aligned}$$

A uma ação de \mathbb{T}^n em M corresponde a respetiva ação *infinitesimal*:

$$\hat{\varphi} : \mathbb{R}^n \rightarrow \mathfrak{X}(M) = \{\text{campos vetoriais em } M\}$$

Ação de um toro

Seja $\mathbb{T}^n = (S^1)^n$ o toro de dimensão n . Uma ação de \mathbb{T}^n em M é um homomorfismo de grupos:

$$\begin{aligned}\varphi : \mathbb{T}^n &\rightarrow \text{Diff}(M) \\ \varphi_t \circ \varphi_s &= \varphi_{ts}\end{aligned}$$

A uma ação de \mathbb{T}^n em M corresponde a respetiva ação *infinitesimal*:

$$\hat{\varphi} : \mathbb{R}^n \rightarrow \mathfrak{X}(M) = \{\text{campos vetoriais em } M\}$$

Se estes *campos vetoriais fundamentais* forem todos hamiltonianos, temos uma aplicação:

$$\mathbb{R}^n \rightarrow C^\infty(M), \quad \text{i.e.,} \quad \mathbb{R}^n \rightarrow (M \rightarrow \mathbb{R})$$

Ação de um toro

Seja $\mathbb{T}^n = (S^1)^n$ o toro de dimensão n . Uma *ação* de \mathbb{T}^n em M é um homomorfismo de grupos:

$$\begin{aligned}\varphi : \mathbb{T}^n &\rightarrow \text{Diff}(M) \\ \varphi_t \circ \varphi_s &= \varphi_{ts}\end{aligned}$$

A uma ação de \mathbb{T}^n em M corresponde a respetiva *ação infinitesimal*:

$$\hat{\varphi} : \mathbb{R}^n \rightarrow \mathfrak{X}(M) = \{\text{campos vetoriais em } M\}$$

Se estes *campos vetoriais fundamentais* forem todos hamiltonianos, temos uma aplicação:

$$\mathbb{R}^n \rightarrow C^\infty(M), \quad \text{i.e.,} \quad \mathbb{R}^n \rightarrow (M \rightarrow \mathbb{R})$$

Trocando a ordem dos argumentos, definimos a *aplicação momento* $\mu : M \rightarrow (\mathbb{R}^n)^* \cong \mathbb{R}^n$.

Variedades tóricas

Neste caso, se μ for invariante pela ação, diz-se que a ação é *hamiltoniana*.

Uma **variedade tórica (simplética)** é uma variedade simplética (M, ω) compacta conexa de dimensão $2n$, equipada com uma ação hamiltoniana efetiva de \mathbb{T}^n e aplicação momento $\mu : M \rightarrow \mathbb{R}^n$.

Juntando uma estrutura complexa compatível e invariante pela ação do toro, temos uma **variedade tórica Kähler**.

Exemplo: $\mathbb{C}P^n$ (dimensão real $2n$, dimensão complexa n)

Exemplo: $\mathbb{C}P^n$ (dimensão real $2n$, dimensão complexa n)

$$(t_1, \dots, t_n) \cdot [Z_0 : Z_1 : \dots : Z_n] = [Z_0 : t_1 Z_1 : \dots : t_n Z_n]$$

Exemplo: $\mathbb{C}\mathbb{P}^n$ (dimensão real $2n$, dimensão complexa n)

$$(t_1, \dots, t_n) \cdot [Z_0 : Z_1 : \dots : Z_n] = [Z_0 : t_1 Z_1 : \dots : t_n Z_n]$$

A aplicação momento neste caso é $\mu([Z_0 : Z_1 : \dots : Z_n]) =$

$$= -\frac{1}{2} \left(\frac{|Z_1|^2}{|Z_0|^2 + |Z_1|^2 + \dots + |Z_n|^2}, \dots, \frac{|Z_n|^2}{|Z_0|^2 + |Z_1|^2 + \dots + |Z_n|^2} \right)$$

Aplicação momento e convexidade

Teorema de Convexidade (Atiyah-Guillemin-Sternberg-Delzant)

Seja (M, ω, μ) uma variedade tórica. Então, $\mu(M) \subset \mathbb{R}^n$ é um polítopo de Delzant (*polítopo momento*).

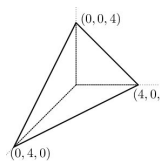
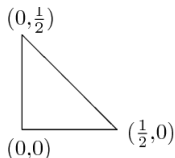
Aplicação momento e convexidade

Teorema de Convexidade (Atiyah-Guillemin-Sternberg-Delzant)

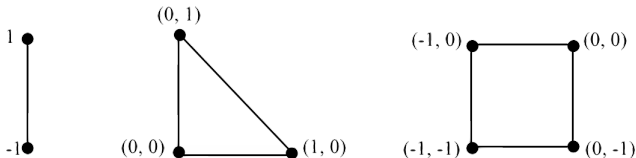
Seja (M, ω, μ) uma variedade tórica. Então, $\mu(M) \subset \mathbb{R}^n$ é um polígono de Delzant (*polígono momento*).

Um polígono convexo $P \subset \mathbb{R}^n$ diz-se *Delzant* se:

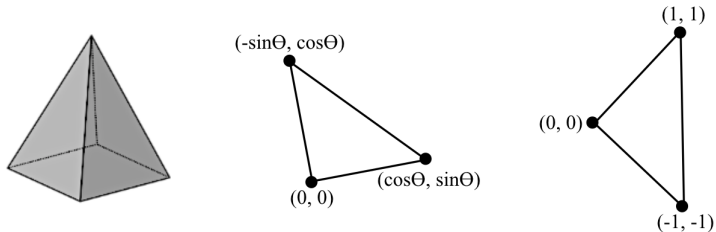
- em cada vértice incidem n arestas;
- os vetores diretores das arestas podem ser tomados com entradas inteiras;
- em cada vértice, os vetores diretores das arestas incidentes podem ser escolhidos de forma a formarem uma base de \mathbb{Z}^n .



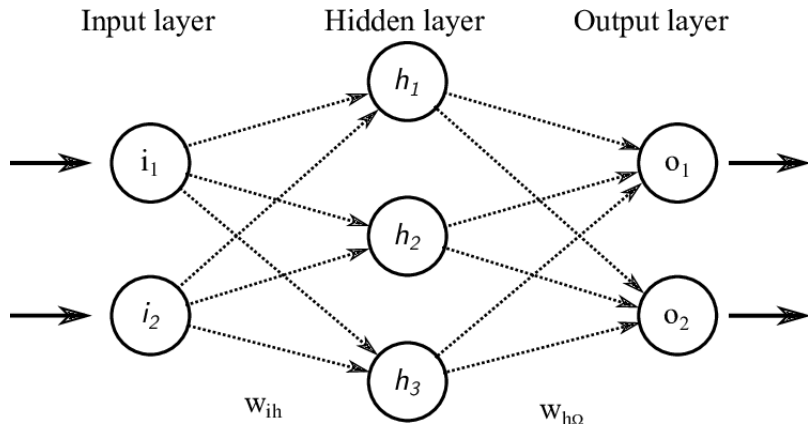
Exemplos de polítopos convexos que são Delzant:

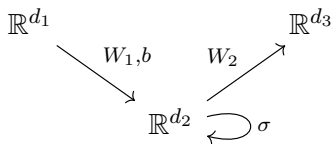


Exemplos de polítopos convexos que não são Delzant:



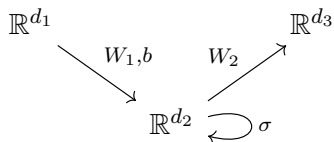
Redes neurais artificiais





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

onde σ é uma *função de ativação* não linear.



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

onde σ é uma *função de ativação* não linear.

O exemplo que nos vai interessar é

$$\sigma(x_1, \dots, x_n) = \left(\frac{e^{2x_1}}{1 + e^{2x_1} + \dots + e^{2x_n}}, \dots, \frac{e^{2x_n}}{1 + e^{2x_1} + \dots + e^{2x_n}} \right)$$

Onde entra a geometria tórica

Recordemos a aplicação momento da variedade tórica Kähler $\mathbb{C}\mathbb{P}^n$:

$$\left(\frac{|Z_1|^2}{|Z_0|^2 + |Z_1|^2 + \dots + |Z_n|^2}, \dots, \frac{|Z_n|^2}{|Z_0|^2 + |Z_1|^2 + \dots + |Z_n|^2} \right)$$

Onde entra a geometria tórica

Recordemos a aplicação momento da variedade tórica Kähler $\mathbb{C}\mathbb{P}^n$:

$$\left(\frac{|Z_1|^2}{|Z_0|^2 + |Z_1|^2 + \dots + |Z_n|^2}, \dots, \frac{|Z_n|^2}{|Z_0|^2 + |Z_1|^2 + \dots + |Z_n|^2} \right)$$

Dividindo por Z_0 , i.e., $z_j = \frac{Z_j}{Z_0} = e^{x_j + i\theta_j}$:

$$[Z_0 : Z_1 : \dots : Z_n] = [1 : z_1 : \dots : z_n] = [1 : e^{x_1 + i\theta_1} : \dots : e^{x_n + i\theta_n}]$$

Onde entra a geometria tórica

Recordemos a aplicação momento da variedade tórica Kähler $\mathbb{C}\mathbb{P}^n$:

$$\left(\frac{|Z_1|^2}{|Z_0|^2 + |Z_1|^2 + \dots + |Z_n|^2}, \dots, \frac{|Z_n|^2}{|Z_0|^2 + |Z_1|^2 + \dots + |Z_n|^2} \right)$$

Dividindo por Z_0 , i.e., $z_j = \frac{Z_j}{Z_0} = e^{x_j + i\theta_j}$:

$$[Z_0 : Z_1 : \dots : Z_n] = [1 : z_1 : \dots : z_n] = [1 : e^{x_1 + i\theta_1} : \dots : e^{x_n + i\theta_n}]$$

a aplicação momento toma a forma

$$\left(\frac{e^{2x_1}}{1 + e^{2x_1} + \dots + e^{2x_n}}, \dots, \frac{e^{2x_n}}{1 + e^{2x_1} + \dots + e^{2x_n}} \right) = \sigma(x)$$

Teorema de Aproximação Universal (com geometria tórica)

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

O facto de a função de ativação ser a aplicação momento de uma variedade tórica permite recorrer à geometria tórica para demonstrar este teorema (em particular, usando técnicas de geometria tropical).