Sobre o Teorema da Aproximação Universal

Caio Lins

Orientador: Yuri Saporito

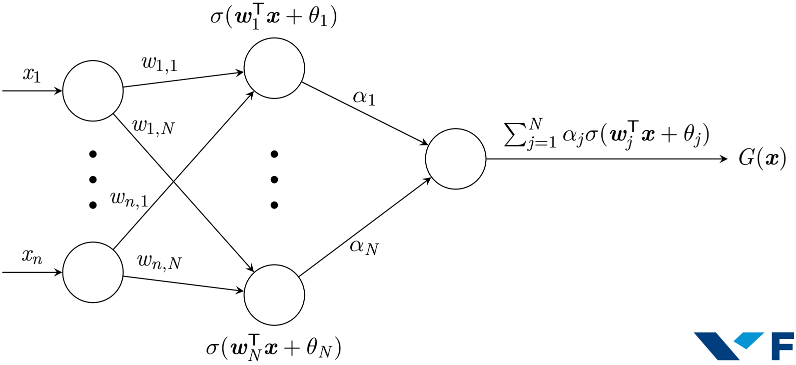
FGV - EMAp

Outubro 2021

Overview de Redes Neurais

O que são redes neurais?

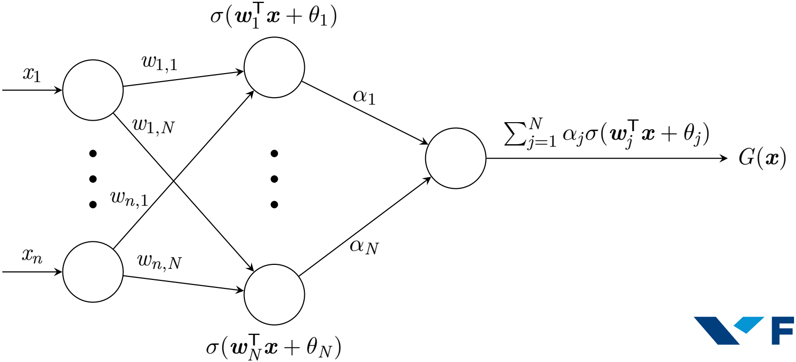
- Redes neurais são funções que espelham o funcionamento do cérebro humano. É organizada em camadas de "neurônios".



Overview de Redes Neurais

O que são redes neurais?

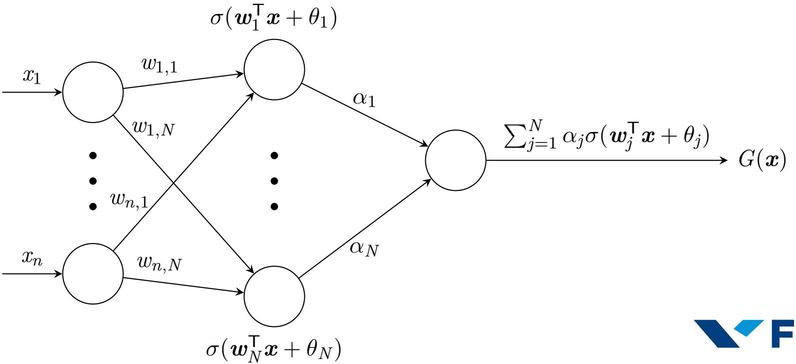
- Input: $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$.
- Pesos (*weights*): $\boldsymbol{w}_j = (w_{1,j}, \ldots, w_{N,j}) \in \mathbb{R}^n, j = 1, \ldots, N$.
- Viéses (*biases*): $\theta_j \in \mathbb{R}, j = 1, \ldots, N$.
- Função de ativação: $\sigma : \mathbb{R} \to \mathbb{R}$.



O j-ésimo neurônio da camada oculta computa a soma

$$\boldsymbol{w}_j^\top \boldsymbol{x} + \theta_j = w_{1,j}x_1 + w_{2,j}x_2 + \cdots + w_{n,j}x_n + \theta_j$$

e passa o resultado pela ativação $\sigma : \mathbb{R} \to \mathbb{R}$, que é *não linear*.

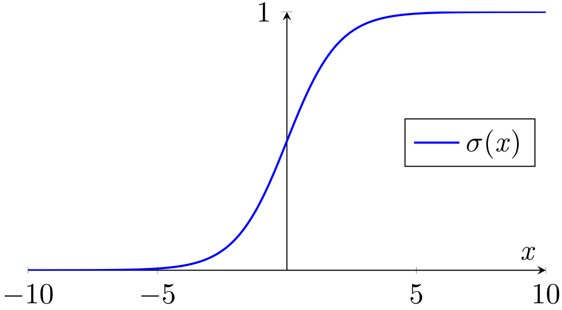


Overview de Redes Neurais

O que são redes neurais?

Existem várias funções de ativação possíveis, mas tenham em mente a *função logística*:

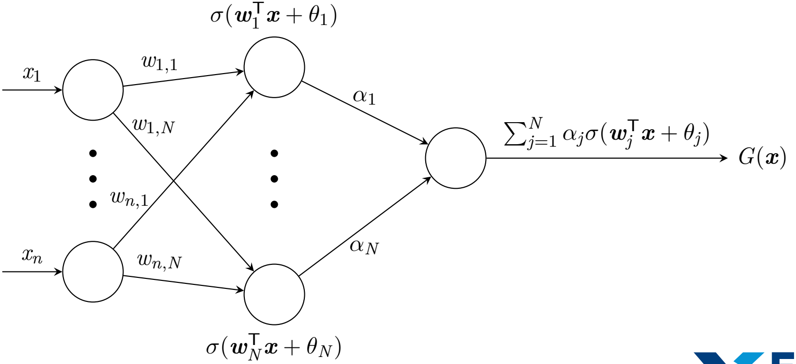
$$\sigma(x) = \frac{1}{1 + e^{-x}}.$$



Overview de Redes Neurais

O que são redes neurais?

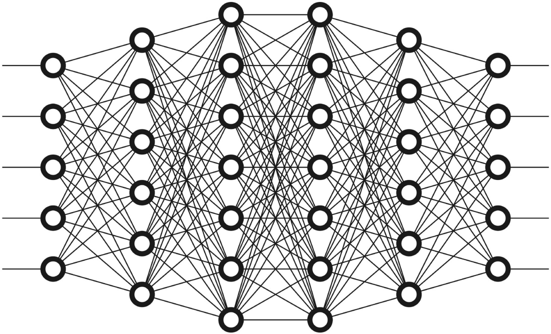
Neste caso, o output da rede é um escalar, dado pela soma: $G(\boldsymbol{x}) = \sum_{j=1}^{N} \alpha_j \sigma(\boldsymbol{w}_j^\mathsf{T} \boldsymbol{x} + \theta_j)$, mas isso não é regra.



Overview de Redes Neurais

O que são redes neurais?

A rede mostrada até o momento tem apenas uma camada oculta, mas redes neurais em geral têm várias.



Diversas aplicações:

- Classificação de imagens, incluindo reconhecimento facial;

Diversas aplicações:

- Classificação de imagens, incluindo reconhecimento facial;
- Predição de séries temporais;

Overview de Redes Neurais

Para que servem?

Diversas aplicações:

- Classificação de imagens, incluindo reconhecimento facial;
- Predição de séries temporais;
- Análise de regressão;

Overview de Redes Neurais

Para que servem?

Diversas aplicações:

- Classificação de imagens, incluindo reconhecimento facial;
- Predição de séries temporais;
- Análise de regressão;
- O algoritmo usado pela inteligência artificial *AlphaGo* emprega redes neurais convolucionais (CNNs).

Diversas aplicações:

- Classificação de imagens, incluindo reconhecimento facial;
- Predição de séries temporais;
- Análise de regressão;
- O algoritmo usado pela inteligência artificial *AlphaGo* emprega redes neurais convolucionais (CNNs).
- Deep Galerkin Method (DGM): Algoritmo de Deep Learning para resolver equações diferenciais parciais.

Redes Neurais Funcionam?

- Podemos nos questionar se Redes Neurais são capazes de realizar essas tarefas arbitrariamente bem (no fundo, tudo se resume a aproximar funções).

Redes Neurais Funcionam?

- Podemos nos questionar se Redes Neurais são capazes de realizar essas tarefas arbitrariamente bem (no fundo, tudo se resume a aproximar funções).
- A resposta é sim.

Redes Neurais Funcionam?

Teorema da Aproximação Universal

Seja $\sigma : \mathbb{R} \to \mathbb{R}$ *uma função contínua e discriminatória qualquer. Então, dada uma função* $f : [0,1]^n \to \mathbb{R}$ *contínua, e um* $\varepsilon > 0$, *existe uma função* $G : [0,1]^n \to \mathbb{R}$, *da seguinte forma:*

$$G(\boldsymbol{x}) = \sum_{j=1}^{N} \alpha_j \sigma(\boldsymbol{w}_j^\mathsf{T} \boldsymbol{x} + \theta_j),$$

onde $\alpha_j, \theta_j \in \mathbb{R}$ *e* $\boldsymbol{w}_j \in \mathbb{R}^n$, $j = 1, \ldots, N$, *tal que*

$$|G(\boldsymbol{x}) - f(\boldsymbol{x})| < \varepsilon$$

para todo $\boldsymbol{x} \in [0,1]^n$.

Redes Neurais Funcionam?

- Mais à frente especificaremos o que significa "discriminatória".

Redes Neurais Funcionam?

- Mais à frente especificaremos o que significa "discriminatória".
- Ou seja, redes neurais com uma camada oculta são aproximadores universais de funções reais contínuas definidas em $[0,1]^n =: \mathbb{I}^n$.

Redes Neurais Funcionam?

- Artigo onde esse resultado foi provado: [Cyb89]

Theorem 1. *Let σ be any continuous discriminatory function. Then finite sums of the form*

$$G(x) = \sum_{j=1}^{N} \alpha_j \sigma(y_j^{\mathrm{T}} x + \theta_j) \tag{2}$$

are dense in $C(I_n)$. In other words, given any $f \in C(I_n)$ and $\varepsilon > 0$, there is a sum, $G(x)$, of the above form, for which

$$|G(x) - f(x)| < \varepsilon \qquad \textit{for all} \quad x \in I_n.$$

Proof. Let $S \subset C(I_n)$ be the set of functions of the form $G(x)$ as in (2). Clearly S is a linear subspace of $C(I_n)$. We claim that the closure of S is all of $C(I_n)$.

Assume that the closure of S is not all of $C(I_n)$. Then the closure of S, say R, is a closed proper subspace of $C(I_n)$. By the Hahn–Banach theorem, there is a bounded linear functional on $C(I_n)$, call it L, with the property that $L \neq 0$ but $L(R) = L(S) = 0$.

By the Riesz Representation Theorem, this bounded linear functional, L, is of the form

$$L(h) = \int_{I_n} h(x)\, d\mu(x)$$

for some $\mu \in M(I_n)$, for all $h \in C(I_n)$. In particular, since $\sigma(y^{\mathrm{T}} x + \theta)$ is in R for all y and θ, we must have that

$$\int_{I_n} \sigma(y^{\mathrm{T}} x + \theta)\, d\mu(x) = 0$$

for all y and θ.

However, we assumed that σ was discriminatory so that this condition implies that $\mu = 0$ contradicting our assumption. Hence, the subspace S must be dense in $C(I_n)$. ∎

- Demonstração utiliza resultados de Análise Funcional e Teoria da Medida.

Interlúdio

Outros resultados relacionados

- Demonstração utiliza resultados de Análise Funcional e Teoria da Medida.
- Estudar aproximações em outros contextos

- Demonstração utiliza resultados de Análise Funcional e Teoria da Medida.
- Estudar aproximações em outros contextos
- Teorema da Aproximação de Weierstrass:

- Demonstração utiliza resultados de Análise Funcional e Teoria da Medida.
- Estudar aproximações em outros contextos
- Teorema da Aproximação de Weierstrass:
    - *Aproximar* funções contínuas de um intervalo na reta por meio de *polinômios*.

Interlúdio

Outros resultados relacionados

- Demonstração utiliza resultados de Análise Funcional e Teoria da Medida.
- Estudar aproximações em outros contextos
- Teorema da Aproximação de Weierstrass:
  - *Aproximar* funções contínuas de um intervalo na reta por meio de *polinômios*.
- Um desdobramento do 13° problema de Hilbert:

- Demonstração utiliza resultados de Análise Funcional e Teoria da Medida.
- Estudar aproximações em outros contextos
- Teorema da Aproximação de Weierstrass:
  - *Aproximar* funções contínuas de um intervalo na reta por meio de *polinômios*.
- Um desdobramento do 13° problema de Hilbert:
  - Representar *de maneira exata* funções contínuas de $\mathbb{I}^n$ em $\mathbb{R}$ por meio de composições e adições de *funções contínuas de* $\mathbb{R}$ *em* $\mathbb{R}$.

Um pouco de notação:

- Dado $X \subseteq \mathbb{R}^n$, definimos

$$C(X) = \{\varphi : X \to \mathbb{R} \mid \varphi \text{ é contínua } \}.$$

- $C(X)$ é um espaço vetorial no qual introduzimos a seguinte norma:

$$\|\varphi\|_\infty := \sup\{|\varphi(x)| : x \in X\}.$$

Teorema (Aproximação de Weierstrass)

Dada $f \in C([a, b])$, *para todo* $\varepsilon > 0$ *existe um polinômio* $p : [a, b] \to \mathbb{R}$ *tal que*

$$\|f - p\|_\infty < \varepsilon.$$

Ideia da demonstração [You06]:

- Reduzir ao caso em que $[a, b] = [0, 1]$.

Ideia da demonstração [You06]:

- Reduzir ao caso em que $[a, b] = [0, 1]$.
- Aproximar f pelos seus *Polinômios de Bernstein*:

$$B_n(x, f) := \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^k (1 - x)^{n-k}.$$

Ideia da demonstração [You06]:

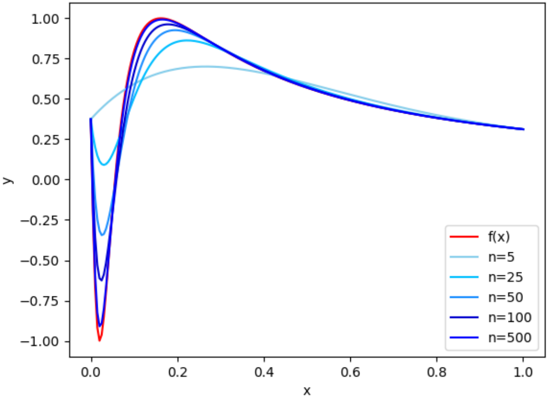
- Reduzir ao caso em que $[a, b] = [0, 1]$.
- Aproximar f pelos seus *Polinômios de Bernstein*:

$$B_n(x, f) := \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}.$$

- De fato, temos

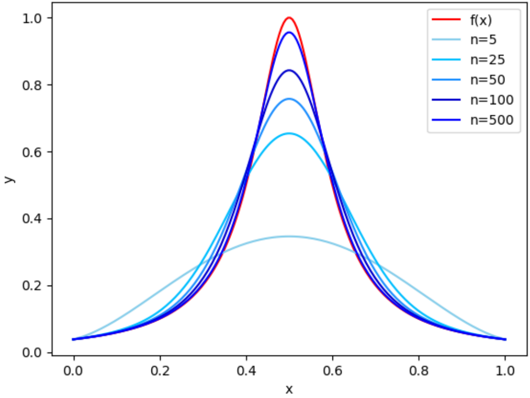
$$\lim_{n\to\infty} \|B_n(\cdot, f) - f\|_\infty = 0.$$

Ilustrando o Teorema:



$$f(x) = \text{sen}\left(\frac{1}{3(x+0.05)}\right)$$

Ilustrando o Teorema:

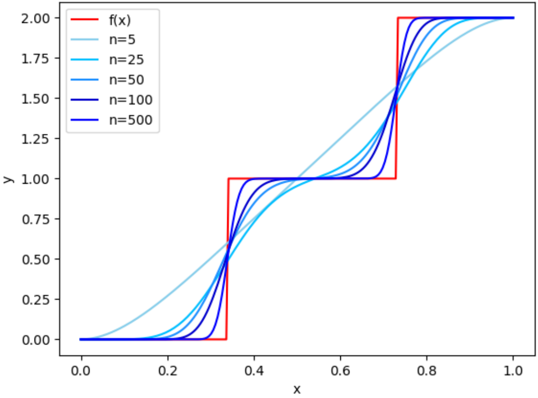


$$f(x) = \frac{1}{100(x-0.5)^2+1}$$

Interlúdio

Teorema da Aproximação de Weierstrass

Um exemplo com f descontínua:



$$f(x) = |\lfloor 3\,\text{sen}(x) \rfloor|$$

Interlúdio

13° problema de Hilbert

- Com a linguagem da época, Hilbert conjecturou que existem funções contínuas $f : \mathbb{I}^3 \to \mathbb{R}$ que não podem ser representadas como composições e somas de funções contínuas de $\mathbb{R} \to \mathbb{R}$ [Mor21].
- Exemplo em $\mathbb{I}^2$: $f(x, y) = xy$. Observe que

$$
\begin{aligned}
f(x,y) = xy &= (x+1)(y+1) - \left(x + \frac{1}{2}\right) - \left(y + \frac{1}{2}\right) \\
&= e^{[\log(x+1)+\log(y+1)]} - \left(x + \frac{1}{2}\right) - \left(y + \frac{1}{2}\right).
\end{aligned}
$$

(Somamos 1 para evitar $\log 0$).

- Essa conjectura era **falsa**.

Interlúdio

13° problema de Hilbert

- Essa conjectura era **falsa**.
  - Provado por Vladimir Igorevich Arnol'd em 1957, baseado no trabalho de seu orientador, Andrej Nikolajewitseh Kolmogorov.

- Essa conjectura era **falsa**.
  - Provado por Vladimir Igorevich Arnol'd em 1957, baseado no trabalho de seu orientador, Andrej Nikolajewitseh Kolmogorov.
- Teorema da Superposição de Kolmogorov.

Interlúdio

13° problema de Hilbert

- Essa conjectura era **falsa**.
  - Provado por Vladimir Igorevich Arnol'd em 1957, baseado no trabalho de seu orientador, Andrej Nikolajewitseh Kolmogorov.
- Teorema da Superposição de Kolmogorov.
- Foi generalizado posteriormente.

Teorema (Kolmogorov, Arnol'd, Kahane, Lorentz e Sprechner)

Para todo $n \in \mathbb{N}$ *com* $n \geq 2$, *existem números reais* $\lambda_1, \lambda_2, \ldots, \lambda_n$ *e funções contínuas* $\varphi_k : \mathbb{I} \to \mathbb{R}$, *para* $k = 1, \ldots, 2n+1$, *com a propriedade de que para toda função contínua* $f : \mathbb{I}^n \to \mathbb{R}$ *existe uma função contínua* $g : \mathbb{R} \to \mathbb{R}$ *tal que, para todo* $(x_1, \ldots, x_n) \in \mathbb{I}^n$,

$$f(x_1, \ldots, x_n) = \sum_{k=1}^{2n+1} g(\lambda_1 \varphi_k(x_1) + \cdots + \lambda_n \varphi_k(x_n)).$$

FGV

Diferenças com o Teorema da Aproximação Universal:

- A representação é *exata*, não aproximada.

Diferenças com o Teorema da Aproximação Universal:

- A representação é *exata*, não aproximada.
- As funções g permitem utilizar uma gama de não-linearidades muito maior do que apenas a função de ativação σ.

Voltando

Voltando ao Teorema da Aproximação Universal, sua demonstração pode ser vista como uma simples aplicação dos seguintes Teoremas:

Voltando

Voltando ao Teorema da Aproximação Universal, sua demonstração pode ser vista como uma simples aplicação dos seguintes Teoremas:

- Teorema de Hahn-Banach [Oli12].

Voltando

Voltando ao Teorema da Aproximação Universal, sua demonstração pode ser vista como uma simples aplicação dos seguintes Teoremas:

- Teorema de Hahn-Banach [Oli12].
- Teorema da Representação de Riesz-Markov [Roy88].

Voltando

Voltando ao Teorema da Aproximação Universal, sua demonstração pode ser vista como uma simples aplicação dos seguintes Teoremas:

- Teorema de Hahn-Banach [Oli12].
- Teorema da Representação de Riesz-Markov [Roy88].

Há, também, o lema que relaciona funções discriminatórias com as funções de ativação de redes neurais.

O Teorema de Hahn-Banach

Definição

Dado um espaço vetorial real V, uma função $p : V \to \mathbb{R}$ é chamada de *funcional sublinear* se para todos $\boldsymbol{x}, \boldsymbol{y} \in V$ e $\lambda \in \mathbb{R}$ temos

$$p(\boldsymbol{x} + \boldsymbol{y}) \leq p(\boldsymbol{x}) + p(\boldsymbol{y})$$
$$p(\lambda \boldsymbol{x}) = \lambda p(\boldsymbol{x}),$$

e, além disso, $p(\mathbf{0}) = 0$.

O Teorema de Hahn-Banach

Teorema

Sejam V um espaço vetorial real, M um subespaço de V e p um funcional sublinear em V. Se f é um funcional linear em M dominado por p, ou seja, tal que $f(\boldsymbol{v}) \leq p(\boldsymbol{v})$ para todo $\boldsymbol{v} \in M$, então existe um funcional linear F em V, que coincide com f em M e que também é dominado por p. O funcional F é a extensão de Hahn-Banach de f.

O Teorema de Hahn-banach

O Teorema de Hahn-banach

Sua demonstração é elementar, constituindo de basicamente uma aplicação do Lema de Zorn:

Axioma (Lema de Zorn)

Todo conjunto parcialmente ordenado, tal que todos seus subconjuntos totalmente ordenados possuem cota superior, possui elemento maximal.

O Teorema de Hahn-Banach

Na verdade, o que realmente utilizaremos é um corolário desse Teorema:

Definição

Dado um espaço vetorial V, definimos

$$V^* = \{f : V \to \mathbb{R} \mid f \text{ é linear limitada}\}.$$

Corolário

Um subespaço M de um espaço vetorial V é denso se, e somente se, o único elemento de V^* que se anula em M é o funcional nulo.

O Teorema da Representação de Riesz

- Um nome para vários teoremas, mas todos se preocupam com representar funcionais lineares utilizando alguma estrutura do espaço onde eles estão definidos.

O Teorema da Representação de Riesz

Em Análise Funcional

Geralmente, no contexto de Análise Funcional (mais especificamente, na teoria de espaços de Hilbert), esse teorema trata da associação natural entre um espaço de Hilbert $\mathbb{H}$ e seu dual $\mathbb{H}^*$:

Teorema (Representação de Riesz - Análise Funcional)

Dado um espaço de Hilbert $\mathbb{H}$ e seu dual $\mathbb{H}^*$, a função

$$\begin{aligned} \gamma : & \mathbb{H} \to \mathbb{H}^* \\ & \boldsymbol{v} \mapsto \gamma(\boldsymbol{v}) = f_{\boldsymbol{v}}, \end{aligned}$$

tal que $f_{\boldsymbol{v}} = \langle \boldsymbol{v}, \cdot \rangle$, é uma isometria antilinear e sobrejetiva em $\mathbb{H}^*$.

Já no contexto de Teoria da Medida, há mais de uma possibilidade. Em geral, ele se refere à dualidade entre os espaços L_p e L_q:

Teorema (Representação de Riesz - Teoria da Medida)

Se $(X, \boldsymbol{X}, \mu)$ *é um espaço de medida arbitrário e* G *é um funcional linear limitado em* $L_p(X, \boldsymbol{X}, \mu)$, $1 < p < \infty$, *então existe uma* g *em* $L_q(X, \boldsymbol{X}, \mu)$, *onde* $1/p + 1/q = 1$, *tal que*

$$G(f) = \int fg \, \mathrm{d}\mu$$

para toda $f \in L_p$.

Se μ é σ-finita, o resultado vale para L_1, cujo dual é L_∞.

O Teorema da Representação de Riesz

Em Teoria da Medida

Entretanto, também pode ser referir à representação de funcionais lineares por meio de medidas. Nesse sentido, também é chamado de Teorema de Riesz-Markov.

A versão exata que usaremos posteriormente é esta:

Teorema (Representação de Riesz-Markov)

Seja X um espaço compacto Hausdorff e $C(X)$ o espaço das funções reais contínuas em X. Então, a cada funcional linear limitado $F : C(X) \to \mathbb{R}$ corresponde uma única medida de Baire, finita, com sinal, ν em X tal que

$$F(f) = \int f \, \mathrm{d}\nu$$

para cada f em $C(X)$.

Na prática, como $\mathbb{I}^n$ é um espaço métrico compacto, a medida ν será uma medida de Borel regular.

O Teorema da Representação de Riesz

Em Teoria da Medida

Esta é uma versão mais simples desse Teorema, cuja demonstração foi estudada em maior detalhe:

Teorema (Representão de Riesz)

Seja $J = [a, b]$. Se G é um funcional linear limitado no espaço $C(J)$ das funções reais contínuas definidas em J, então existe uma medida γ, definida nos subconjuntos de Borel de $\mathbb{R}$, tal que

$$G(f) = \int_J f \, \mathrm{d}\gamma$$

para toda $f \in C(J)$.

Tendo os Teoremas de Hahn-Banach e de Riesz-Markov em mãos, vamos provar o Teorema da Aproximação Universal. Mas, antes, devemos dizer o que significa σ ser discriminatória:

Definição

Seja $M(\mathbb{I}^n)$ o espaço das medidas de Borel finitas, com sinal e regulares em $\mathbb{I}^n$. Dizemos que σ é *discriminatória* se a única medida $\mu \in M(\mathbb{I}^n)$ que satisfaz

$$\int_{\mathbb{I}^n} \sigma(\boldsymbol{y}^\mathsf{T}\boldsymbol{x} + \theta)\, \mathrm{d}\mu(\boldsymbol{x}) = 0$$

para todos $\boldsymbol{y} \in \mathbb{R}^n$ e $\theta \in \mathbb{R}$ é a medida nula.

Relembramos o enunciado do Teorema:

Teorema da Aproximação Universal

Seja σ uma função contínua discriminatória qualquer. Então as somas finitas da forma

$$G(\boldsymbol{x}) = \sum_{j=1}^{N} \alpha_j \sigma(\boldsymbol{w}_j^\mathsf{T} \boldsymbol{x} + \theta_j).$$

são densas em $C(\mathbb{I}^n)$. Em outras palavras, dada qualquer $f \in C(\mathbb{I}^n)$ e $\varepsilon > 0$, existe uma soma, $G(\boldsymbol{x})$, com a forma acima, tal que

$$\|G - f\|_\infty < \varepsilon.$$

O Teorema da Aproximação Universal

Demonstração

- Seja $S \subseteq C(\mathbb{I}^n)$ o subespaço formado pelas somas G do enunciado do Teorema.

O Teorema da Aproximação Universal

Demonstração

- Seja $S \subseteq C(\mathbb{I}^n)$ o subespaço formado pelas somas G do enunciado do Teorema.
- Suponha que o Teorema seja **falso**, ou, seja, que S não é denso em $C(\mathbb{I}^n)$.

O Teorema da Aproximação Universal
Demonstração

- Seja $S \subseteq C(\mathbb{I}^n)$ o subespaço formado pelas somas G do enunciado do Teorema.
- Suponha que o Teorema seja **falso**, ou, seja, que S não é denso em $C(\mathbb{I}^n)$.
- Por Hahn-Banach (corolário) existe $L : C(\mathbb{I}^n) \to \mathbb{R}$ linear e limitado tal que L **é não nulo**, porém $L(S) = 0$.

O Teorema da Aproximação Universal
Demonstração

- Seja $S \subseteq C(\mathbb{I}^n)$ o subespaço formado pelas somas G do enunciado do Teorema.
- Suponha que o Teorema seja **falso**, ou, seja, que S não é denso em $C(\mathbb{I}^n)$.
- Por Hahn-Banach (corolário) existe $L : C(\mathbb{I}^n) \to \mathbb{R}$ linear e limitado tal que L **é não nulo**, porém $L(S) = 0$.
- Por Riesz-Markov, existe $\mu \in M(\mathbb{I}^n)$ tal que

$$L(f) = \int_{\mathbb{I}^n} f(\boldsymbol{x}) \, \mathrm{d}\mu(\boldsymbol{x})$$

para toda $f \in C(\mathbb{I}^n)$.

O Teorema da Aproximação Universal

Demonstração

- Entretanto, perceba que para todo $\boldsymbol{y} \in \mathbb{R}^n$ e $\theta \in \mathbb{R}$, a função $G_{\boldsymbol{y},\theta} : \mathbb{I}^n \to \mathbb{R}$ dada por

$$G_{\boldsymbol{y},\theta}(\boldsymbol{x}) = \sigma(\boldsymbol{y}^\mathsf{T}\boldsymbol{x} + \theta)$$

pertence a S.

- Entretanto, perceba que para todo $\boldsymbol{y} \in \mathbb{R}^n$ e $\theta \in \mathbb{R}$, a função $G_{\boldsymbol{y},\theta} : \mathbb{I}^n \to \mathbb{R}$ dada por

$$G_{\boldsymbol{y},\theta}(\boldsymbol{x}) = \sigma(\boldsymbol{y}^\mathsf{T}\boldsymbol{x} + \theta)$$

pertence a S.

- Logo, para todo $\boldsymbol{y} \in \mathbb{R}^n$ e $\theta \in \mathbb{R}$ temos

$$\begin{aligned} 0 = F(G_{\boldsymbol{y},\theta}) &= \int_{\mathbb{I}^n} G_{\boldsymbol{y},\theta}(\boldsymbol{x}) \, \mathrm{d}\mu(\boldsymbol{x}) \\ &= \int_{\mathbb{I}^n} \sigma(\boldsymbol{y}^\mathsf{T}\boldsymbol{x} + \theta) \, \mathrm{d}\mu(\boldsymbol{x}). \end{aligned}$$

- Porém, supomos que σ é discriminatória, ou seja, isso implica $\mu = 0$.

O Teorema da Aproximação Universal

- Porém, supomos que σ é discriminatória, ou seja, isso implica $\mu = 0$.
- Com isso, F é o funcional nulo, e obtemos uma contradição.

- Porém, supomos que σ é discriminatória, ou seja, isso implica $\mu = 0$.
- Com isso, F é o funcional nulo, e obtemos uma contradição.
- Portanto, S é denso em $C(\mathbb{I}^n)$, e acabamos. $\square$

Limitações do Teorema

- Não é especificado quantos neurônios são necessários para aproximar uma dada função

Limitações do Teorema

- Não é especificado quantos neurônios são necessários para aproximar uma dada função
- Nem um *algoritmo* para encontrar a aproximação

Limitações do Teorema

- Não é especificado quantos neurônios são necessários para aproximar uma dada função
- Nem um *algoritmo* para encontrar a aproximação
  - Atualmente utiliza-se variações de *Stochastic Gradient Descent*, junto com *Backpropagation* para o cálculo dos gradientes.

Possíveis Generalizações

- Em anos posteriores, foram publicados reultados que generalizam o Teorema apresentado.

Possíveis Generalizações

- Em anos posteriores, foram publicados reultados que generalizam o Teorema apresentado.
- Em particular, há resultados que garantem a proximidade das derivadas da rede com as da função objetivo.

Possíveis Generalizações

- Em anos posteriores, foram publicados reultados que generalizam o Teorema apresentado.
- Em particular, há resultados que garantem a proximidade das derivadas da rede com as da função objetivo.
- Também há estudos sobre outras funções de ativação, em particular, a ReLU:

$$\text{ReLU}(x) = \max\{x, 0\},$$

que também dá origem a uma família de redes com a propriedade de serem aproximadores universais.

FGV

Próximos passos

Estudo de Redes Neurais como ferramentas para obter soluções de Equações Diferenciais Parciais.

Próximos passos

Estudo de Redes Neurais como ferramentas para obter soluções de Equações Diferenciais Parciais.

- DGM: Deep Galerkin Method.

Agradecimentos

- Orientador: Yuri Saporito

Agradecimentos

- Orientador: Yuri Saporito
- CNPq: bolsa PICME

Repositório

Link para o reposiório no GitHub com o relatório e esta apresentação.

Referências I

[Roy88] H. L. Royden. *Real Analysis*. Ed. por Inc. Collier Macmillian Canada. Macmillian Publishing Company, 1988.

[Cyb89] George Cybenko. «Approximation by superpositions of a sigmoidal function». Em: *Math. Control. Signals Syst.* 2 (1989), pp. 303–314. DOI: 10.1007/BF02551274. URL: https://doi.org/10.1007/BF02551274.

[You06] Matt Young. *The Stone-Weierstrass theorem*. Jan. de 2006. URL: https://mast.queensu.ca/~speicher/Section14.pdf.

[Oli12] César R. de Oliveira. *Introdução à análise funcional*. Ed. por IMPA. IMPA, 2012.

Referências II

[Mor21] Sidney A. Morris. «Hilbert 13: are there any genuine continuous multivariate real-valued functions?» Em: *Bulletin (New Series) of the American Mathematical Society* 58.1 (jan. de 2021), pp. 107–118. URL: https://doi.org/10.1090/bull/1698.