





Reconhecimento de padrões e aprendizagem computacional

Redes neurais artificiais



Redes neurais

Definição

Técnica inspirada no funcionamento do cérebro, onde neurônios artificiais, conectados em rede, conseguem aprender e de generalizar.



Generalização

Isso significa que se a rede aprende a lidar com um certo problema, e lhe é apresentado um similar, mas não o mesmo, ela tende a reconhecer esse novo problema, oferecendo a mesma solução.



Aproximação de função

A característica mais significativa de redes neurais está em sua habilidade de aproximar qualquer função contínua ou não contínua com um grau de correção desejado.



Aprendizagem

Objetivo da Aprendizagem: descobrir a função $f(x)$ dado um número finito (desejável pequeno) de pares entrada-saída x, y .



Teorema da projeção linear

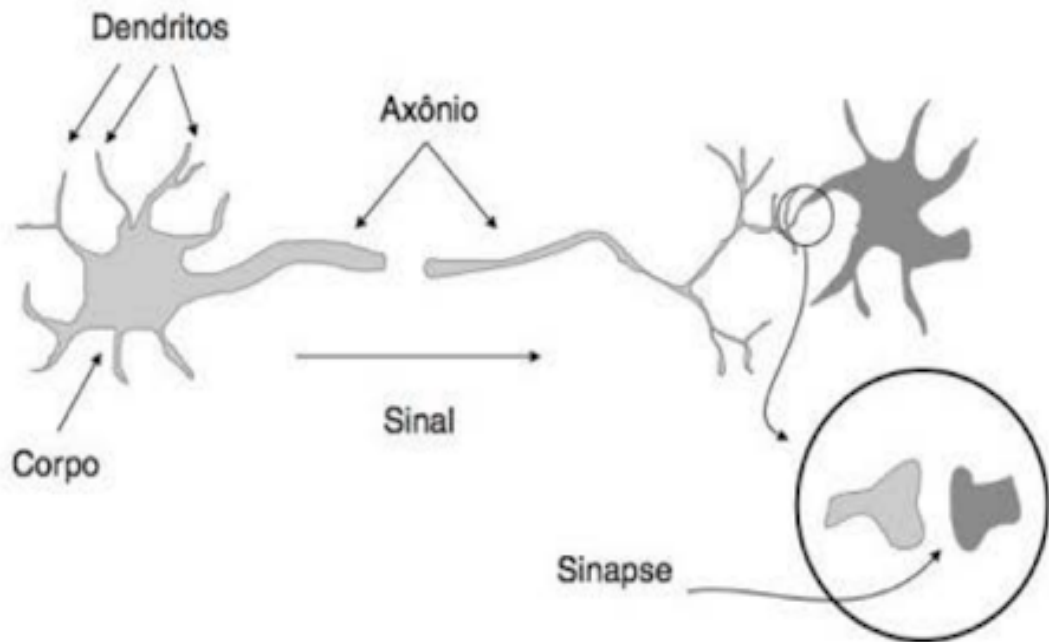
É possível descrever uma função $f(x)$, pela combinação de funções $\varphi_i(x)$ mais simples, em uma área compacta S do espaço de entrada:

$$f(x, w) = \sum_{i=1}^N w_i \varphi_i(x) \quad (1)$$

em que w_i são elementos reais do vetor $w = [w_1, \dots, w_N]$ tais que

$$f(x) - f(x, w) < \varepsilon \quad (2)$$

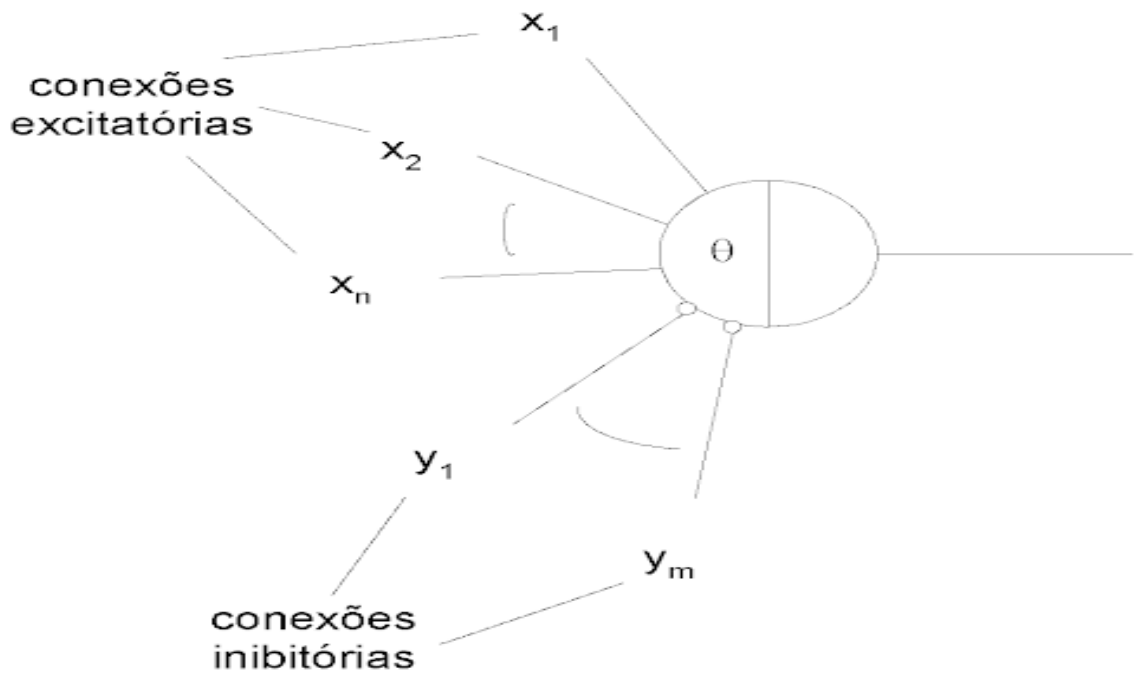
e ε sendo arbitrariamente pequeno. A função $f(x, w)$ é chamada **aproximante** e as funções $\varphi_i(x)$ são chamadas funções elementares.





O neurônio natural

- Dendritos tem por função receber os estímulos transmitidos pelos outros neurônios;
- Corpo é responsável por coletar e combinar informações vindas de outros neurônios;
- Axônio é constituído de uma fibra tubular que pode alcançar até alguns metros, e é responsável por transmitir os estímulos para outras células.





O neurônio artificial

- Executa uma função lógica.
- Os nós produzem resultados binários e as conexões transmitem exclusivamente zeros e uns.
- As redes são compostas de conexões sem peso, de tipos excitatórios e inibitórios.
- Cada unidade é caracterizada por um limiar ('threshold') q .



O aprendizado

O psicólogo Donald Hebb, demonstrou que a capacidade da aprendizagem em redes neurais biológicas vem da alteração da eficiência sináptica, isto é, a conexão somente é reforçada se tanto as células pré-sinápticas quanto as pós-sinápticas estiverem excitadas. Em 1986, Rumelhart, Hinton e Williams introduziram o poderoso método de treinamento denominado “Backpropagation”



A rede neural artificial

As unidades são dispostas em uma ou mais camadas e interligadas por um grande número de conexões, geralmente unidirecionais.

Na maioria das arquiteturas, essas conexões, que simulam as sinapses biológicas, possuem pesos associados, que ponderam a entrada recebida por cada neurônio da rede.

Os pesos podem assumir valores positivos ou negativos, dependendo de o comportamento da conexão ser excitatório ou inibitório, respectivamente.



Componentes básicos do neurônio artificial

- Sinapses (entradas), com seus pesos associados (mais um limiar)
- Junção somadora;
- A função de ativação.



Funcionamento básico do neurônio artificial

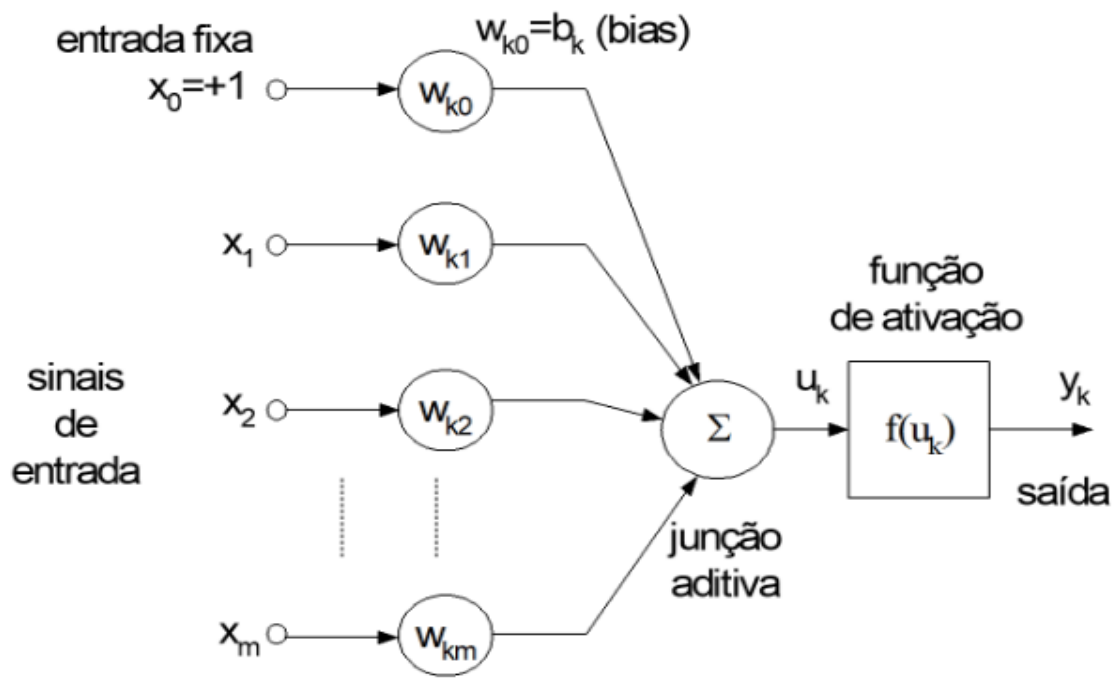
- Cada terminal de entrada do neurônio, simulando os dendritos, recebe um valor.
- Os valores recebidos são ponderados, somados e passam por uma função de ativação.
- A saída da função é a resposta do neurônio para a entrada.
- Várias funções diferentes podem ser utilizadas.
- A função de ativação limita a saída e introduz a não-linearidade ao modelo.



O limiar

O limiar tem o papel de aumentar ou diminuir a influência do valor das entradas.

É possível considerar o limiar como uma entrada de valor constante 1, multiplicado por um peso.





Expressão matemática do neurônio artificial

Considerando o k -ésimo neurônio recebendo j -ésimas entradas x , em que x_0 e w_{k0} representam o bias.

$$y_k = f(u_k) = f\left(\sum (w_{kj}x_j)\right) \quad (3)$$



Funções de ativação

- Função tangente hiperbólica (\tanh)
- Função logística (logistic)
- Função sinal (threshold)
- Função semi-linear



Exemplo de um neurônio

Inicia-se um neurônio com duas entradas ($w_1 = 0,5$ e $w_2 = 0,6$) além de um limiar ($w_0 = -0,3$)

$$u_k = 0,5x_1 + 0,6x_2 - 0,3x_0 \quad (4)$$

A função de ativação será uma função sinal, cuja saída dispara quando $u \geq 0$.



Exemplo de um neurônio

Será que o neurônio criado consegue separar os pontos
 $x_1 = 0; x_2 = 0$, $x_1 = 0; x_2 = 1$, $x_1 = 1; x_2 = 0$, do ponto $x_1 = 1; x_2 = 1$?



Exemplo de um neurônio

Se $x_1 = 0$; $x_2 = 0$, $u = 0,5 * 0 + 0,6 * 0 - 0,3 * 1 = -0,3$.

Como $u < 0$, $Y = 0$.

Se $x_1 = 0$; $x_2 = 1$, $u = 0,5 * 0 + 0,6 * 1 - 0,3 * 1 = 0,3$.

Como $u > 0$, $Y = 1$.

Se $x_1 = 1$; $x_2 = 0$, $u = 0,5 * 1 + 0,6 * 0 - 0,3 * 1 = 0,2$.

Como $u > 0$, $Y = 1$.

Se $x_1 = 1$; $x_2 = 1$, $u = 0,5 * 1 + 0,6 * 1 - 0,3 * 1 = 0,8$.

Como $u > 0$, $Y = 1$.



Exemplo de um neurônio

Adotaremos um processo de ajuste dos pesos. Testaremos as entradas uma a uma, em ciclos. A cada erro, reduziremos em 0,1 o peso da respectiva entrada e do limiar.



Exemplo de um neurônio

Se $x_1 = 0; x_2 = 0, u = 0,5 * 0 + 0,6 * 0 - 0,3 * 1 = -0,3$.

Como $u < 0$, $Y = 0$. Correto.

Se $x_1 = 0; x_2 = 1, u = 0,5 * 0 + 0,6 * 1 - 0,3 * 1 = 0,3$.

Como $u > 0$, $Y = 1$. Incorreto.

Corrigem-se os pesos: $w_1 = 0,5, w_2 = 0.5, w_0 = -0,4$.

Se $x_1 = 1; x_2 = 0, u = 0,5 * 1 + 0,5 * 0 - 0,4 * 1 = 0,1$.

Como $u > 0$, $Y = 1$. Incorreto.

Corrigem-se os pesos: $w_1 = 0,4, w_2 = 0.5, w_0 = -0,5$.

Se $x_1 = 1; x_2 = 1, u = 0,4 * 1 + 0,5 * 1 - 0,5 * 1 = 0,4$.

Como $u > 0$, $Y = 1$. Correto.



Exemplo de um neurônio

Recomeça.

Se $x_1 = 0$; $x_2 = 0$, $u = 0,4 * 0 + 0,5 * 0 - 0,5 * 1 = -0,5$.

Como $u < 0$, $Y = 0$. Correto.

Se $x_1 = 0$; $x_2 = 1$, $u = 0,4 * 0 + 0,5 * 1 - 0,5 * 1 = 0$.

Como $u = 0$, $Y = 1$. Incorreto.

Corrigem-se os pesos: $w_1 = 0,4$, $w_2 = 0,4$, $w_0 = -0,6$.

Se $x_1 = 1$; $x_2 = 0$, $u = 0,4 * 1 + 0,4 * 0 - 0,6 * 1 = -0,2$.

Como $u < 0$, $Y = 0$. Correto.

Se $x_1 = 1$; $x_2 = 1$, $u = 0,4 * 1 + 0,4 * 1 - 0,6 * 1 = 0,2$.

Como $u > 0$, $Y = 1$. Correto.



Exemplo de um neurônio

Recomeça.

Se $x_1 = 0$; $x_2 = 0$, $u = 0,4 * 0 + 0,4 * 0 - 0,6 * 1 = -0,6$.

Como $u < 0$, $Y = 0$. Correto.

Se $x_1 = 0$; $x_2 = 1$, $u = 0,4 * 0 + 0,4 * 1 - 0,6 * 1 = 0$.

Como $u < 0$, $Y = 0$. Correto.



Erro mínimo quadrático

$$E = \sum e_k^2 = \sum (\hat{y}_k - y_k)^2 \quad (5)$$

O peso w_{kj} será atualizado na direção oposta ao gradiente de E .

$$w'_{kj} = w_{kj} - \alpha \frac{\partial E}{\partial w_{kj}} \quad (6)$$

em que k é o k -ésimo neurônio e j é a j -ésima entrada. α indica a taxa de aprendizagem.



Erro mínimo quadrático

$$\frac{\partial E}{\partial w_{kj}} = \frac{\partial}{\partial w_{kj}} \sum (\hat{y}_k - y_k)^2 = \frac{\partial}{\partial w_{kj}} (\hat{y}_k - y_k)^2 \quad (7)$$

Sendo:

$$y_k = f\left(\sum (w_{kj} x_j)\right) \quad (8)$$

Tem-se que:

$$\frac{\partial E}{\partial w_{kj}} = -2(\hat{y}_k - y_k) \frac{\partial y_k}{\partial w_{kj}} = -2(\hat{y}_k - y_k) x_j \quad (9)$$



Regra delta para atualização dos pesos

$$w_{kj} = w_{kj} + \alpha(\hat{y}_k - y_i)x_j \quad (10)$$

$$w_{k0} = w_{k0} + \alpha(\hat{y}_k - y_i) \quad (11)$$



Erro mínimo quadrático matricial

$$W = W + \alpha \mathbf{e}_i \mathbf{x}_i^T \quad (12)$$

$$\mathbf{b} = \mathbf{b} + \alpha \mathbf{e}_i \quad (13)$$



Backpropagation

Para a camada de saída:

$$\delta_k(t) = f'_k[u_k(t)]e_k(t) \quad (14)$$

Para neurônios da camada oculta:

$$\delta_k(t) = f'_k[u_k(t)] \sum [\delta_j(t)w_{ji}] \quad (15)$$



Arquitetura de RNA

Usualmente as redes neurais são formadas por três tipos de camadas:

- Camada de entrada: são apenas sensoriais, e não fazem processamento.
- Camadas intermediárias: sempre fazem processamento.
- Camada de saída: podem fazer processamento.



Arquitetura de RNA

Arquitetura e superfície de decisão:

- Rede de uma camada: pode resolver problemas limitadas por um hiperplano.
- Rede de duas camadas: Divisão do espaço segue a complexidade limitada pelo número de neurônios da camada escondida.
- Rede de três camadas: Agrupamentos no espaço de solução segue complexidade limitada pelo número de neurônios da camada escondida.



Paradigmas de aprendizagem

- Supervisionado: fornece o conjunto de exemplos de entrada-resposta desejada.
- Não-supervisionado: os parâmetros livres da rede são otimizados a medida que os dados são apresentados.
- Com reforço: a única informação fornecida à rede é se uma determinada saída está correta ou não.



Problemas de aprendizagem

- Overfitting: Sobre-ajuste aos dados, inclusive dos ruídos presentes na amostra de treinamento.
- Saturação da função de ativação: para valores grandes de argumento, a função opera numa região de saturação.



Critérios de parada

- Número total de iterações.
- Um valor de erro suficientemente pequeno.
- A variação do erro de uma época para outra atingir um valor suficientemente pequeno.
- Capacidade de generalização.