Tópicos Especiais em Computação I

Redes Neurais

Patrícia Lucas

Bacharelado em Sistemas de Informação
IFNMG - Campus Salinas

Salinas
Março 2021

INSTITUTO FEDERAL
Norte de Minas Gerais
Campus Salinas

Versão 1.2021

Referência

Redes Neurais

Capítulo 7: Modelos baseados em otimização.

Inteligência Artificial: Uma abordagem de aprendizado de máquina. Katti Faceli...[et al.]. - Rio de Janeiro: LTC, 2011.

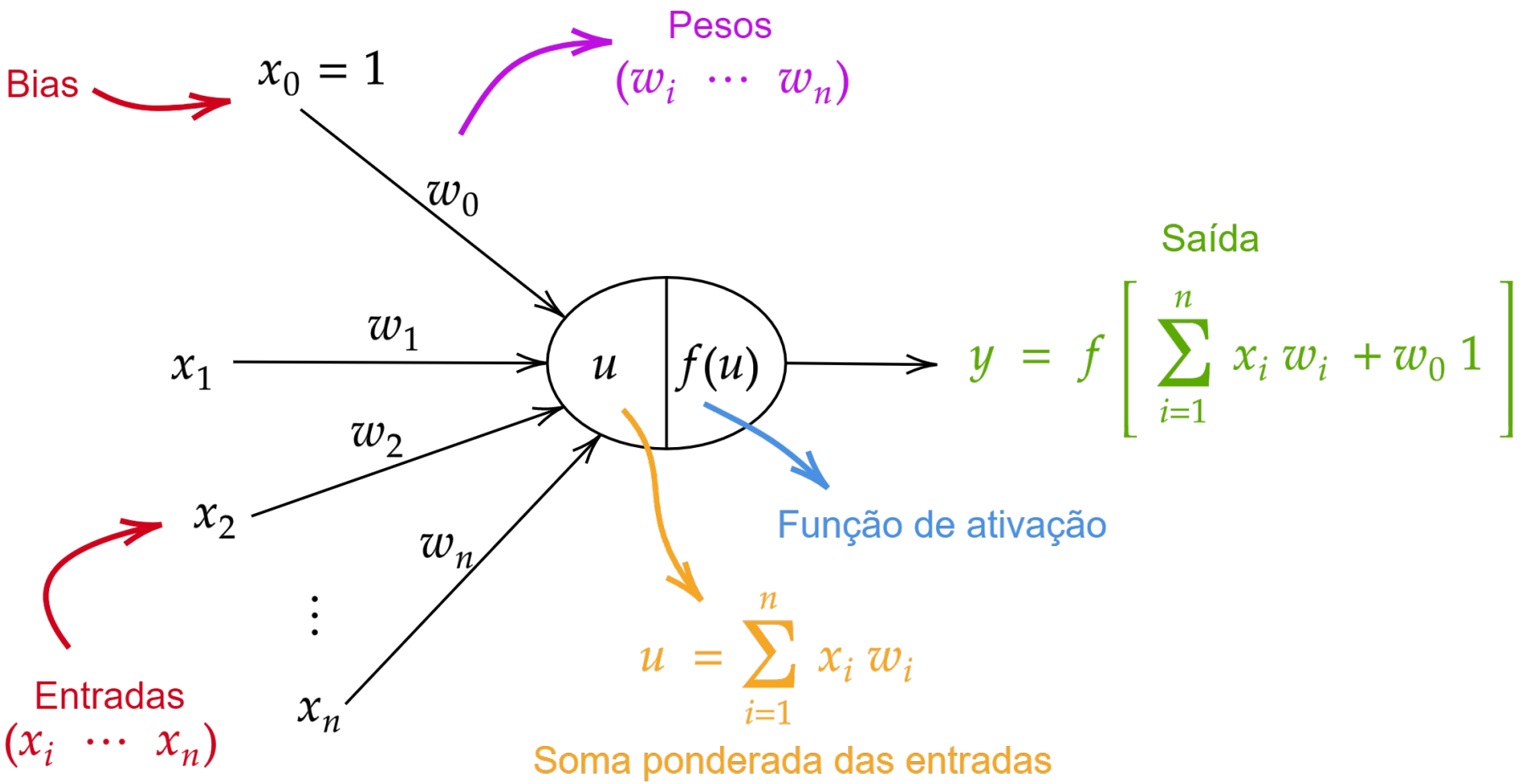
Roteiro

Redes Neurais

- O neurônio artificial
- Funções de ativação
- Aprendizado supervisionado
- Perceptron
- Adaline
- Perceptron de múltiplas camadas - MLP

O neurônio artificial

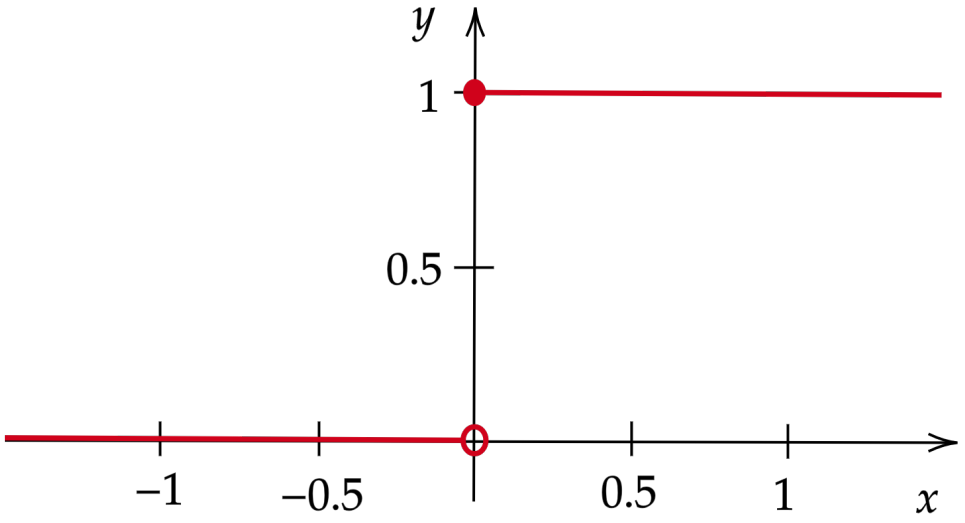
Redes Neurais



Funções de ativação

Redes Neurais

Função degrau: usada na camada de saída para problemas de classificação.

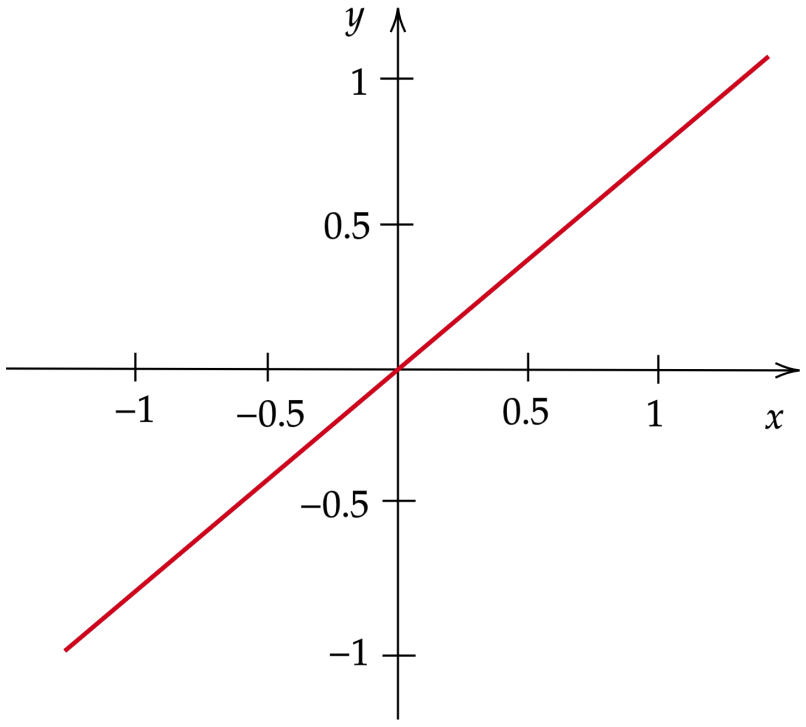


$$f(u) = \begin{cases} 1 & \text{se } u \geq 0 \\ 0 & \text{se } u < 0 \end{cases} \tag{1}$$

Funções de ativação

Redes Neurais

Função linear: usada na camada de saída para problemas de regressão.



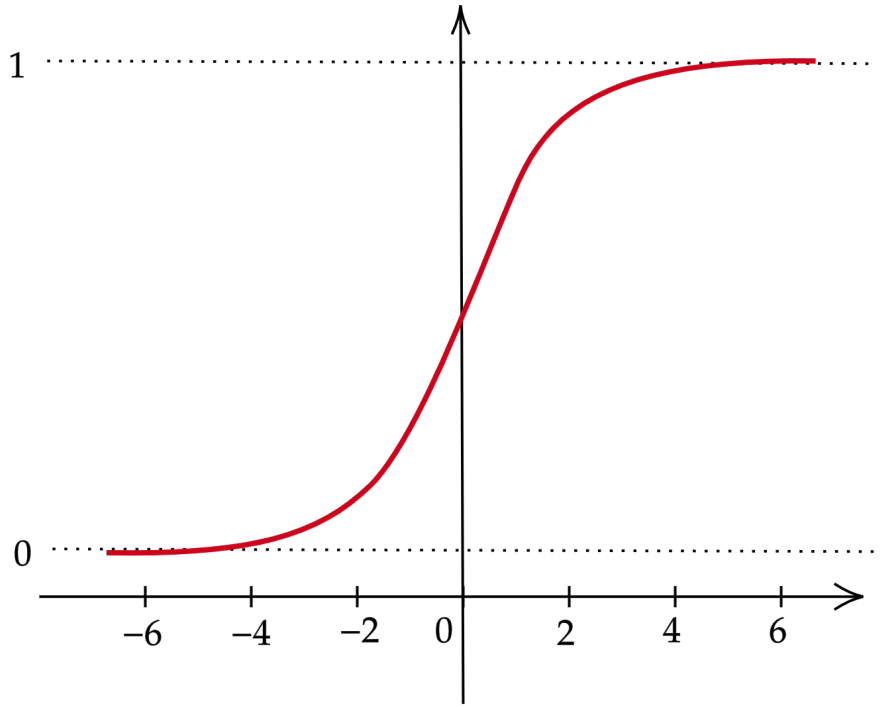
$$f(u) = u \tag{2}$$

Intervalo: $[-\infty, +\infty]$

Funções de ativação

Redes Neurais

Função sigmóide: usada na camada de saída para problemas de classificação binária.



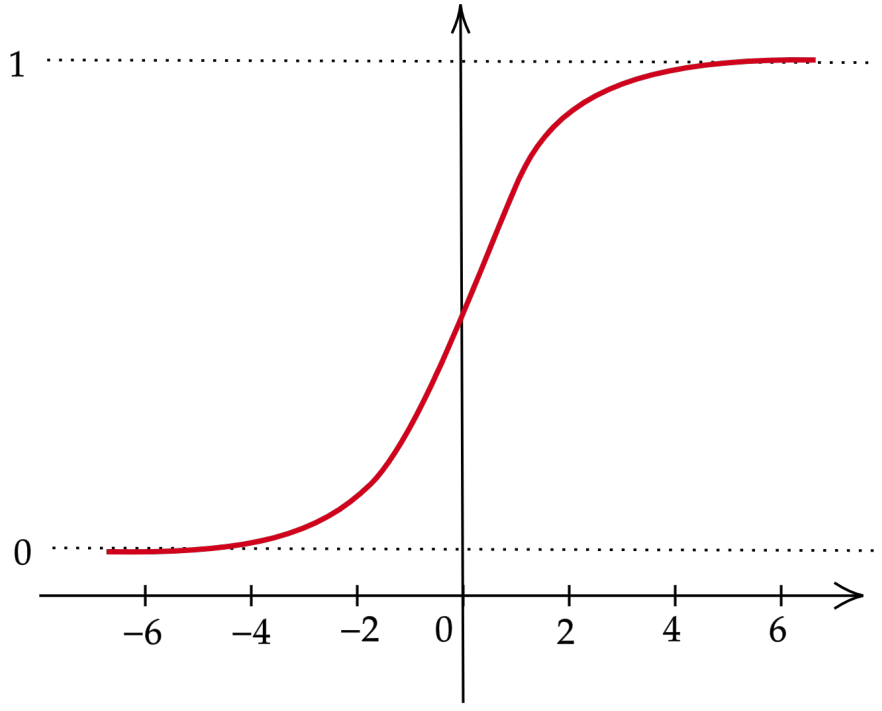
$$f(u) = \frac{1}{1 + e^{-u}} \tag{3}$$

Intervalo: $[0, 1]$

Funções de ativação

Redes Neurais

Função softmax: usada na camada de saída para problemas de classificação multiclasse.



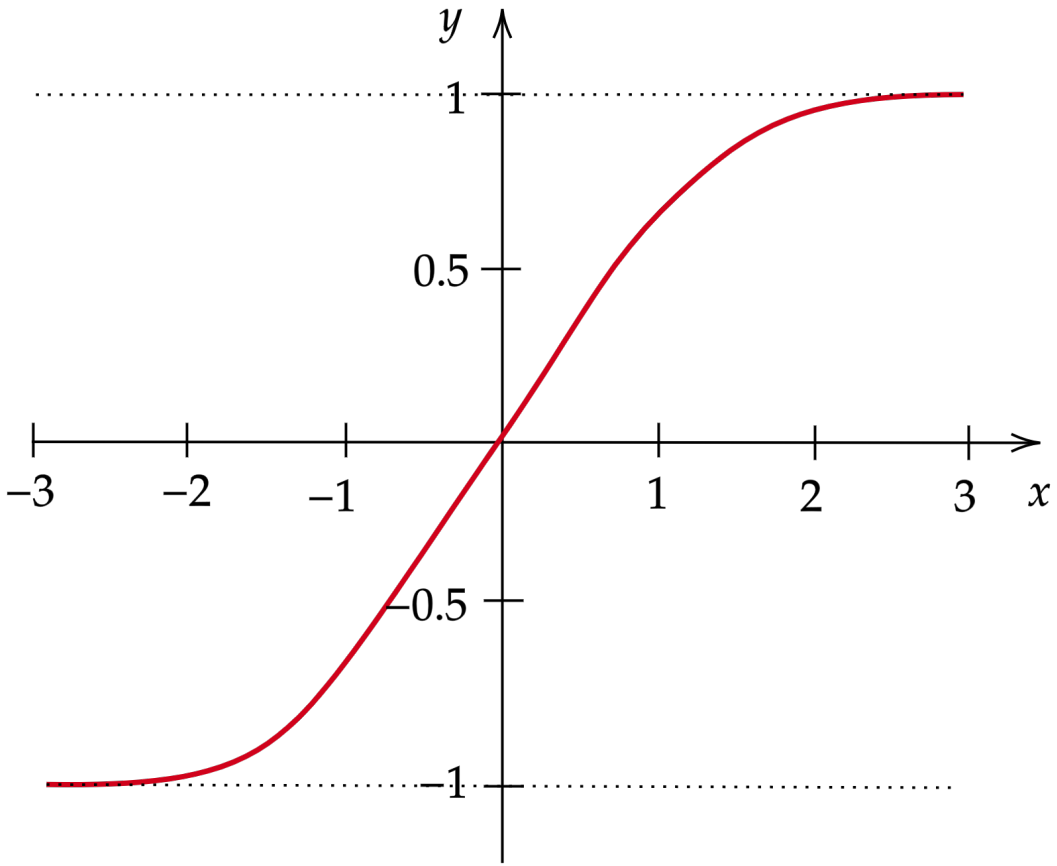
$$f(u) = \frac{e^{u_i}}{\sum e^{u_i}} \tag{4}$$

Intervalo: $[0, 1]$

Funções de ativação

Redes Neurais

Função tangente hiperbólica: usada na camada escondida.



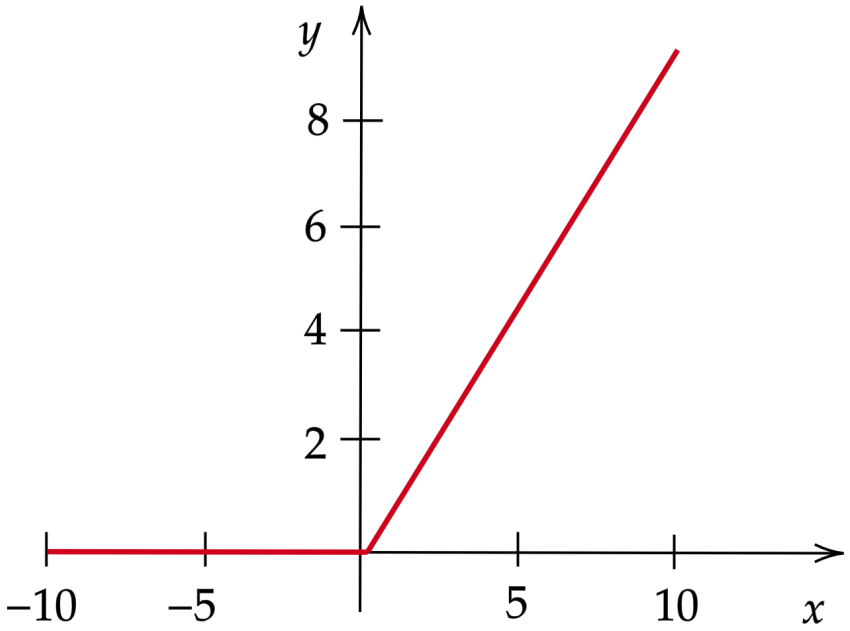
$$f(u) = \frac{e^{u} - e^{-u}}{e^{u} + e^{-u}} \tag{5}$$

Intervalo: $[-1, 1]$

Funções de ativação

Redes Neurais

Função Relu (unidade linear retificada): usada na camada escondida.



$$f(u) = max(0, u) \tag{6}$$

$$f(u) = \begin{cases} 0 & \text{se } u < 0 \\ u & \text{se } u \geq 0 \end{cases}$$

Aprendizado supervisionado

Redes Neurais

Nesse tipo de aprendizado, os pesos são ajustados através da comparação da saída da rede com a saída desejada.

O método mais comum de aprendizado supervisionado é o aprendizado por correção de erros, em que procura-se minimizar o erro da resposta atual da rede em relação à saída desejada.

O erro:

$$e = y - \hat{y} \tag{7}$$

Aprendizado supervisionado

Redes Neurais

Para a atualização dos pesos por correção dos erros usa-se a equação 8, chamada de Regra Delta:

$$w_i(t+1) = w_i(t) + \eta e x_i \tag{8}$$

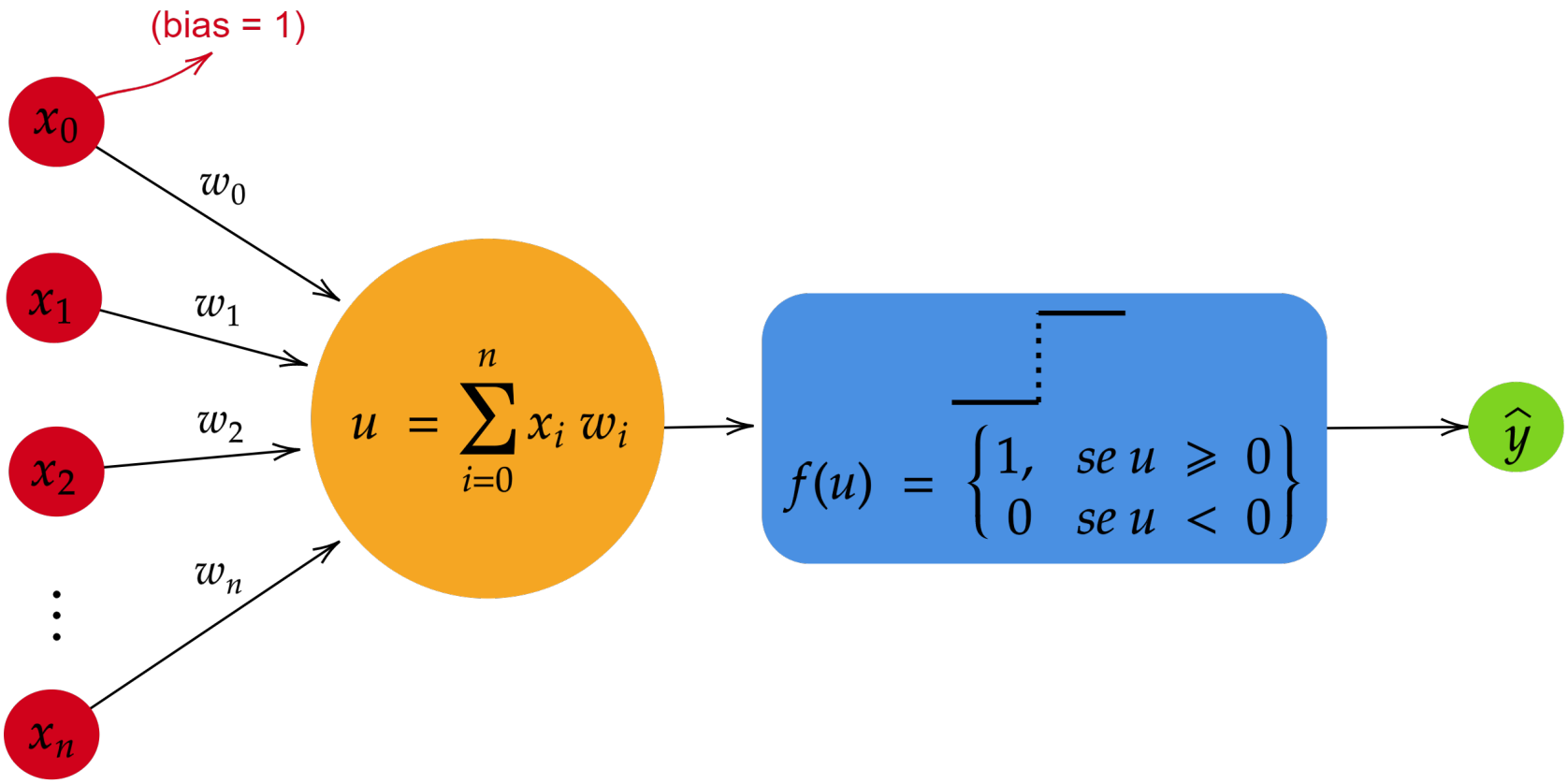
onde:

- $w_i(t+1)$ = novos pesos
- $w_i(t)$ = pesos atuais
- η = taxa de aprendizagem
- e = erro
- x_i = entradas

Perceptron

Redes Neurais

O perceptron é uma rede neural de camada única que resolve problemas de classificação linear (binária).



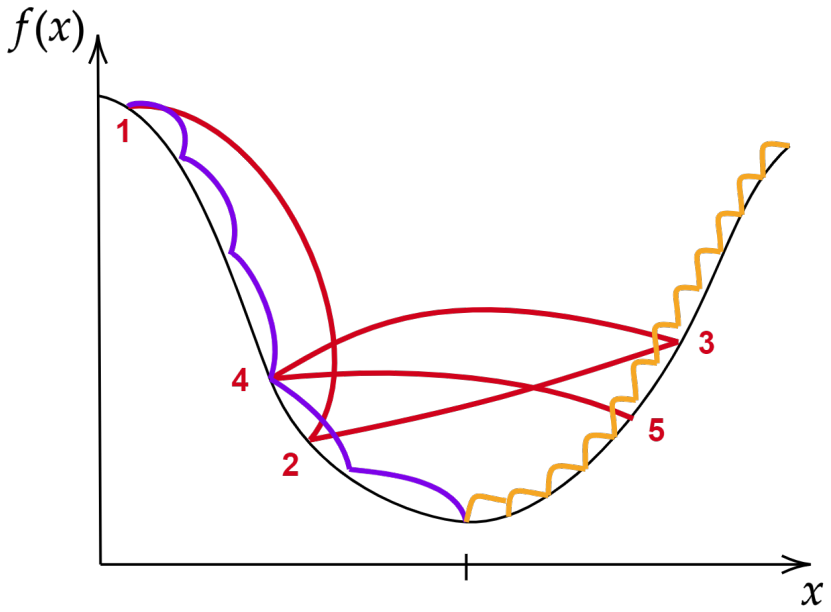
O algoritmo Perceptron

1. Inicializar η
2. Inicializar os pesos w com valores aleatórios
3. Repita até $\sum e^2 = 0$
   - 3.1 Calcule $f(u)$
   - 3.2 Calcule e
   - 3.3 Aplique a regra de atualização dos pesos para todos os pares (x_i, y_i) do conjunto de treinamento
   - 3.4 Calcule $\sum e^2$

Algumas considerações...

Redes Neurais

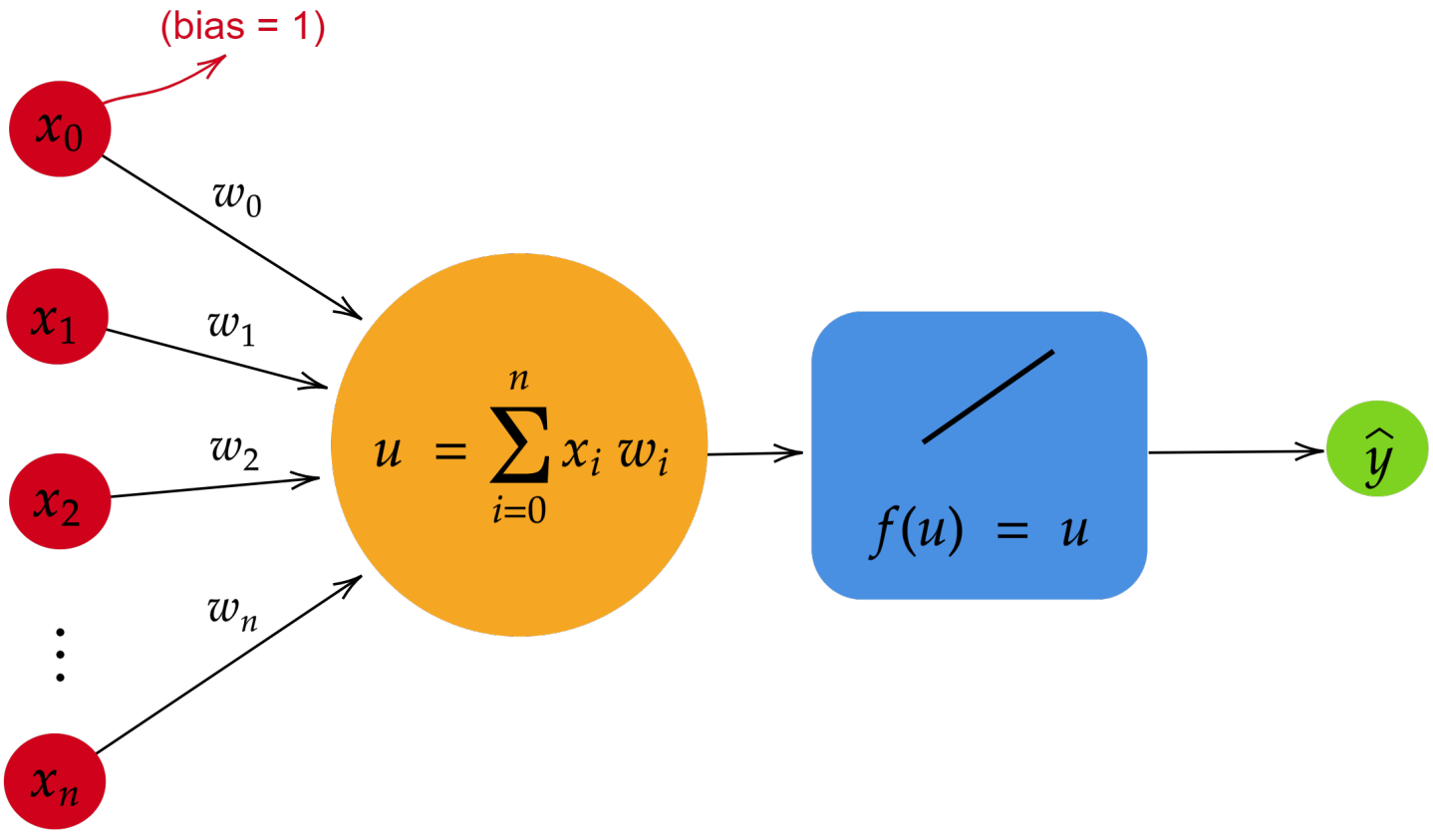
- Valor da taxa de aprendizado (η): valores muito pequenos podem levar a um tempo de convergência muito alto, enquanto que valores muito altos podem levar a instabilidade no treinamento.
- Inicialização dos pesos: em geral devem ser inicializados com valores no intervalo $[-0.5, 0.5]$.
- Normalizar e embaralhar as entradas.



Adaline

Redes Neurais

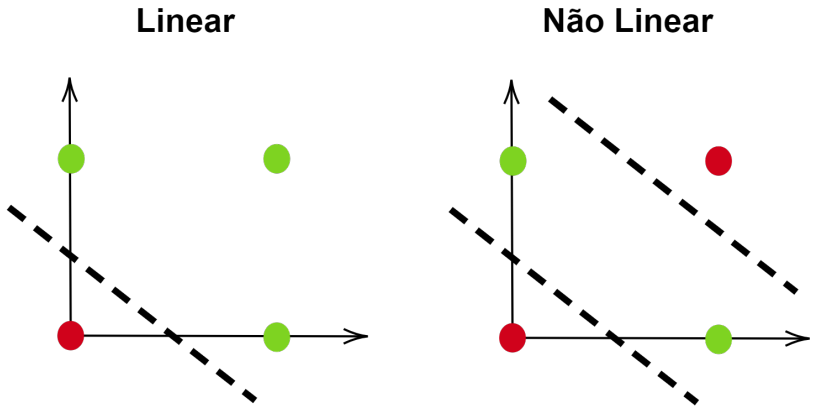
O adaline é uma rede neural de camada única que resolve problemas de regressão.



Perceptron de múltiplas camadas - MLP

Redes Neurais

- As redes vista até aqui, com uma única camada, têm a limitação de resolver apenas problemas com características lineares.



- No entanto, a não-linearidade é inerente à maioria das situações e problemas reais, sendo necessária a utilização de estruturas com características não-lineares para resolução de problemas de maior complexidade.

- As não-linearidades são incorporadas a modelos neurais através das funções de ativação (não-lineares) de cada neurônio da rede e da composição da sua estrutura em camadas sucessivas.

Perceptron de múltiplas camadas - MLP

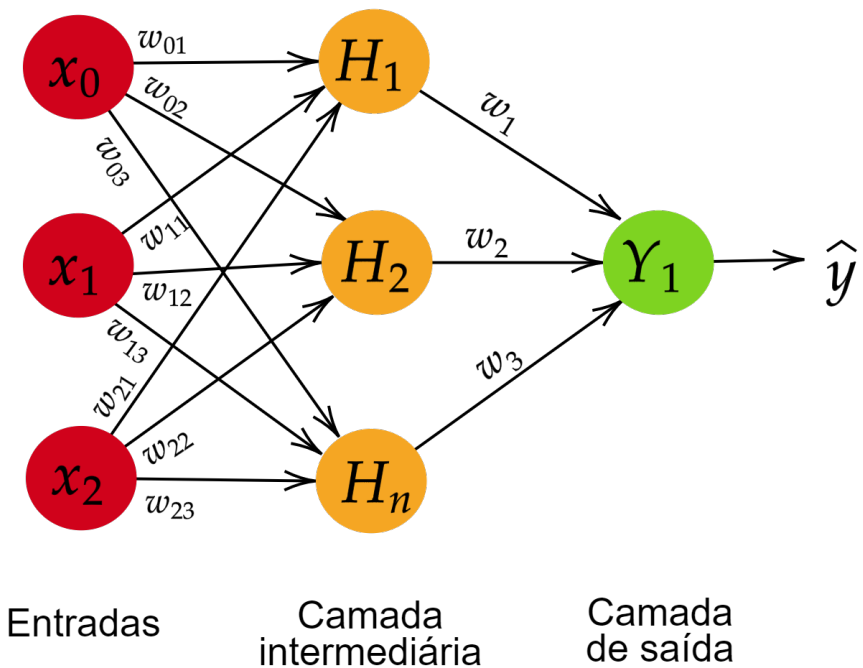
Redes Neurais

Arquitetura: o comportamento de uma MLP com 1 camada intermediária, pode ser descrito por meio de duas transformações sucessivas:

$$Y(H(x, w_i), w_s) \tag{9}$$

onde:

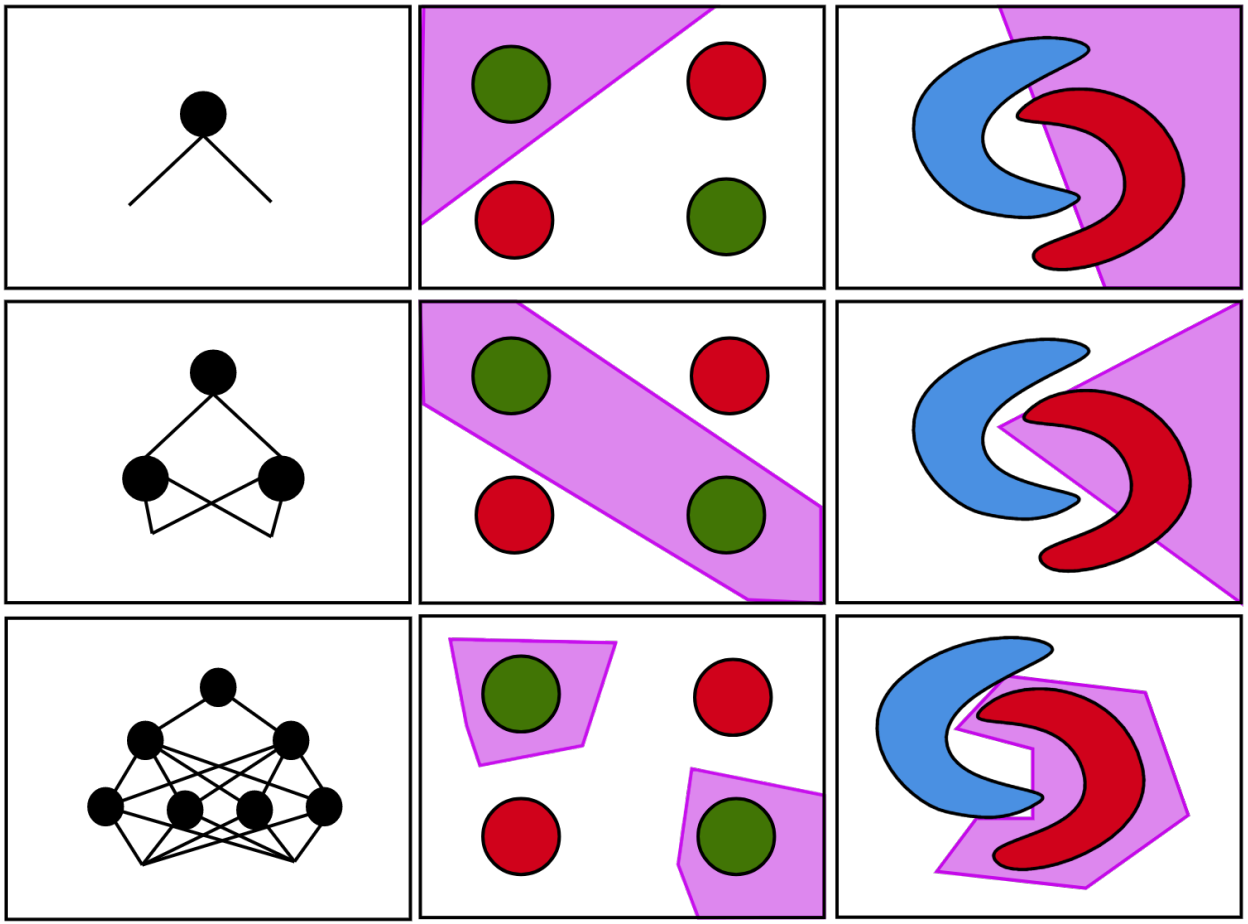
- w_i são os pesos da camada intermediária.
- w_s são os pesos da camada de saída.



Perceptron de múltiplas camadas - MLP

Redes Neurais

Número de camadas:



Número de neurônios:

- O número de neurônios determina a capacidade da rede em resolver problemas de determinada complexidade.
- Quanto maior o número de neurônios, maior a complexidade da rede e maior a sua abrangência em termos de soluções possíveis.
- A determinação do número de neurônios é o problema mais fundamental em aprendizado de redes neurais, mas ainda não existe uma regra geral para resolvê-lo.

Perceptron de múltiplas camadas - MLP

O treinamento de redes de uma única camada por meio do aprendizado supervisionado e correção de erros é realizado através da equação:

$$w_i(t+1) = w_i(t) + \eta e x_i \tag{10}$$

$$e = y - \hat{y} \tag{11}$$

No entanto, para redes de múltiplas camadas esse procedimento pode ser aplicado somente na camada de saída, já que não existem saídas desejadas nas camadas intermediárias!

Perceptron de múltiplas camadas - MLP

Solução: algoritmo de retropropagação de erros (back-propagation).

- O princípio desse algoritmo é, utilizando-se do gradiente descendente, estimar o erro das camadas intermediárias por meio de uma estimativa do efeito que estas causam no erro da camada de saída.
- Assim, o erro de saída da rede é calculado e este é retroalimentado para as camadas intermediárias, possibilitando o ajuste dos pesos proporcionalmente aos valores das conexões entre as camadas.
- Requer o uso de funções de ativação contínuas e diferenciáveis.