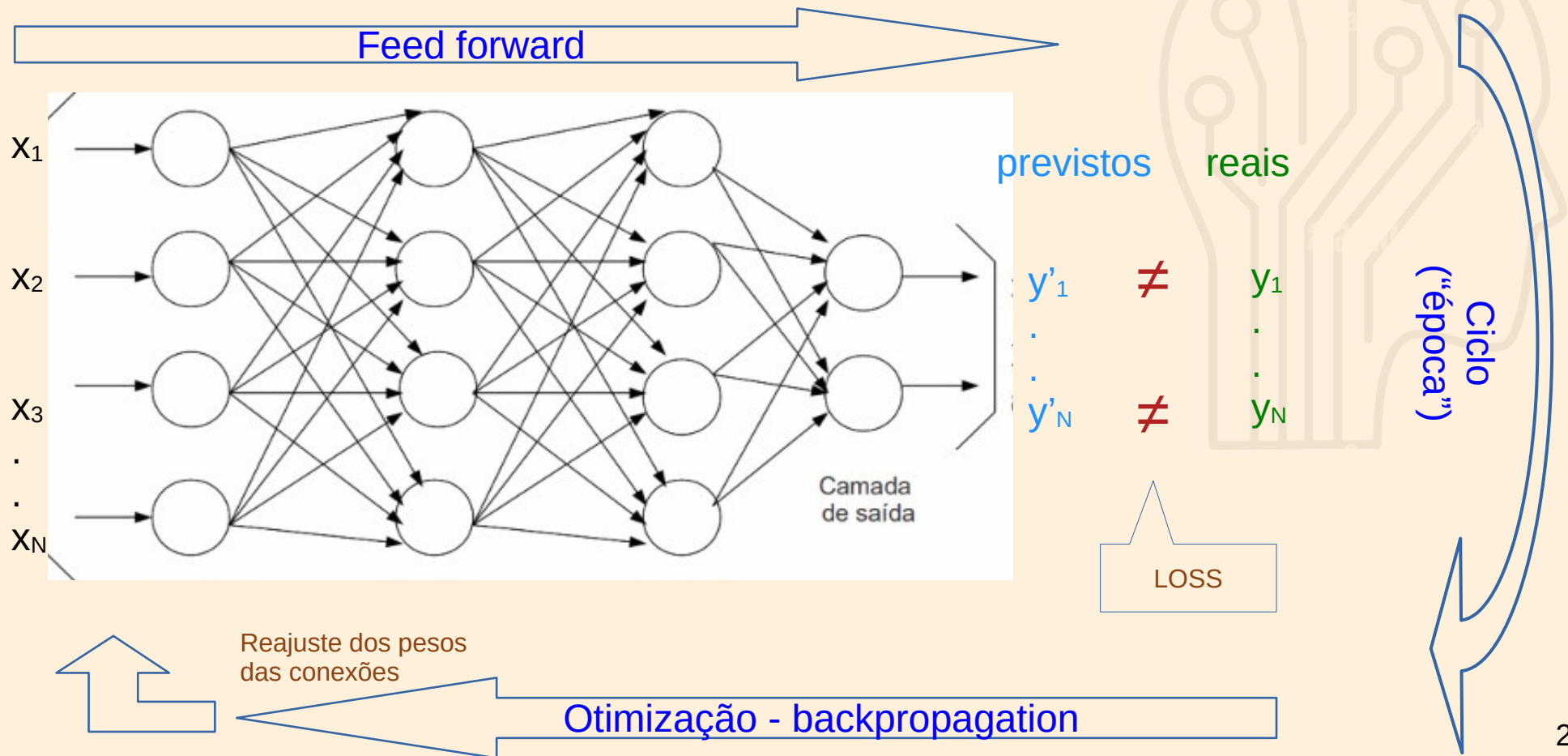


Como a rede neural “aprende”

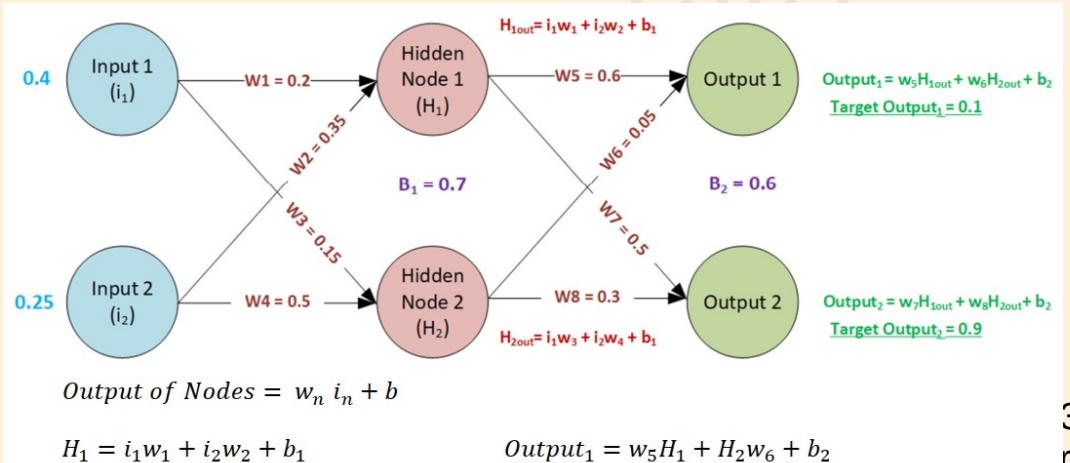
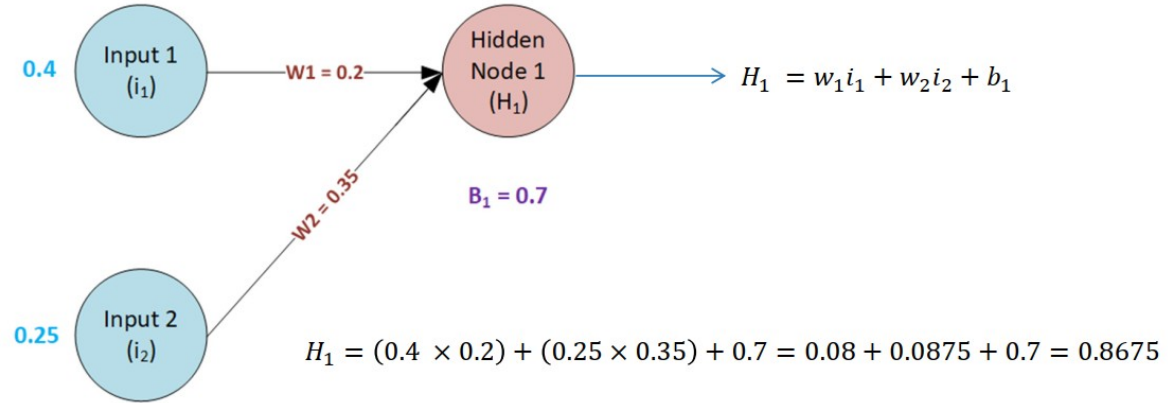
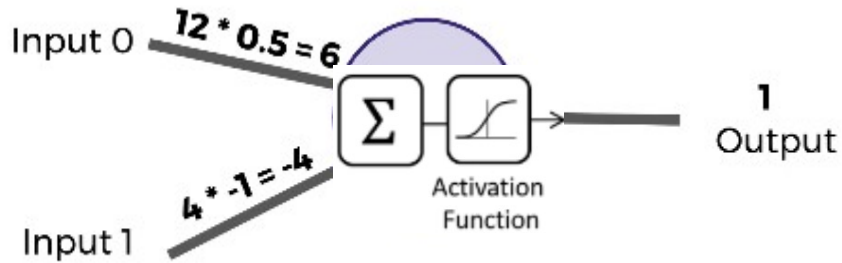
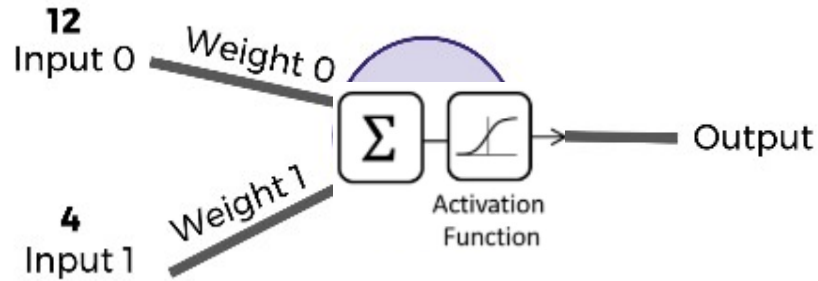


- Uma RNA é considerada “treinada” quando
 - os pesos das conexões entre neurônios atingem um valor tal que
 - a rede é capaz de associar corretamente os dados da camada de entrada
 - com os neurônios “ativados” na camada de saída
- O processo de fazer esses pesos convergirem ao valor ideal é chamado “treinamento” da RNA

Como a rede neural “aprende”



Feed forward



$$Output\ of\ Nodes = w_n i_n + b$$

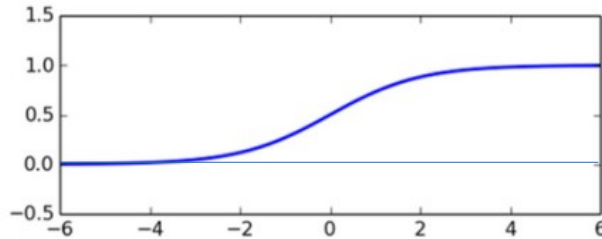
$$H_1 = i_1 w_1 + i_2 w_2 + b_1$$

$$Output_1 = w_5 H_1 + H_2 w_6 + b_2$$

Funções de ativação

$$\text{saída do neurón} = f_{ATIV}\left(\sum (w_i x_i + b_i)\right)$$

ReLU

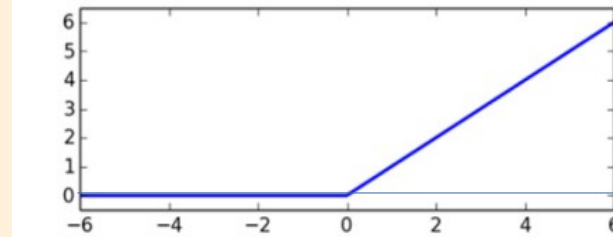
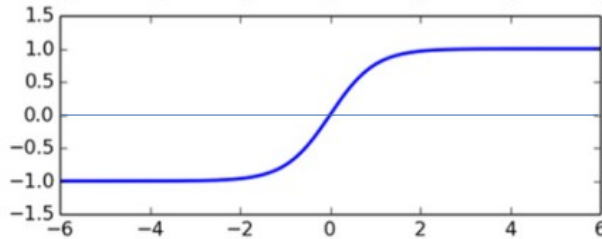


Sigmoid

$$\phi(z) = \frac{1}{1 + e^{-z}}$$

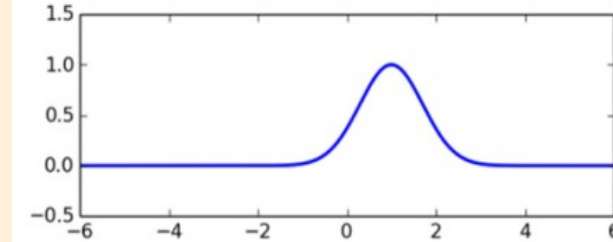
Hyperbolic Tangent

$$\phi(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$



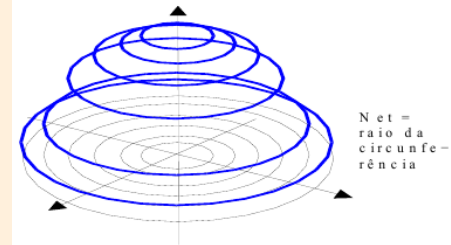
Rectified Linear

$$\phi(z) = \begin{cases} 0 & \text{if } z < 0 \\ z & \text{if } z \geq 0 \end{cases}$$



Radial Basis Function

$$\phi(z, c) = e^{-(\epsilon \|z - c\|)^2}$$



Backpropagation = processo para minimizar o erro

- Ajuste dos pesos das conexões
- O processo de ajuste de pesos por backpropagation envolve os seguintes passos:
 - 1) **Forward Pass**: a rede neural processa a entrada e produz uma saída prevista.
 - 2) **Cálculo do Erro**: o erro entre a saída prevista e a saída real é calculado utilizando uma função de erro, como o erro quadrático médio (**MSE**) ou a entropia cruzada.
 - 3) **Backward Pass**: o erro é propagado de volta pela rede neural, de forma a calcular o gradiente do erro em relação a cada peso e viés (bias) dos neurônios.
 - 4) **Ajuste de Pesos**: os pesos e vieses dos neurônios são ajustados com base no gradiente do erro, utilizando uma taxa de aprendizado (**learning rate**) para controlar a magnitude do ajuste.

“Erros” (losses)

<https://keras.io/api/losses/>

- MSE (minimum square error)

$$MSE = \frac{1}{2}(\text{previsto} - \text{real})^2$$

- O $\frac{1}{2}$ é incluído para cancelar o expoente quando diferenciado (parece melhor)
- Binary Crossentropy
 - Quantifica a diferença entre os rótulos de duas classes (onde $y=0$ falso ou $y=1$ verdadeiro) e as **probabilidades previstas** geradas pelo modelo.

$$BCE = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)]$$

- Categorical Cross-Entropy (CCE)

- também conhecida como **softmax**
- Diferença entre a probabilidade prevista e a probabilidade real dentre várias categorias

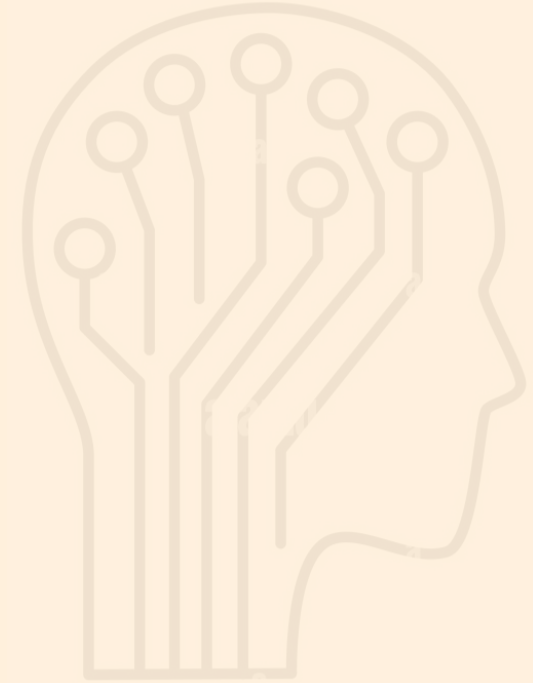
$$L(y, \hat{y}) = - \sum_{i=1}^C y_i \log(\hat{y}_i)$$

- Sparse Categorical Crossentropy

- Nos casos em que para y não foi usado One Hot Encoded
- y é um índice inteiro

Pseudo algoritmo

- Inicialize os pesos das conexões com valores pequenos e aleatórios
- Repita:
 - Entrar com os dados de treinamento X na camada de entrada
 - Embaralhe o conjunto de dados de treinamento
 - Executar o processamento da rede (*forward propagation*)
 - Tomar os valores na camada de saída
 - Se não corresponder à resposta desejada abaixo de uma tolerância
 - cálculo do erro ou LOSS
 - REAJUSTAR os pesos (*backpropagation*) de acordo com uma das estratégias
 - Após ler todo o conjunto de dados de treinamento - *Naive Gradient Decent*
 - A cada dado x_i de treinamento lido - *Stochastic gradient descent (SGD)*
 - Após um lote (*batch*) subconjunto dos dados X de treinamento - *Mini Batch Gradient Descent*
- Até que não haja mais REAJUSTE de pesos ou atingir o limite definido de ciclos
- Quando a RNA “acertar” (quase) todas as saídas => TREINADA

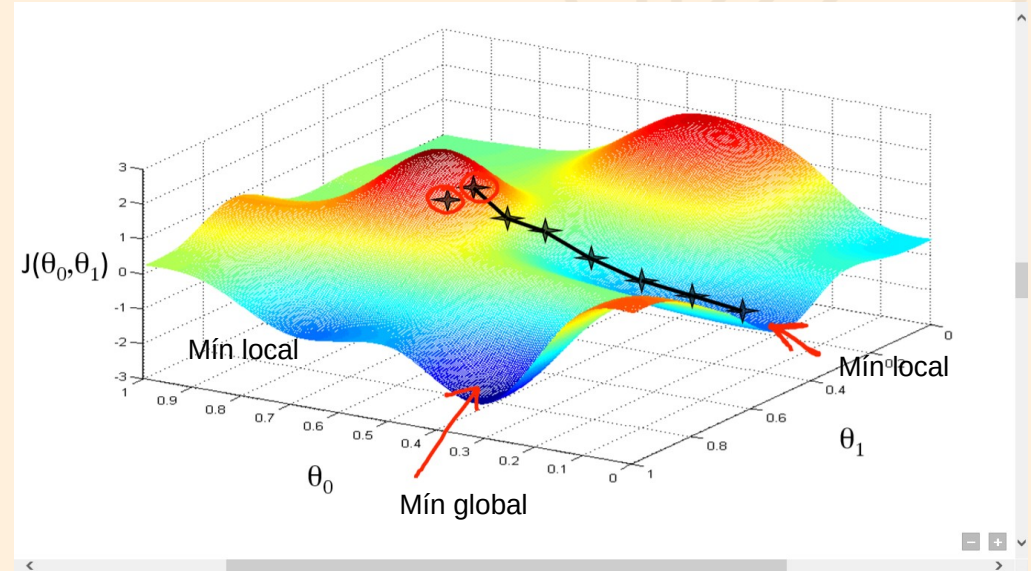


Minimizar o erro

- Minimizar o erro entre a classe prevista e a classe real
- No programa:
- Gradiente descendente sobre a “função erro” (*loss*)

Avalie o desempenho do modelo

```
accuracy =  
accuracy_score(y_test, y_pred)
```



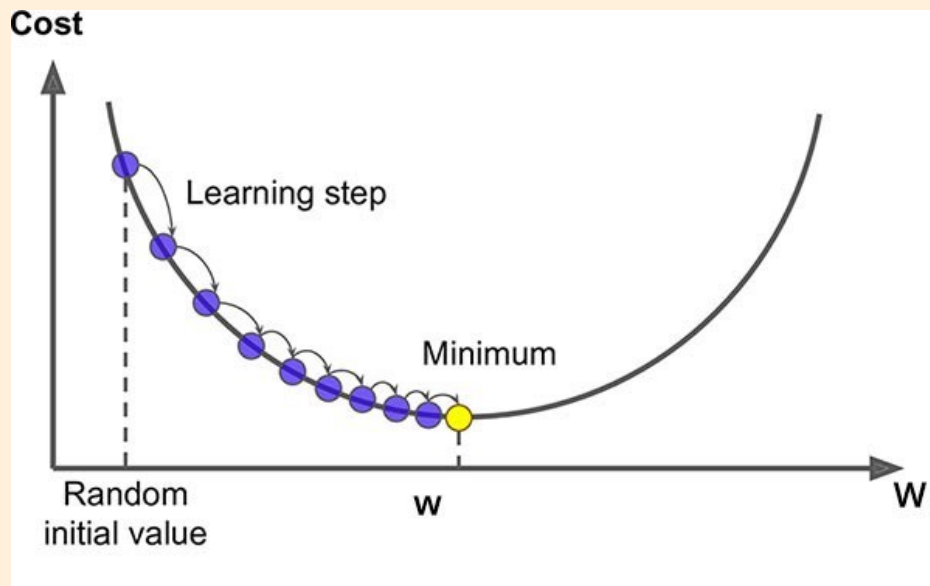
Reajuste das conexões

- A fórmula para ajustar os pesos (*weights*) por backpropagation é a seguinte:
- $w'_{ij} = w_{ij} - \alpha \cdot \partial E / \partial w_{ij}$
- em que:
- w_{ij} é o peso da conexão entre o neurônio i e o neurônio j
- α é a taxa de aprendizado (*learning rate*)
- E é o erro total da rede neural
- $\partial E / \partial w_{ij}$ é o gradiente do erro em relação ao peso w_{ij}

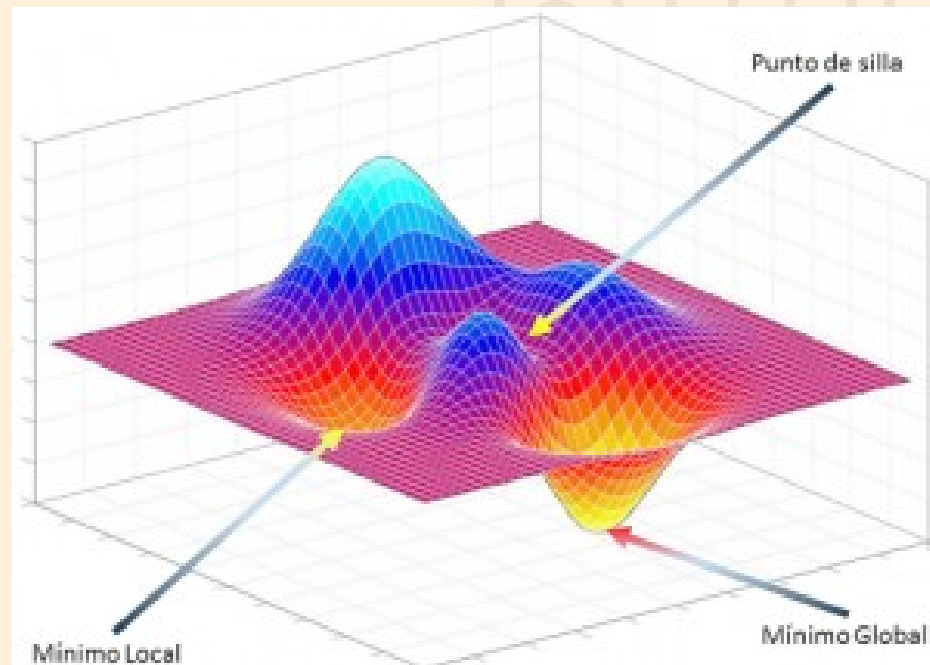
Learning rate

- A taxa de aprendizado simplesmente controla o salto de magnitude na direção de $\partial E / \partial w_{ij}$
- As taxas de aprendizado são sempre positivas e variam entre 0 e 1
- Uma taxa de aprendizado grande permitirá um treinamento mais rápido,
 - mas pode ultrapassar o mínimo global (ficando preso em um mínimo local).
- Um aprendizado pequeno levará mais tempo para ser treinado, mas encontrará o mínimo global de forma mais confiável.

Atingir o mínimo do erro



Taxa de aprendizado α variável



Ideal = mínimo global
Mas pode cair em mínimo local

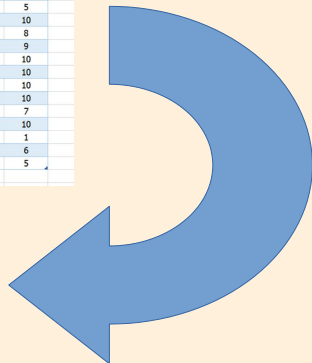
Estratégias de aplicação do backpropagation

- REAJUSTAR os pesos (*backpropagation*) de acordo com uma das estratégias
 - Após ler todo o conjunto de dados de treinamento - *Naïve Gradient Decent*
 - A cada dado x_i de treinamento lido - *Stochastic gradient descent (SGD)*
 - Após um lote (*batch*) subconjunto dos dados X de treinamento - *Mini Batch Gradient Descent*

Naïve gradient descent

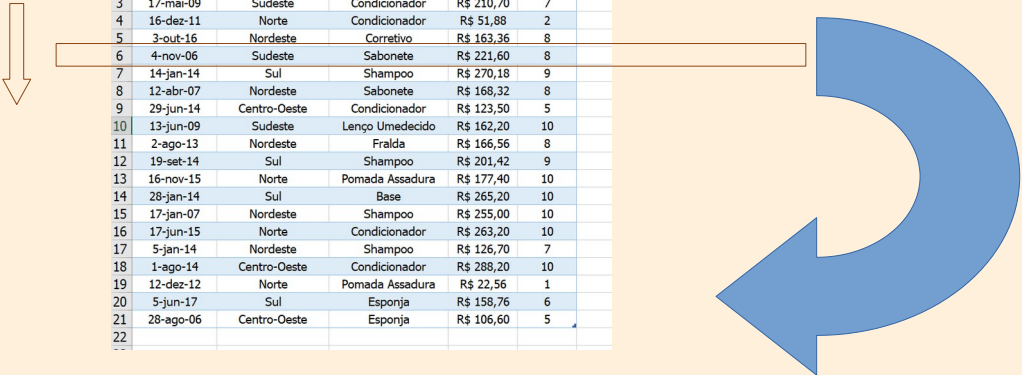
- Calcula o erro para todos os dados e depois atualiza os pesos das conexões
- Em cada ciclo (época)
- Lento, demora a convergir
- Maior acurácia
- Taxa de aprendizado fixa
- Armadilha em mínimo local

	A	B	C	D	E	F
1	Orig	Região	Nome do Produto	Preço	Volum	
2	22-ago-09	Sudeste	Perfume	R\$ 170,56	8	
3	17-mai-09	Sudeste	Condicionador	R\$ 210,70	7	
4	16-dez-11	Norte	Condicionador	R\$ 51,88	2	
5	3-out-16	Nordeste	Corretivo	R\$ 163,36	8	
6	4-nov-06	Sudeste	Sabonete	R\$ 221,60	8	
7	14-jan-14	Sul	Shampoo	R\$ 270,18	9	
8	12-fev-07	Nordeste	Sabonete	R\$ 166,32	8	
9	29-jun-14	Centro-Oeste	Condicionador	R\$ 123,50	5	
10	13-jun-09	Sudeste	Lenço Umedecido	R\$ 162,20	10	
11	2-ago-13	Nordeste	Fralda	R\$ 166,56	8	
12	19-set-14	Sul	Shampoo	R\$ 201,42	9	
13	16-nov-15	Norte	Pomada Assadura	R\$ 177,40	10	
14	28-jan-14	Sul	Bate	R\$ 265,20	10	
15	17-jan-07	Nordeste	Shampoo	R\$ 255,00	10	
16	17-jun-15	Norte	Condicionador	R\$ 263,20	10	
17	5-jan-14	Nordeste	Shampoo	R\$ 126,70	7	
18	1-ago-14	Centro-Oeste	Condicionador	R\$ 288,20	10	
19	12-dez-12	Norte	Pomada Assadura	R\$ 22,56	1	
20	5-jun-17	Sul	Espanja	R\$ 158,76	6	
21	28-ago-06	Centro-Oeste	Espanja	R\$ 106,60	5	
22						



Stochastic gradient descent (SGD)

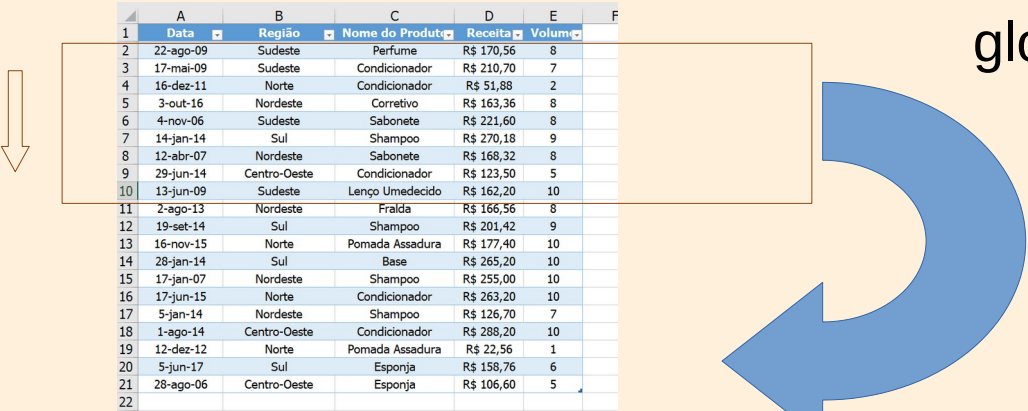
- Calcula o erro dado-a-dado (ou pequeno subconjunto)
- Converge mais rápido
- Consegue escapar do mínimo local
- Para garantir natureza estocástica, os dados devem ser embaralhados a cada ciclo (época)
- Taxa de aprendizado ajustada dinamicamente



	A	B	C	D	E	F
1	Data	Região	Nome do Produto	Receita	Volum	
2	22-ago-09	Sudeste	Perfume	R\$ 170,56	8	
3	17-mai-09	Sudeste	Condicionador	R\$ 210,70	7	
4	16-dez-11	Norte	Condicionador	R\$ 51,88	2	
5	3-out-16	Nordeste	Corretivo	R\$ 163,36	8	
6	4-nov-06	Sudeste	Sabonete	R\$ 221,60	8	
7	14-jan-14	Sul	Shampoo	R\$ 270,18	9	
8	12-abr-07	Nordeste	Sabonete	R\$ 168,32	8	
9	29-jun-14	Centro-Oeste	Condicionador	R\$ 123,50	5	
10	13-jun-09	Sudeste	Lenço Umedecido	R\$ 162,20	10	
11	2-ago-13	Nordeste	Fralda	R\$ 166,56	8	
12	19-set-14	Sul	Shampoo	R\$ 201,42	9	
13	16-nov-15	Norte	Pomada Assadura	R\$ 177,40	10	
14	28-jan-14	Sul	Base	R\$ 265,20	10	
15	17-jan-07	Nordeste	Shampoo	R\$ 255,00	10	
16	17-jun-15	Norte	Condicionador	R\$ 263,20	10	
17	5-jan-14	Nordeste	Shampoo	R\$ 126,70	7	
18	1-ago-14	Centro-Oeste	Condicionador	R\$ 288,20	10	
19	12-dez-12	Norte	Pomada Assadura	R\$ 22,56	1	
20	5-jun-17	Sul	Esponja	R\$ 158,76	6	
21	28-ago-06	Centro-Oeste	Esponja	R\$ 106,60	5	
22						

Mini Batch Gradient Descent

- É uma combinação dos dois.
- Necessário um lote de amostras de entrada
 - e atualiza o gradiente depois que esse lote é processado
- Os lotes são tipicamente potências de 2 (32, 64, 128, 256, 512) embora não exista uma regra clara
- Isso leva a um treinamento mais rápido (ou seja, convergência mais rápida para os mínimos globais)



	A	B	C	D	E	F
1	Data	Região	Nome do Produto	Receita	Volum	
2	22-ago-09	Sudeste	Perfume	R\$ 170,56	8	
3	17-mai-09	Sudeste	Condicionador	R\$ 210,70	7	
4	16-dez-11	Norte	Condicionador	R\$ 51,88	2	
5	3-out-16	Nordeste	Corretivo	R\$ 163,36	8	
6	4-nov-06	Sudeste	Sabonete	R\$ 221,60	8	
7	14-jan-14	Sul	Shampoo	R\$ 270,18	9	
8	12-abr-07	Nordeste	Sabonete	R\$ 168,32	8	
9	29-jun-14	Centro-Oeste	Condicionador	R\$ 123,50	5	
10	13-jun-09	Sudeste	Lenço Umedecido	R\$ 162,20	10	
11	2-ago-13	Nordeste	Fralda	R\$ 166,56	8	
12	19-set-14	Sul	Shampoo	R\$ 201,42	9	
13	16-nov-15	Norte	Pomada Assadura	R\$ 177,40	10	
14	28-jan-14	Sul	Base	R\$ 265,20	10	
15	17-jan-07	Nordeste	Shampoo	R\$ 255,00	10	
16	17-jun-15	Norte	Condicionador	R\$ 263,20	10	
17	5-jan-14	Nordeste	Shampoo	R\$ 126,70	7	
18	1-ago-14	Centro-Oeste	Condicionador	R\$ 288,20	10	
19	12-dez-12	Norte	Pomada Assadura	R\$ 22,56	1	
20	5-jun-17	Sul	Espanja	R\$ 158,76	6	
21	28-ago-06	Centro-Oeste	Espanja	R\$ 106,60	5	
22						

Conclusão

- Este processo de aprendizado da rede neural é geralmente utilizado para vários tipos delas:
 - Multilayer perceptron (densa)
 - Convolutional Neural Network (muito usada em visão computacional)
 - Recurrent Neural Network (muito usada em previsão no tempo)
 - Autoencoders (autoaprendizado, sem rótulos e geração de códigos compactados)