

BACKPROPAGATION

RESUMO TEÓRICO

Soma das entradas (da camada de entrada):

$$a_{lj} = \sum_{i=1}^{n(l-1)} w_{ij} * x_i$$

Soma das entradas (de outros neurônios):

$$a_{lj} = \sum_{i=1}^{n(l-1)} w_{ij} * z_i$$

z é uma função não linear:

$$z_i = \sigma(a_i)$$

Função de ativação (exemplo):

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

Derivada do exemplo de função de ativação:

$$\sigma'(x) = \sigma(x)(1 - \sigma(x))$$

Loss function (exemplo)

$$loss = (y - \hat{y})^2$$

Erro de uma saída:

$$E = loss(y, \hat{y})$$

Objetivo:

$$\min |E|$$

Variação de E para um dado peso u_{ij} :

$$\frac{\partial E}{\partial u_{ij}}$$

Regra da cadeia para a variação entre o peso u_{ij} e o erro E :

$$\frac{\partial E}{\partial u_{ij}} = \frac{\partial E}{\partial a_i} * \frac{\partial a_i}{\partial u_{ij}}$$

$$\frac{\partial a_i}{\partial u_{ij}} = z_j$$

$$\frac{\partial E}{\partial a_i} = \delta_i$$

$$\frac{\partial E}{\partial u_{ij}} = \delta_i * z_j$$

Repassando o delta i:

$$\delta_i = \sum_j \frac{\partial E}{\partial a_j} * \frac{\partial a_j}{\partial a_i}$$

$$\delta_i = \sum_j \delta_j * \frac{\partial a_j}{\partial a_i}$$

$$\frac{\partial a_j}{\partial a_i} = \frac{\partial a_j}{\partial z_j} * \frac{\partial z_j}{\partial a_i}$$

$$\frac{\partial a_j}{\partial a_i} = u_{jk} * \sigma'(a_i)$$

Ou seja:

$$\delta_i = \sigma'(a_i) * \sum_j \delta_j * u_{jk}$$

Excepcionalmente na camada de SAÍDA:

$$z_k = a_k$$

$$\frac{\partial z_k}{\partial a_k} = 1$$

$$\delta_k = \frac{\partial E}{\hat{y}}$$

$$\begin{aligned}\frac{\partial E}{\hat{y}} &= loss'(y, \hat{y}) \\ loss'(y, \hat{y}) &= -2(y - \hat{y}) \\ \delta_k &= -2(y - \hat{y})\end{aligned}$$

Depois de saber como calcular $\frac{\partial E}{\partial u_{ij}}$, calcula-se o novo u_{ij}

$$u_{ij}^* = u_{ij} - \rho * \frac{\partial E}{\partial u_{ij}}$$

No caso do backpropagation para treinamento com batches, o gradiente do peso é acumulado para cada entrada do batch, e depois dividido pelo tamanho do batch. Ou seja, é uma média dos gradientes calculados para aquele peso, por cada item do batch.

B : *Batch size*

$$u_{ij}^* = u_{ij} - \frac{\rho}{B} * \sum_{b=1}^B \frac{\partial E_b}{\partial u_{ij}}$$