1 Activation Functions

1.1 Linear Function

- $f(x) = x$
- $\frac{\partial f}{\partial x} = 1$

1.2 Relu

- $f(x) = \begin{cases} 0, & \text{dla } x < 0 \\ x, & \text{dla } x \geq 0 \end{cases}$
- $\frac{\partial f}{\partial x} = \begin{cases} 0, & \text{dla } x < 0 \\ 1, & \text{dla } x \geq 0 \end{cases}$

1.3 Sigmoid Function

- $f(x) = \frac{1}{1+e^{-x}}$
- $\frac{\partial f}{\partial x} = \frac{e^x}{(1+e^x)^2}$

1.4 Softmax Function

- $f(\mathbf{x})_i = \frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}}$
- $J = S_i(\delta_{ij} - S_j)$

2 Loss Functions

2.1 Mean Squared Error

- $f(x, y) = \frac{1}{2}(y - x)^2$
- $\frac{\partial f}{\partial x} = x - y$

2.2 Logarithmic Loss

- $f(x, y) = -(y \ln x + (1 - y) \ln 1 - x)$
- $\frac{\partial f}{\partial x} = \frac{1-y}{1-x} - \frac{y}{x}$

2.3 Categorical Cross-Entropy (Numerically stable)

- $f(x, y) = -y \ln \left(x + 10^{-100}\right)$
- $\frac{\partial f}{\partial x} = -\frac{y}{x+10^{-100}}$

3 Neuron

3.1 Impulse

$$z = \mathbf{x} \times \mathbf{w} + b$$

x input tensor

w weights vector

b bias

3.2 Activation

$$a = \phi(z(\mathbf{x}))$$

z impulse

ϕ activation function attached to layer

3.3 Weights actualization

$$\mathbf{w} = \mathbf{w} - \boldsymbol{\Delta}\mathbf{w}$$

w weights

Δw actualization vector

3.4 Bias actualization

$$b = b - \Delta b$$

b bias

Δb actualization scalar

4 Feed forward

4.1 For first layer

$$\mathbf{a^0} = \phi(\mathbf{x})$$

x tensor with input data

4.2 For n-th layer

$$\mathbf{a^n} = \phi(\mathbf{a^{n-1}})$$

a^n activation of all neurons in n-th layer collected in one vector

5 Backpropagation

5.1 Last layer error

$$\delta^{\mathbf{out}} = \frac{\partial C(a^{out})}{\partial a^{out}} \odot \frac{\partial \phi(z^{out})}{\partial z^{out}}$$

C loss function

$\odot$ Hadamard product

5.2 N-th layer error

$$\delta^{\mathbf{n}} = \delta^{\mathbf{n+1}} \times \left(\mathbf{w^{n+1}}\right)^{\mathbf{T}} \odot \frac{\partial \phi(z^{n})}{\partial z^{n}}$$

5.3 Update weights vector

$$\Delta \mathbf{w^{n}} = \left(\mathbf{a^{n-1}}\right)^{\mathbf{T}} \times \delta^{\mathbf{n}}$$

5.4 Update bias scalar

$$\Delta b^{n} = \sum_{i=0}^{k} \delta_i^n$$