Text Mining Home Assignment 01

May 6, 2024

Christoph Hlava - 10027820
Dominik Rittner - 10019667

1 Gradient Descent

1.

Derivative calculation is a mathematical approach to examine how strongly a parameter influences the target function (in a specified interval). By this we can approximate how to adjust a weight in a neural network, to minimize the loss.

2.

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

$$\frac{\mathrm{d}\sigma}{\mathrm{d}x} = \frac{\mathrm{d}\sigma}{\mathrm{d}u}\frac{\mathrm{d}u}{\mathrm{d}x}$$

$$\frac{\mathrm{d}u}{\mathrm{d}x} = -e^{-x}$$

$$\frac{\mathrm{d}\sigma}{\mathrm{d}u} = -(u(x))^{-2} = -\frac{1}{(u(x))^2} = -\frac{1}{(1+e^{-x})^2}$$

$$\frac{\mathrm{d}\sigma}{\mathrm{d}x} = -\frac{1}{(1+e^{-x})^2} * (-e^{-x}) = \frac{e^{-x}}{(1+e^{-x})^2}$$

Erweitern:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}x} = \frac{1+e^{-x}-1}{(1+e^{-x})^2}$$

$$\frac{\mathrm{d}\sigma}{\mathrm{d}x} = \frac{1+e^{-x}}{(1+e^{-x})^2} - \frac{1}{(1+e^{-x})^2}$$

$$\frac{\mathrm{d}\sigma}{\mathrm{d}x} = \frac{1}{(1+e^{-x})} - \frac{1}{(1+e^{-x})^2}$$

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

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

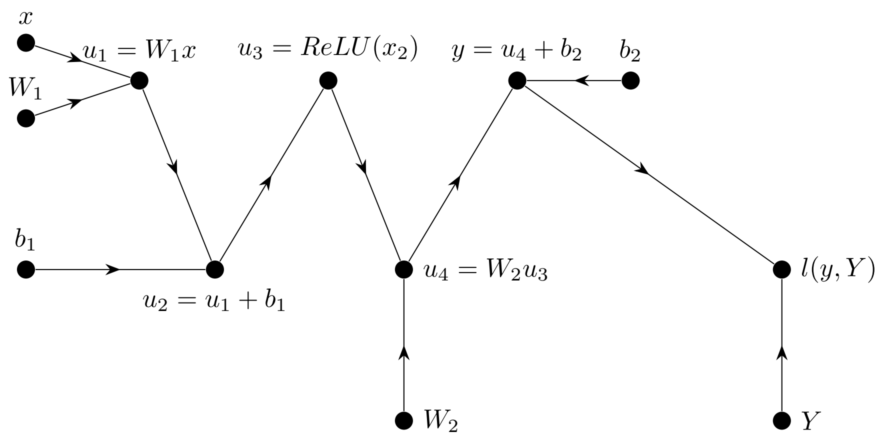
$$\frac{\mathrm{d}\sigma}{\mathrm{d}x} = \sigma(x)(1-\sigma(x))$$

2 Neural Network Basics

Our formula is the following:

$$y = W_2 * ReLU(W_1 * x + b_1) + b_2 \tag{1}$$

First we build the directed graph:



Then we compute the derivates for each Weight and each Bias (W_1, W_2, b_1, b_2).

$$\frac{\mathrm{d}l}{\mathrm{d}W_1} = \frac{\mathrm{d}l}{\mathrm{d}y}\frac{\mathrm{d}y}{\mathrm{d}u_4}\frac{\mathrm{d}u_4}{\mathrm{d}u_3}\frac{\mathrm{d}u_3}{\mathrm{d}u_2}\frac{\mathrm{d}u_2}{\mathrm{d}u_1}\frac{\mathrm{d}u_1}{\mathrm{d}W_1} = \frac{\mathrm{d}l}{\mathrm{d}y} * W_2 * \frac{\mathrm{d}}{\mathrm{d}ReLU}(W_1 * x + b_1) * x$$
$$\frac{\mathrm{d}l}{\mathrm{d}W_2} = \frac{\mathrm{d}l}{\mathrm{d}y}\frac{\mathrm{d}y}{\mathrm{d}u_4}\frac{\mathrm{d}u_4}{\mathrm{d}W_2} = \frac{\mathrm{d}l}{\mathrm{d}y} * ReLU(W_1 * x + b_1)$$
$$\frac{\mathrm{d}l}{\mathrm{d}b_1} = \frac{\mathrm{d}l}{\mathrm{d}y}\frac{\mathrm{d}y}{\mathrm{d}u_4}\frac{\mathrm{d}u_4}{\mathrm{d}u_3}\frac{\mathrm{d}u_3}{\mathrm{d}u_2}\frac{\mathrm{d}u_2}{\mathrm{d}b_1} = \frac{\mathrm{d}l}{\mathrm{d}y} * W_2 * \frac{\mathrm{d}}{\mathrm{d}ReLU}(W_1 * x + b_1)$$
$$\frac{\mathrm{d}l}{\mathrm{d}b_2} = \frac{\mathrm{d}l}{\mathrm{d}y}\frac{\mathrm{d}y}{\mathrm{d}b_2} = \frac{\mathrm{d}l}{\mathrm{d}y}$$

3 Embeddings

1.

Embeddings are a way to map words to vectors of numbers, such that words that are semantically similar also have similar location in the corresponding vectorspace (nearer together the more they encode similar concepts). Therefore we can calculate the similarity between words not only on the syntactical level but with embeddings also on the level of meaning, which is highly useful regarding AI tasks.