2 Understanding word2vec

2.a

YOUR ANSWER HERE

2.b

YOUR ANSWER HERE

2.c

YOUR ANSWER HERE

2.d

YOUR ANSWER HERE

2.e

YOUR ANSWER HERE

2.f.i

YOUR ANSWER HERE

2.f.ii

YOUR ANSWER HERE

2.f.iii

YOUR ANSWER HERE

3 Implementing word2vec

4 Paraphrase Detection (theoretical)

4.a

We note that because of the relu all of the elements in both $relu(x_1)$ and $relu(x_2)$ are positive, because of that the dot product $relu(x_1)^T relu(x_2)$ is composed of a sum of positive arguments, hence

$$relu(x_1)^T relu(x_2) > 0.5$$

Now, we note that for every positive scalar i, it holds that $\sigma(i) > 0.5$. Hence, we arrive at the conclusion that the probability for every sample is

$$p(Paraphrases|x_1, x_2) = \sigma(relu(x_1)^T relu(x_2)) > 0.5$$

Assuming that our decision threshold is 0.5 it holds that we predict True for every pair. Hence, we predict correctly for True pairs and wrong for False pairs. Our accuracy is

$$\frac{1}{1+3} = 25\%$$

4.b

5 Paraphrase Detection (theoretical)

5.a

We note that because of the relu all of the elements in both $relu(x_1)$ and $relu(x_2)$ are positive, because of that the dot product $relu(x_1)^T relu(x_2)$ is composed of a sum of positive arguments, hence

$$relu(x_1)^T relu(x_2) > 0.5$$

Now, we note that for every positive scalar i, it holds that $\sigma(i) > 0.5$. Hence, we arrive at the conclusion that the probability for every sample is

$$p(Paraphrases|x_1, x_2) = \sigma(relu(x_1)^T relu(x_2)) > 0.5$$

Assuming that our decision threshold is 0.5 it holds that we predict True for every pair. Hence, we predict correctly for True pairs and wrong for False pairs. Our accuracy is

$$\frac{1}{1+3} = 25\%$$

5.b

If we have to keep the relu functions at the last layer, then we can map the sigmoid into $[0, 1]$ by adding this computation

$$2 * (\sigma(relu(x_1)^T relu(x_2)) - 0.5)$$

(This is equivalent to changing the threshold from 0.5 to 0.75.)