Aufgabe 1

a) Why do randomly initialized, untrained sentence embeddings produce usable results?

The randomly initialized sentence embeddings are used as features for a collection of downstream tasks.

Sentence representations are randomly computed from pre-trained word embeddings

b) How does the Bag of random embedding projections (BOREP) approach create a sentence representation?

First, should a matrix W, where D is the dimension of the projection and d is the dimension of input word embedding, randomly initialize and sampled uniformly.

Second, the sentence representation is obtained through pooling function.

Aufgabe 2

What are the effects of the four parameters size, window, negative, and cbow?

-size <int>

Set size of word vectors; default is 100

-window <int>

Set max skip length between words; default is 5

-negative <int>

Number of negative examples; default is 5, common values are 3 - 10 (0 = not used)

-cbow <int>

Use the continuous bag of words model; default is 1 (use 0 for skip-gram model)

use the distance tool in the word2vec directory and compute:

man:

love 0.952634

my 0.936406

woman 0.933501

saying 0.915894

person 0.913718

woman:

child 0.978644

got 0.951689

Zeus 0.948846

Achilles 0.940926

Bakr 0.939643

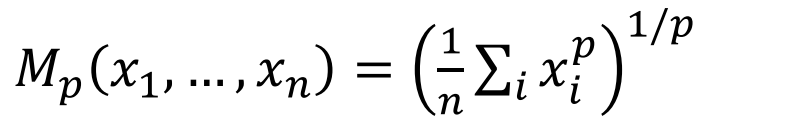
Aufgabe 3.1

3) Which dimensionality does the average word embedding have? Which dimensionality does the concatenated power mean word embedding have?

len(reviews)\*300

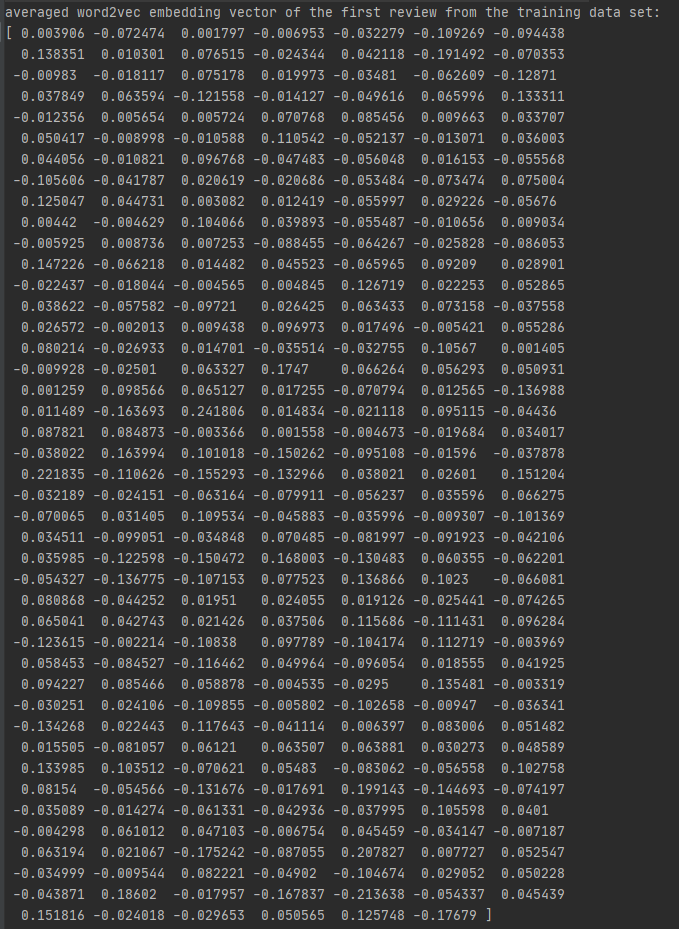
len(reviews)\*1200

4) Which problem arises when you want to use a power mean with negative parameter p such as the harmonic mean?

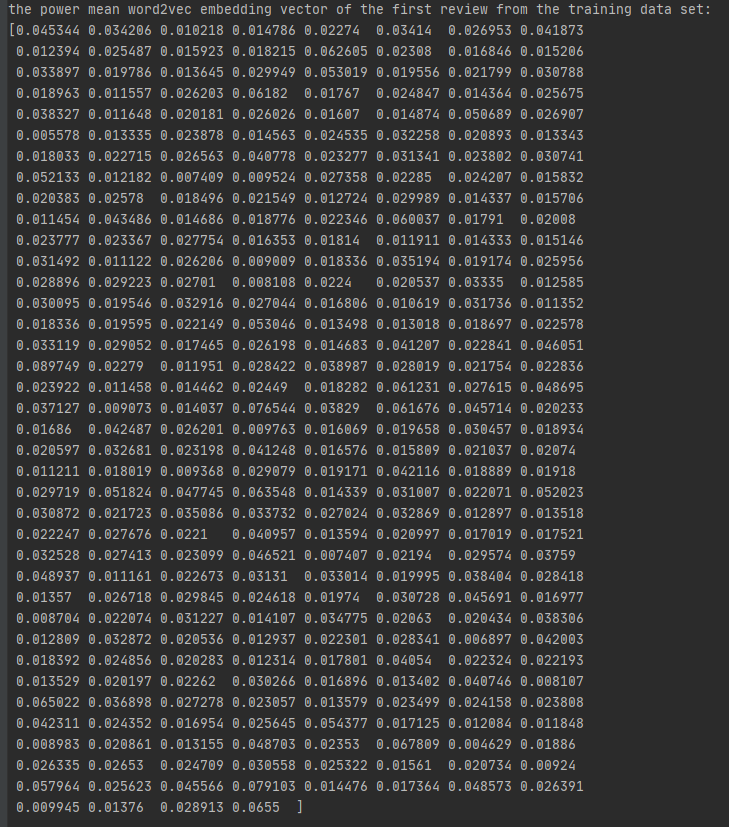


If use with negative parameter p, OOC Vector (Zeros) could cause zero-cross problem.

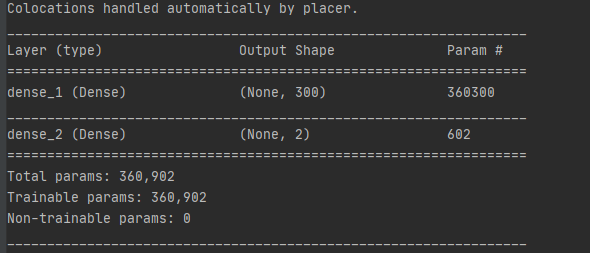
averaged word2vec embedding vector



the power mean and the averaged word2vec embedding vector



Aufgabe 3.2

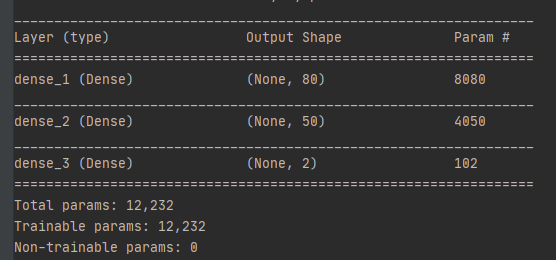






Result from 3. Homework





Network from 3. Homework has 3 Layers, which mean the more complexity in model. It could get better result. Moreover, the activation function in 3. Homework is ReLu, which has a huge acceleration effect on the convergence of stochastic gradient descent.