Home Exercise 04

Viet-Hoa Nguyen
Martikelnummer: 2581721

20-00-0947 Deep Learning für Natural Language Processing
Technische Universität Darmstadt

May 18, 2020

Problem 1: Mandatory Paper

One major limitation of popular word vector models is that they can cannot deal with sense disambiguation of polysemous words. The word vector model trained on a certain context cannot perform when it is applied on another context.

Problem 2: Softmax

2.1 Compare to other activation functions

Similar to sigmoid and tanh, the softmax funtion squashes numerical values into a small interval [0,1], however its out put also equivalent to a categorical probability distribution, which tells us the probability that any of the classes are true.

2.2 Mathematical Properties

1.

2. It is easy to see that

$$\begin{aligned} y_1 + y_2 + y_3 + ... + y_n &= \frac{exp(z_1/T)+exp(z_2/T)+exp(z_3/T)+...+exp(z_n/T)}{\sum_n^i exp(z_i/T)} \\ &= \frac{\sum_n^i exp(z_i/T)}{\sum_n^i exp(z_i/T)} \\ &= 1 \end{aligned} \tag{1}$$

3.
When $T \to 0$, the probability of the class with the highest expectation will tend to over-weight other classes
When $T \to \infty$, the probability distribution over classes will be softer, all classes will have almost the same probability

Problem 3: Word similarity with word embedding

The Pearson's coefficient of -0.4361 implies a moderate negative relationship between the human prediction and the prediction based on euclidean distance, as the human prediction increases the euclidean distance based prediction decreases and vice versa. This might indicates a bad performance of the similarity algorithm, since it does not correspond well to the human prediction.