

## NLP HW 2 - Dry Part

Daniel Shemesh 211388251 & Eylon Efraim 207427428

3. a:

Unlike in the standard Word2Vec model, here every word in the model is associated with a single vector of dimension 500. The coordinates of this vector are the parameters of the model, which means there are  $500 \times 500,000 = 250,000,000$  parameters.

3. b:

The model is not computationally feasible. In models in which the parameters are shared between all terms, we would only have had to calculate 500 parameters, which would have been feasible. However, that is not the case in this model, and calculating 250 million parameters (in a single layer, unlike the many parameters of some deep learned models) using SGD will take far too much time.

4:

Without negative sampling, the model could have been biased towards increasing  $p(D=1 | x, y, \theta)$  regardless of the accuracy of the prediction, as no loss would have occurred from assigning the highest probability possible to all word pairs. Introducing negative sampling eliminates this bias, this bias is eliminated, as the model will now have to maximize the probability on some word pairs and minimize it on others, preventing the situation previously described.

5:

We would expect words with similar Word2Vec vectors to be highly similar, rather than highly associated. The Word2Vec algorithm attempts to match words to vectors based on the language model provided. As such, similar vectors will be matched to words that may be represented similarly w.r.t. the language model we use.

For example, in this exercise we used the bigram language model, which means a pair of words will have similar vectors if we could switch one for the other in our language model and receive results that still make sense (and thus may have showed up in the training vocabulary). To use the provided examples, "fast car" may be replaced with "fast vehicle", and thus car and vehicle are likely to have similar representative vectors. However, "football field" cannot be replaced with "Maradona field", and as such football and Maradona are unlikely to be represented by similar vectors.