

Name:

Student ID number:

- (1) Given co-occurrence probabilities $P(c|t)$ between target words t and context words c , the crucial insight in creating GloVe word embeddings (Pennington et al. 2014) is: *(choose one)*
- (a) using the difference between co-occurrence probabilities $P(c|t_1) - P(c|t_2)$
 - (b) using the ratio of co-occurrence probabilities $P(c|t_1)/P(c|t_2)$
 - (c) using co-occurrence probabilities $P(c|t)$ instead of counts of c and t .
- (2) Is the following statement true or false: In Skip-gram with negative sampling, a classifier is trained on a multi-class prediction task.
- (3) Suppose that you are using an RNN for POS tagging and that the output at each time step is a distribution over the POS tagset, as generated by a softmax layer. Is this output sufficient to determine the optimal POS tag sequence for the whole sentence? Say why or why not, and make some suggestions for improvement if necessary. (3-5 sentences)
- (4) What is the difference between “similarity” and “relatedness” of words? Illustrate using examples.
- (5) Describe in 1-2 sentences one improvement made in Abnar et al. (2018) “Experiential, Distributional and Dependency based Word Embeddings have Complementary Roles in Decoding Brain Activity” over the original experiments in Mitchell et al. (2008) “Predicting Human Brain Activity Associated with the Meanings of Nouns”?
- (6) Name at least three possible applications of Recurrent Neural Networks for language processing (1.5 points). For one of them, describe the input and the output of the network at one given timestep (2 points).

1 points

1 points

3 points

1 points

2 points

3.5 points