Notes: word2vec
(and related distributed embeddings)

References:

1. high level overview: https://blog.acolyer.org/2016/04/21/the-amazing-power-of-word-vectors/
2. details with math: http://arxiv.org/abs/1402.3722v1

Related methods:

1. fastText: https://fasttext.cc
2. doc2vec: https://arxiv.org/abs/1405.4053

Applications:

1. translating with word2vec https://arxiv.org/abs/1309.4168
2. gender bias: http://wordbias.umiacs.umd.edu/
3. racial bias: https://www.pnas.org/content/115/16/E3635/tab-figures-data
4. histwords: https://nlp.stanford.edu/projects/histwords
5. temporal word analogies: https://www.aclweb.org/anthology/P17-2071
6. political words: https://arxiv.org/abs/1711.0560

1 Word2Vec

Problem 1. What is a word analogy?

Problem 2. What is the algorithm for solving word analogies?

Problem 3. Why can 1-hot encoded vectors not be used for solving word analogies?

Problem 4. Describe the following two word2vec learning problems:

1. continuous bag of words

2. skipgram

Problem 5. We saw in our previous notes that the softmax cross entryopy is the standard loss for multivariate classification problems. The word2vec models do not use this loss function, however. Instead, they use something called *negative sampling*.

1. Why can word2vec models not use the standard cross entropy loss function for classification?

2. Describe the negative sampling loss function.

Problem 6. What are the hyperparameters for learning a word2vec model?

1. choice of input dataset

2. how should words be tokenized?

3. context window size (c)

4. which word2vec model should we pick?

5. vocabulary size (v)

6. number of dimension (d)

7. number of data points (m)

8. learning rate (η)

9. number of negative samples (s)

Problem 7. What is the relationship between neural networks/deep learning and word2vec?

2 Related models

Problem 8. What is the fastText model?

Problem 9. What is the doc2text model?

3 Applications

Problem 10. Other than word analogies, what problems can word vectors solve?