

Tutorial 7

Question 1 (Word Vectors)

You are working with a small corpus of text from a children's storybook. Here's a sample sentence from the book: "The cat sat on the mat." You decide to use the Skip-Gram model of Word2Vec to learn word vectors for this text.

- 1) Given the target word "sat", and a window size of 1, identify the (target, context) word pairs you would use to train the Skip-Gram model.
- 2) Write the objective function to be maximized with respect to this target word and the context window given in the previous question, ignoring negative sampling.
- 3) Write the objective function to be maximized with respect to this target word "sat" and context window 1, with negative sampling.
- 4) Suppose after training, you obtain the following word vectors. Which word is most similar to "cat" based on cosine similarity? Are these vectors well learned?
- 5) Explain whether cosine similarity is used in the training objective function, and discuss why (open question).

Word	Vector Representation	L2-norm
cat	[0.5, 1.0, -0.3]	1.16
mat	[0.6, 0.8, -0.2]	1.02
sat	[-0.1, 0.9, 0.3]	0.95
hat	[0.6, 1.0, -0.4]	1.23
dog	[-1.0, -0.5, 0.8]	1.46

Question 2 (Word Vector Evaluations)

Assume we have simplified vectors for the words: happy, joyful, sad, angry, calm, peaceful, tense, and frustrated. Their vectors are:

- "happy" = [6, 4]

- “joyful” = [5, 5]
- “sad” = [1, 2]
- “angry” = [2, 6]
- “calm” = [4, 1]
- “peaceful” = [4, 2]
- “tense” = [1, 5]
- “Frustrated” = [2, 5]

- 1) Use vector arithmetic to find a word that should be as opposite to “happy” as “angry” is to “calm”.
- 2) Use vector arithmetic to find a word that is to “happy” as “peaceful” is to “calm”.

Question 3 (Language Modeling)

Imagine you're working with a simple neural language model designed to predict the next word in a sentence. The model has been trained on a small dataset of children's stories. Here's a sample from the corpus: "Jack loves climbing trees".

- 1) Give the mathematical formulation on the probability of the above sentence.
- 2) If Markov Assumption with window size 1 is used, what is the revised probability of the above sentence?
- 3) Describe the transformation process of the input words "Jack loves" into a probability distribution over the next possible word in the vocabulary. Outline the key stages in this transformation.

Coding exercises (for your own interest)

- <https://colab.research.google.com/drive/164dB-Vemzwav1ffqDDVNtx7Y5VtcmQh?usp=sharing>

More references

- Word2vec paper: <https://arxiv.org/abs/1301.3781>,
<https://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf>
- Glove paper: <https://nlp.stanford.edu/pubs/glove.pdf>