CSE 4392 Special Topic: Natural Language Processing

Homework 5 - Spring 2025

Due Date: Feb 24, 2025, 11:59 p.m. Central Time

Problem 1 - 30%

Derive the gradient of the negative sampling objective function with respect to the parameters $(\mathbf{u}_t \text{ and } \mathbf{v}_t)$ used in Skip-Gram Word2Vec.

$$y = -log(\sigma(\mathbf{u}_t \cdot \mathbf{v}_c)) - \sum_{i=1}^{K} \mathbb{E}_{j \sim P(w)} log(\sigma(-\mathbf{u}_t \cdot \mathbf{v}_j))$$

Any context word vector v_t might be the actual context vector t = c or one of those negatively sampled t = j (where $c \neq j$) or none of these. The gradient w.r.t v_t should be derived for each of these three cases.

Before solving the problem, write all the differentiation rules you expect to use except for the most trival ones and solve in latex or clear handwriting.

Problem 2 - 70%

In this problem, you are required to re-implement feature extraction for the E-commerce dataset, but this time utilize advanced word embedding techniques instead of traditional Bag of Words with Logistic Linear Regression. Measure the impact on performance metric.

- For each of Word2Vec, GloVe, and FastText, you will define function(s) in a new file **embeddings.py** to be able to extract the embeddings of given word(s) for any of the methods above.
- Moreover, you will define another function that given a document or paragraph of text, returns a single dense vector representation by pooling its word vectors
- Define a function that computes a bag-of-words vector given any document
- By this, including your feature extraction method from last time, there are five feature extraction methods in total.
- Ensure that the your feature vectors only depend on the training set or are pretrained.
- Make a table to compare the five methods together in terms of accuracy, F1 score, precision and recall (macro scores).
- Provide insights into the comparison and your justification to their performance ranking.
- Submit your Python files and PDF analogues as usual. No zipping required.
- As a bonus you can perform nonlinear dimensionality reduction and visualize the different features