Homework 2

Learning Goals

- Understand skip-gram with negative sampling in more detail

  - Compute various derivatives in order to get the gradient of the loss with respect to the parameters

- Learn how to translate math into code, for

  - The model forward pass

  - Gradient computations

Understanding Word2Vec

- Count parameters

- Understand sigmoid, and the role it plays in SGNS
  - Compute its derivative

- Compute the gradient of L_{CE} with respect to parameters
  - Done in stages
  - Uses:
    - Logarithm rules
    - Derivative of logarithm
    - Addition / product / chain rule for derivatives

Implementing Word2Vec

- SGNS will be implemented in raw numpy

- We provide the entire training loop, but various methods that are called need to be filled in

  - Data processing: generating positive and negative samples

  - Model computation: implement the $P(1 \mid w, c; \theta)$ computation

  - Gradient computation: compute ∇L_{CE} w/r/t each of the relevant parameters

Training Word Vectors

- Finally, you will train word vectors by iterating through the SST training set

- Plot the vectors of a list of words, using PCA for dimensionality reduction
  - We provide all of this code!

- Describe any trends you see in the embeddings