

Mini-project 1: Training a CBoW Model

Janmesh - u1367465

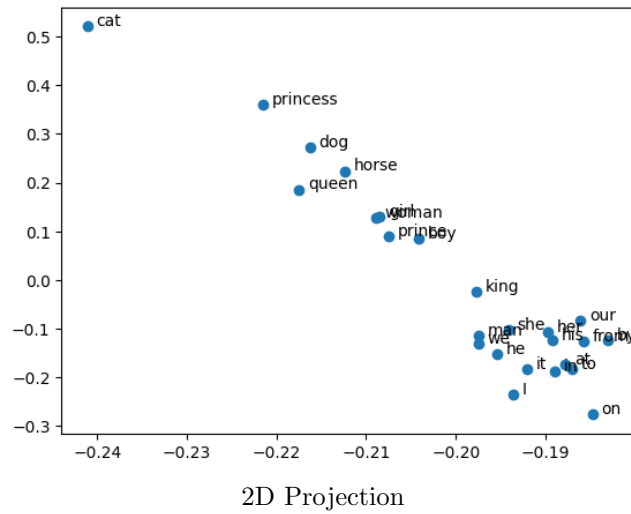
September 18, 2023

Report

1. Best learning rate = 0.001 with development loss = 6.1571.
2. (a) Cosine Similarity closest pair -
 - i. [cat, tiger] = 0.9065176 & [plane, human] = 0.86195904 ; [cat, tiger] is closer which is expected.
 - ii. [my, mine] = 0.9056487 & [happy, human] = 0.9235775 ; [my, mine] should have been more closer.
 - iii. [happy, cat] = 0.8628401 & [king, princess] = 0.86253476 ; [king, princess] should have been more closer.
 - iv. [ball, racket] = 0.88345003 & [good, ugly] = 0.8391444 ; [ball, racket] is closer which is expected.
 - v. [cat, racket] = 0.92847943 & [good, bad] = 0.9054469 ; [good, bad] should have been more closer.
- (b) Word Analogy - (Top 5 matches provided)
 - i. king:queen, man:['friend', 'deep', 'person', 'beauty', 'always']
 - ii. king:queen, prince:['irène', 'äfven', 'tid', 'cœurs', 'bride']
 - iii. king:man, queen:['friend', 'deep', 'person', 'beauty', 'always']
 - iv. woman:man, princess:['prince', 'tout', 'discovery', 'author', 'shogun']
 - v. prince:princess, man:['girl', 'woman', 'gentleman', 'person', 'fellow']
3. Generated Examples -
 - (a) Word Similarity Test -
 - i. [she, her] = 0.9065176 & [he, chair] = 0.8290193 ; [she, her] is closer.
 - ii. [europe, france] = 0.9272915 & [fish, read] = 0.8765209 ; [europe, france] is closer.
 - iii. [england, british] = 0.90075195 & [cow, philosophy] = 0.8600543 ; [england, british] is closer.
 - (b) Word Analogy - (Top 5 matches)
 - i. england:english, france:['french', 'themselves', 'popular', 'altogether', 'nevertheless']
 - ii. london:city, england:['country', 'state', 'land', 'night', 'law']
 - iii. she:her, he:['his', 'him', 'the', 'them', 'and']
4. Evaluation Results -
 - (a) Word Similarity Test Pearson Correlation: 0.0851135200986339
 - (b) Accuracy on Analogy Test: 0.03986902927580894 (3.98%).
5. In the paper given the loss used is negative log likelihood loss over the classes whereas we have used cross entropy loss in this CBoW project, the paper also uses hierarchical softmax which is not used in the project. Most we assign a label corresponding to each word in this project, where as this paper uses character n-gram, i.e. it divides the word into different list of characters and the final word is represented by the sum of this set of n-character list vectors.

Extra Credit

1. 2-D projection of the embeddings -



2D projection may be misleading as projecting a a high dimension data into low dimension does not preserves the structure of data and is prone to losing some important information. When talking specifically about PCA or SVD they assume the data to be linear and don't work that well with non linear data.