A Neural Probabilistic Language Model

1 Main Idea

This is the seminal paper on neural language modeling that first proposed learning distributed representations of words. There is an obvious distinction made for predictions in a discrete vocabulary space vs. predictions in a continuous space, and the solution proposed is to have real-valued word feature vectors that are learnt along with the joint probability function of their occurrance in sequences in the corpus.

2 Method

- The probability function is expressed as the product of the conditional probabilities of the next word given the current word.

3 Observations

-