Word2Vec paper

* Word2Vec maps each word to a vector of real numbers in such a way that words used in similar contexts end up close together in that space.
* Word2Vec proposes two models to learn these word vectors:

1. **Continuous Bag-of-Words (CBOW)**: Predict the target word based on its context words It takes the average of the context word vectors and predicts the center word.

2. **Skip-Gram**: predict context words from the target word.

* they use two techniques to make training efficient:

1. **Hierarchical Softmax** : Uses a binary tree (like a Huffman tree) ,Frequent words are near the root → faster predictions.

2. **Negative Sampling** : Instead of updating all weights for all words, it only updates a few "negative" examples + the correct one.