Week 2: Word2Vec Deep Dive

Understanding CBOW, Skip-gram, and Negative Sampling
Pre-Lab Exercise (No Programming Required)

NLP Course 2025 - Student Version

Time: 30-40 minutes
Objective: Master the core mechanisms of Word2Vec by understanding what goes in (context), how it's processed (method), and what comes out (prediction).

Part 1: Understanding Context Windows (10 minutes)

Context Window Discovery

Consider this sentence: **"The quick brown fox jumps over the lazy dog"**
With a context window size of 2 (two words on each side), let's analyze the word "fox":
Questions:

1. Circle the context words for "fox" (window size = 2):

   `The` `quick` `brown` `[fox]` `jumps` `over` `the` `lazy` `dog`

2. What are the context words? _____

3. For the word "jumps", what would be the context words (window size = 2)?
   Context words: _____

4. **Key Question:** In a context window approach:
   - What is the **CONTEXT**? _____
   - What is the **METHOD**? _____
   - What is the **PREDICTION**? _____

Part 2: CBOW - Continuous Bag of Words (10 minutes)

CBOW Principle

CBOW predicts a target word given its surrounding context words. Think of it as a "fill in the blank" task.

CBOW in Action

Given the sentence: "The cat sat on the _____ mat"

Task 1: Manual CBOW

1. Context words given: ["cat", "sat", "on", "the", "mat"]

   What word would you predict for the blank? _____

2. Now consider: "I love to _____ pizza for dinner"

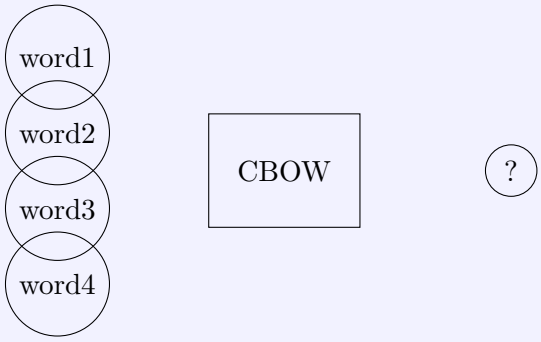
   Context words: ["I", "love", "to", "pizza"]

   Your prediction: _____

Task 2: CBOW Analysis Complete the following table for CBOW:

Aspect	**Your Answer**
INPUT (Context)	_____
METHOD (How does it work?)	_____ _____
OUTPUT/PREDICTION	_____

Visual Representation: Draw arrows showing the flow:



Draw arrows to show data flow

Part 3: Skip-gram - Predicting Context (10 minutes)

Skip-gram Principle

Skip-gram does the opposite of CBOW: given a center word, predict the surrounding context words.

Skip-gram in Action

Given the center word "**coffee**" in: "I drink coffee every morning"
Task 1: Manual Skip-gram

1. What context words would you expect around "coffee"? List 4 likely words:
   - _____
   - _____
   - _____
   - _____
2. Given center word "**king**", what context words might appear?
   - _____
   - _____

Task 2: Skip-gram Analysis Complete the following table for Skip-gram:

Aspect	**Your Answer**
INPUT (Context)	_____
METHOD (How does it work?)	_____ _____
OUTPUT/PREDICTION	_____

Task 3: CBOW vs Skip-gram

1. Which approach has MORE training examples from the same sentence?

   _____
2. Why?

Part 4: Negative Sampling - Making Training Efficient (10 minutes)

Why Negative Sampling?

Training Word2Vec on large vocabularies (e.g., 50,000 words) is computationally expensive. Negative sampling speeds this up by only updating a small subset of weights.

Understanding Negative Sampling

Task 1: Positive vs Negative Pairs
Given the sentence: "The dog barked loudly"
With center word "dog" and context word "barked":

1. This is a _____ pair (positive/negative) because _____
2. Now consider "dog" and "elephant". This is a _____ pair because _____
3. Create 3 negative pairs for center word "dog":
   - (dog, _____)
   - (dog, _____)
   - (dog, _____)

Task 2: The Sampling Process
Instead of updating all 50,000 word weights, negative sampling:

1. Updates the positive pair: (dog, barked) → predict "1" (real pair)
2. Randomly samples k negative words (e.g., k=5)
3. Updates these negative pairs: (dog, random_word) → predict "0" (fake pair)

Complete this table for Negative Sampling:

Aspect	**Your Answer**
INPUT (Context)	_____ _____
METHOD (How does it work?)	_____ _____
OUTPUT/PREDICTION	_____

Task 3: Efficiency Analysis

1. Without negative sampling, how many weights update for each training example?
   _____
2. With negative sampling (k=5), how many weights update?
   _____
3. What is the speedup factor for a 50,000-word vocabulary?
   _____

Hint

Negative sampling converts the problem from multi-class classification (50,000 classes) to binary classification (real vs fake pairs).

Part 5: Putting It All Together (5 minutes)

Synthesis

Complete this comparison table:

Aspect	CBOW	Skip-gram	Negative Sampling
Input	Context words		
Method	Average/combine context		Binary classification
Output		Context words	
Best for	Frequent words		
Training Speed	Faster		Makes both faster

Reflection Questions:

1. Why might Skip-gram work better for rare words than CBOW?

2. How does negative sampling change what the model is learning to predict?

3. If you wanted to find similar words to "doctor", which approach (CBOW or Skip-gram) would likely give better results? Why?

End of Student Exercise