Week 2: Teaching Computers to Understand Word Relationships

Discovering How Machines Learn Language
Pre-Lab Exercise (No Programming Required)

NLP Course 2025 - Student Version

Time: 30-40 minutes
Objective: Discover three fundamental ways computers can learn word meanings from text.

Part 1: How Words Keep Company (8 minutes)

Word Prediction Game

Task 1: Fill in the blanks with the most likely word:

a) The cat sat on the _____

b) I drink _____ every morning

c) The _____ barked loudly at the mailman

d) She wore a beautiful _____ to the wedding

Task 2: Reflection

1. How did you know what words to fill in? What clues did you use?

2. List the words that helped you guess the answer for (a):
   Helper words: _____

3. **Important Discovery:** You used the surrounding words to predict the missing word.
   In your own words, explain why surrounding words help:

Think About It

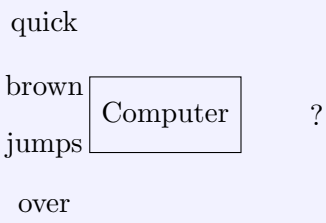
If humans can guess words from their surroundings, can we teach computers to do the same?

Part 2: Two Ways to Learn Words (10 minutes)

Method A: Many Words Predict One

Imagine you're teaching a computer to fill in blanks like you just did.
Scenario: "The quick brown _____ jumps over"
Task 1: Design the Learning

1. What information would you give the computer as INPUT?
   INPUT: _____
2. What should the computer learn to OUTPUT?
   OUTPUT: _____
3. Draw how this works (use arrows):



Draw arrows to show the flow

Task 2: Practice Examples
Using Method A (surrounding words → missing word), predict:

dog, walked, the, park → _____

ate, pizza, for, lunch → _____

Discovery Moment

Congratulations! You've just invented an approach that computer scientists call "CBOW" (Continuous Bag of Words). It uses context words to predict a center word - just like you did!

Method B: One Word Predicts Many

Now let's flip it around!
Scenario: Given the word "coffee", predict what words often appear near it.
Task 1: Word Association

1. Given "coffee", list 4 words that often appear nearby:
   - _____
   - _____
   - _____
   - _____
2. This is the OPPOSITE of Method A. Complete:
   - Method A: Many words → _____
   - Method B: One word → _____
3. Which method creates MORE training examples from one sentence? Why?

Discovery Moment

You've discovered "Skip-gram"! It takes one word and predicts the surrounding context - the reverse of CBOW.

Part 3: Making It Faster - A Clever Trick (10 minutes)

The Speed Problem

The Challenge: English has about 50,000 common words. Every time the computer learns, should it:

- Option A: Check all 50,000 words to find the right one?
- Option B: Check just 5-10 words?

Obviously, Option B is faster! But how can we do this?

Task 1: Real or Fake?

Instead of finding THE right word out of 50,000, let's play a simpler game:
Given word pairs, decide if they're REAL (actually appear together) or FAKE (random pairing):

Word 1	Word 2	Real or Fake?
coffee	drink	_____
coffee	elephant	_____
dog	barked	_____
dog	galaxy	_____
queen	king	_____
queen	bicycle	_____

Task 2: Understanding the Trick

1. Instead of asking "Which of 50,000 words is correct?", we now ask:
   New question: _____
2. For each real pair, how many fake pairs should we create for good learning?
   _____ fake pairs
3. If we use 5 fake pairs + 1 real pair, we only update 6 words instead of 50,000.
   What's the speedup? _____ times faster!

Discovery Moment

This clever trick is called "Negative Sampling"! The real pairs are "positive samples" and the fake pairs are "negative samples". It makes training about 8,000 times faster!

Part 4: Comparing Your Discoveries (7 minutes)

Putting It All Together

Now that you understand all three approaches, let's compare them:

Aspect	**Method A (CBOW)**	**Method B (Skip-gram)**	**The Speed Trick (Negative Sampling)**
What goes IN? (Input)	Multiple context words		
How does it work? (Method)	Combine all context words		
What comes OUT? (Output)		Multiple context words	
Example	[cat, sat, the] → on		(coffee, drink) → Real
Best for	Common words		Making training faster

Critical Thinking Questions:

1. Why might Method B (Skip-gram) work better for rare words than Method A (CBOW)?

   Hint: Think about how many training examples each method creates.

2. The Speed Trick changes the question from "which word?" to "real or fake?"

   Why is this simpler for a computer?

3. If you wanted to find words similar to "doctor", which method would you choose? Why?

Part 5: Real-World Impact (5 minutes)

Reflection

You've just discovered three fundamental techniques that power modern AI language models!

1. These techniques help computers understand that "king" and "queen" are related.

   How do you think the computer learns this relationship?

2. Word prediction is used in your phone's keyboard.

   Which method (A or B) do you think works better for predicting your next word? Why?

3. Before these methods, computers needed humans to manually teach every word relationship.

   What's the advantage of learning from context automatically?

Summary: You discovered three key concepts today:
CBOW (many→one), **Skip-gram** (one→many), and **Negative Sampling** (real vs fake).
These are the foundations of modern language AI!