Word Embeddings in 3D: Post-Class Learning Verification

Checking Your Understanding After the Interactive Lab

From Words to Vectors: Can You Apply What You Learned?

NLP Course 2025 - BSc Level Assessment

Time Required: 45-60 minutes
Purpose: Verify and deepen your understanding of word embeddings after completing the interactive notebook
Format: No coding required - focus on concepts, visualization, and application

Checkpoint

Before Starting: You should have completed the "word_embeddings_3d_bsc.ipynb" notebook. This handout will test your understanding of:

- Why words need to be vectors
- How Word2Vec learns from context
- Word similarity and clustering
- Word arithmetic
- Applications of embeddings

Part A: Conceptual Understanding (20 minutes)

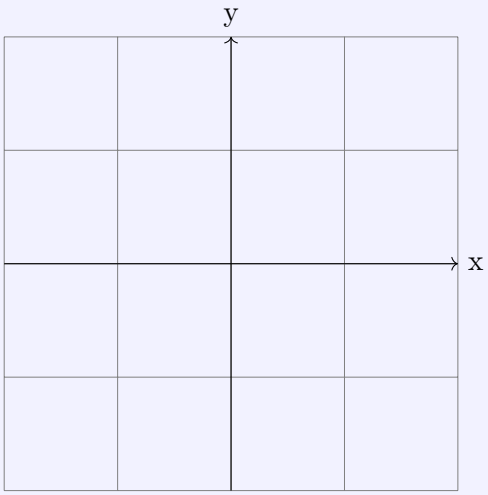
A1: The Embedding Concept (5 minutes)

Why Vectors?

Question 1: Explain in your own words why computers need word embeddings instead of just treating words as text strings.

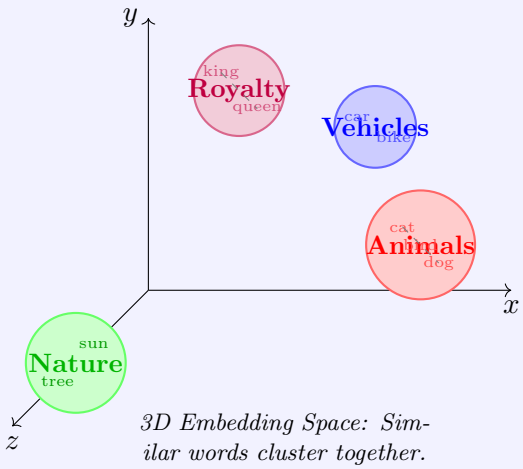
Question 2: Draw simple 2D vectors for these words showing their relationships:

- cat, dog, car
- Show which two should be closer together and why



Draw vectors for cat, dog, and car

Visualization: Real 3D Word Embedding Space



3D Embedding Space: Similar words cluster together.
Distance between clusters = semantic difference

Question 3: Circle the TRUE statements:

☐ Embeddings capture word meaning as numbers

☐ Similar words have similar vectors

A2: Context Windows (5 minutes)

Understanding Context

Given the sentence: **"The quick brown fox jumps over the lazy dog"**
Task 1: If we're training on the word "fox" with window size = 2, circle all context words:

The quick brown $\boxed{fox}$ jumps over the lazy dog

Task 2: How would changing window size affect learning?

Window = 1	_____
Window = 5	_____
Window = 10	_____

Task 3: Which window size would be better for:

- Learning syntax (grammar): _____
- Learning topic/theme: _____

A3: Dimensions and Quality (5 minutes)

Dimension Trade-offs

Scenario: You're choosing embedding dimensions for different applications.
Task 1: Match the dimension size to the use case:

Application	Suggested Dimensions
Simple word similarity	
Complex language model	
Visualization in 3D	
Mobile app (limited memory)	
Research with huge vocabulary	

Options: 3, 10-50, 100-300, 500+, exactly 3
Task 2: Explain the trade-off:

- More dimensions = _____
- Fewer dimensions = _____

A4: Quick Concept Check (5 minutes)

Think About It

Rate your understanding (1 = confused, 5 = confident):

- ☐ Words as vectors [1] [2] [3] [4] [5]
- ☐ Context windows [1] [2] [3] [4] [5]
- ☐ Training process [1] [2] [3] [4] [5]
- ☐ Similarity measurement [1] [2] [3] [4] [5]

Part B: Practical Application (15 minutes)

B1: Word Similarity Exercise (5 minutes)

Computing Similarity

Given these simplified 3D embeddings:

- king = [0.8, 0.2, 0.5]
- queen = [0.7, 0.3, 0.6]
- car = [0.1, 0.9, 0.2]

Task 1: Which pair is more similar? (Use rough estimation)

- king & queen: Distance ≈ _____
- king & car: Distance ≈ _____
- More similar pair: _____

Task 2: Rank these word pairs by expected similarity (1 = most similar):

☐ cat - dog

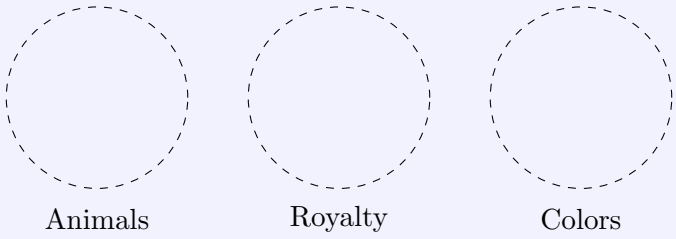
☐ king - queen

☐ happy - sad

☐ computer - laptop

☐ run - blue

Task 3: Draw approximate clusters:



Place these words: cat, king, red, dog, queen, blue, prince, green

B2: Word Arithmetic Magic (5 minutes)

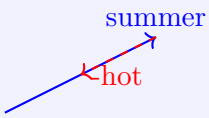
Vector Math with Words

Task 1: Complete these analogies:

- king - man + woman = _____
- paris - france + germany = _____
- cat - kitten + puppy = _____

Task 2: Create your own word equation:
_____ - _____ + _____ = _____

Task 3: Draw the vector arithmetic for: "summer - hot + cold"



Complete the diagram

Task 4: Explain why word arithmetic works:

B3: 3D Visualization Interpretation (5 minutes)

Reading 3D Plots

Imagine you see a 3D plot where:

- "love" and "hate" are far apart
- "cat" and "dog" are close together
- "king" and "queen" form a cluster with "prince"

Task 1: What does distance represent in the plot?

_____

Task 2: Where would you expect to find these words?

- "affection" - Near: _____
- "princess" - Near: _____
- "fish" - Near: _____

Task 3: If words gradually move closer during training, what's happening?

_____

Part C: Hands-On Problem Solving (15 minutes)

C1: Build Your Own Embeddings (5 minutes)

Design Embeddings from Scratch

Given these 5 sentences:

1. The cat sleeps
2. The dog plays
3. Cats and dogs play
4. Birds fly high
5. Fish swim deep

Task 1: Create word-context pairs for "cat" (window=1):

- Context words: _____
- Context words: _____

Task 2: Design simple 2D vectors for these words:

Word	x	y
cat		
dog		
bird		
fish		

Task 3: Which words should cluster together? Why?

_____

C2: Application Design (5 minutes)

Building with Embeddings

Task 1: Design a synonym finder:

1. Input: _____
2. Process: _____
3. Output: _____

Task 2: Sketch a sentiment analyzer:
How would you use embeddings to determine if text is positive/negative?

Task 3: Your creative application:
Design a new use for word embeddings:

- Name: _____
- Purpose: _____
- How embeddings help: _____

C3: Debugging Scenarios (5 minutes)

Common Pitfall

Real problems you might encounter:

Problem Solving

Scenario 1: The word "bank" appears near both "river" and "money".

- Problem: _____
- Solution: _____

Scenario 2: A new word "COVID" doesn't exist in your embeddings.

- Problem: _____
- Solution: _____

Scenario 3: Your embeddings show "doctor"="male, "nurse"="female bias.

- Problem: _____
- Solution: _____

Part D: Reflection & Extension (10 minutes)

D1: Self-Assessment Checklist (3 minutes)

Checkpoint

Check off what you can now do:

- ☐ Explain why words need to be vectors
- ☐ Describe how Word2Vec learns from context
- ☐ Calculate word similarity (roughly)
- ☐ Perform word arithmetic
- ☐ Identify word clusters in 3D space
- ☐ Design applications using embeddings
- ☐ Recognize common problems and solutions

D2: Real-World Connections (3 minutes)

Real World Application

Connect to Industry:

1. How does Google use embeddings in search?

   _____

2. How do embeddings connect to ChatGPT/BERT?

   _____

3. Name one ethical concern with embeddings:

   _____

D3: Challenge Questions (4 minutes)

Think About It

Going Deeper:

1. Could we create embeddings for images? How?

2. What about embedding DNA sequences or music?

3. How would you embed an entire document (not just words)?

Final Reflection

Discovery Moment

The Big Picture:
You've learned that word embeddings transform language into mathematical space where:

- Meaning becomes measurable
- Relationships become computable
- Patterns become visible

This is the foundation of modern NLP - from search engines to chatbots to translation systems!

Areas needing review? List them here:

- _____
- _____
- _____

Most interesting discovery: _____

One question you still have: _____

Next Steps

- Review sections with ratings below 3
- Try coding your own Word2Vec model
- Explore pre-trained embeddings (GloVe, FastText)
- Learn about contextual embeddings (BERT)
- Apply embeddings to a real project

— End of Assessment —