

CSSE 315 – Natural Language Processing  
**Rose-Hulman** Institute of Technology

Worksheet 06

Name (Print): \_\_\_\_\_ Date: \_\_\_\_\_

## 1 Review

1. **Encoder-only** architecture focuses on \_\_\_\_\_, while **Decoder-only** architecture focuses on \_\_\_\_\_.
2. Align tasks with the architecture (decoder-only, encoder-only, decoder-encoder):

Task	Architecture
Sentiment analysis	
Translation	
Text generation	
Text classification	
Text summarization	
QA Extractive	
QA Generative	

3. Match

Model	Architecture
BART, Gemini	Encoder
GPT	Encoder-Decoder
BERT	Decoder

## 2 Embeddings

4. Word embeddings map words to a \_\_\_\_\_ space, where similar words are positioned closer together.
5. word2Vec common methods:  
\_\_\_\_\_ predicts a word based on its surrounding words, while \_\_\_\_\_ predicts surrounding words given a word.
6. True/False  
Word2Vec embeddings are static, meaning the same word always has the same vector representation re
7. Traditional search relies on \_\_\_\_\_ matches, while vector search retrieves results based on \_\_\_\_\_ similarity.
8. What are the advantages of FastText over Word2Vec?
9. Circle the correct answer:  
Contextual embeddings generate (a. **different vectors**) (b. **the same vector**) for a word  
(a. **depending on its context**) (b. **regardless its content** in a sentence)
10. Name at least 3 types of embeddings:  
(a)  
(b)  
(c)
11. What distance metrics are commonly used for Vector Search?
12. Traditional search relies on \_\_\_\_\_ matches, while vector search retrieves results based on \_\_\_\_\_ similarity.
13. Describe one real-world application of vector search:
14. Locality Sensitive Hashing (LSH) groups data into \_\_\_\_\_, making similar items more likely to collide in the same group
15. True/False  
Tree-based methods are highly efficient for high-dimensional vector spaces.
16. What is the main advantage of SCANN compared to linear search? Explain:

### 3 Vector Database

17. A vector database stores \_\_\_\_\_ and retrieve \_\_\_\_\_

18. Complete table

Task	Vector Database	Traditional Database
Query		
Data Representation		
Use case		