CS211 Quiz 3 Review Solutions

True/False

1. Vector embeddings always produce the same output length regardless of input data size.

   Solution: True

   Vector embeddings transform input data into fixed-length numerical vectors regardless of input size. This is a fundamental property of embedding models, they convert variable-length input into fixed-length vectors.

2. Semantic similarity can only be calculated between pieces of text of the same length.

   Solution: False

   Semantic similarity can be calculated between texts of different lengths since they are converted to fixed-length embeddings. Once texts are converted to embeddings, their original lengths become irrelevant for similarity calculations.

3. FLUX requires at least one text prompt to generate images.

   Solution: True

   FLUX needs at least one natural language text prompt to generate images. Like other text-to-image models, FLUX requires natural language guidance to create images.

4. In semantic search, exact keyword matching is more effective than using embeddings.

   Solution: False

   Embeddings can capture semantic relationships and find relevant content even with different vocabulary, making them more effective than exact matching. Embeddings understand meaning and context, allowing them to find relevant matches even when different words are used.

5. The quality of vector embeddings depends on the size of the training dataset used to create them.

   Solution: True

   Larger, more diverse training datasets generally produce better quality embeddings. More training data helps models learn better semantic representations and handle a wider variety of inputs.

Multiple Choice

6. Which statement about vector embeddings is correct?

   a) They can only be used with numerical data
   b) They transform data into fixed-length numerical representations
   c) They always increase the dimensionality of the data
   d) They can only be used for text processing

   Solution: b

   They transform data into fixed-length numerical representations This captures the key property of embeddings - converting variable input into consistent-length vectors.

7. What is a key benefit of semantic search using embeddings?

   a) It requires less computational power than traditional search
   b) It can find related content even with different vocabulary
   c) It only works with short text fragments
   d) It provides faster indexing than keyword-based search

   Solution: b

   It can find related content even with different vocabulary Embeddings capture meaning rather than just matching words, enabling more flexible and intelligent search.

8. When working with FLUX for image generation, which is true?

   a) It can only generate artistic illustrations
   b) It works without any user input
   c) It uses natural language prompts to create images
   d) It only works with black and white images

   Solution: c

   It uses natural language prompts to create images. FLUX is a text-to-image model that generates images based on natural language descriptions.

9. Why do we calculate cosine similarity between embeddings?

   a) To measure the semantic relationship between items
   b) To reduce processing time
   c) To save storage space
   d) To validate the embedding process

   Solution: a

   To measure the semantic relationship between items Cosine similarity measures the angle between vectors, indicating how semantically similar their meanings are.

DataFrames

10. Consider the following code:

    What will be printed from this code?

```
import pandas as pd
data = {
    'student': ['Emma', 'James', 'Sofia'],
    'score': [85, 92, 88],
    'subject': ['Math', 'Science', 'Math']
}
df = pd.DataFrame(data)
print("Original DataFrame:")
print(df)
df = df.drop('score', axis=1)
df.iloc[2, 1] = 'Art'
print("\nModified DataFrame:")
print(df)
```

Solution

```
Original DataFrame:
   student  score subject
0    Emma     85    Math
1   James     92  Science
2   Sofia     88    Math

Modified DataFrame:
   student subject
0    Emma    Math
1   James Science
2   Sofia    Art
```

Explanation

The code first creates a DataFrame, prints it, drops the 'score' column, changes Sofia's subject to 'Art', and prints the modified result.

11. Given the following DataFrame:

```
import pandas as pd
data = {
    'student': ['Emma', 'James', 'Sofia'],
    'score': [85, 92, 88]
}
df = pd.DataFrame(data)
```

Write code to add a new column called 'status' with values ['Fail', 'Pass', 'Pass'] and print the resulting DataFrame.

Solution

```
df['status'] = ['Fail', 'Pass', 'Pass']
print(df)
```

Output:

```
   student  score status
0    Emma     85   Fail
1   James     92   Pass
2   Sofia     88   Pass
```

Explanation

This adds a new 'status' column to the DataFrame using a list of values, showing each student's pass/fail status.