edquest

Knowledge Test Day 4

NLP & EMBEDDINGS

edquest

Q1. Two different words appear in similar contexts and get mapped to similar vector representations. This behavior is most characteristic of:

A. Bag-of-Words.

B. TF-IDF.

C. One-Hot Encoding.

D. Word Embeddings.

edquest

Q1. Two different words appear in similar contexts and get mapped to similar vector representations. This behavior is most characteristic of:

A. Bag-of-Words.

B. TF-IDF.

C. One-Hot Encoding.

D. Word Embeddings.

Only word embeddings capture context-based similarity. BoW and TF-IDF don't.

edquest

Q2. Which of the following might give high importance to the word "the" in a document?

A. Bag-of-Words.

B. TF-IDF.

C. BERT .

D. Word Embeddings.

edquest

Q2. Which of the following might give high importance to the word "the" in a document?

A. Bag-of-Words.

B. TF-IDF.

C. BERT .

D. Word Embeddings.

BoW counts raw frequencies — so common stopwords like "the" may get high values.

edquest

Q3. Which of the following best defines Natural Language Processing (NLP)?

A. Programming computers to read binary code

B. Teaching machines to interpret and generate human language

C. Compressing human language into zip files

D. Translating HTML documents into JSON

edquest

Q3. Which of the following best defines Natural Language Processing (NLP)?

A. Programming computers to read binary code

B. Teaching machines to interpret and generate human language

C. Compressing human language into zip files

D. Translating HTML documents into JSON

edquest

Q4.Why is BERT considered a breakthrough in Natural Language Processing (NLP)?

A. It requires extensive labeled data for each new task

B. It generates one-hot vectors for sentence representation

C. It uses a single pre-trained model that can be fine-tuned for many NLP tasks

D. It only works for machine translation and not classification

edquest

Q4.Why is BERT considered a breakthrough in Natural Language Processing (NLP)?

A. It requires extensive labeled data for each new task

B. It generates one-hot vectors for sentence representation

C. It uses a single pre-trained model that can be fine-tuned for many NLP tasks

D. It only works for machine translation and not classification

edquest

Q5.What is a key feature of the Word2Vec model in NLP?

A. It represents each word as a unique one-hot vector

B. It predicts the next sentence in a paragraph

C. It learns word vectors based on surrounding context and captures semantic similarity

D. It requires labeled data for training

edquest

Q5.What is a key feature of the Word2Vec model in NLP?

A. It represents each word as a unique one-hot vector

B. It predicts the next sentence in a paragraph

C. It learns word vectors based on surrounding context and captures semantic similarity

D. It requires labeled data for training

edquest

Q6.What does the Inverse Document Frequency (IDF) component of TF-IDF help achieve?

A. Increases the weight of frequently occurring words across all documents

B. Penalizes common words and highlights unique terms that differentiate documents

C. Removes all stopwords from a document

D. Measures the number of characters in a word

edquest

Q6.What does the Inverse Document Frequency (IDF) component of TF-IDF help achieve?

A. Increases the weight of frequently occurring words across all documents

B. Penalizes common words and highlights unique terms that differentiate documents

C. Removes all stopwords from a document

D. Measures the number of characters in a word

edquest

Q7.Why is text preprocessing important before feeding text into an NLP model?

A. To randomly shuffle word order for better generalization

B. To reduce text complexity and convert it into a machine-readable format

C. To generate new vocabulary dynamically during inference

D. To ensure that only labeled data is used in training

edquest

Q7.Why is text preprocessing important before feeding text into an NLP model?

A. To randomly shuffle word order for better generalization

B. To reduce text complexity and convert it into a machine-readable format

C. To generate new vocabulary dynamically during inference

D. To ensure that only labeled data is used in training

edquest

Q7.Which of the following is not typically considered part of data preprocessing in NLP

A. To randomly shuffle word order for better generalization

B. To reduce text complexity and convert it into a machine-readable format

C. To generate new vocabulary dynamically during inference

D. To ensure that only labeled data is used in training