🔥 20 One-Hot Encoding Interview Questions & Answers 🔥

1️⃣ What is One-Hot Encoding in NLP?

One-hot encoding converts each word into a binary vector with a 1 at the index of the word and 0s elsewhere.

2️⃣ Why is One-Hot Encoding used in NLP?

To convert categorical text data (words) into numerical format so that machine learning models can process them.

3️⃣ What is the length of a one-hot encoded vector?

It equals the size of the vocabulary.

4️⃣ Can One-Hot Encoding capture word meaning or context?

❌ No, it only shows presence, not meaning or relationship.

5️⃣ Is One-Hot Encoding sparse or dense?

Sparse — most values are zeros.

6️⃣ How does One-Hot Encoding differ from Label Encoding?

Label Encoding assigns integers to words; One-Hot creates binary vectors to avoid implying ordinal relationships.

7️⃣ Give an example of One-Hot Encoding for the word "apple" in vocab ["apple", "banana", "cherry"]

[1, 0, 0]

8️⃣ What are the limitations of One-Hot Encoding?

High dimensionality, no semantic info, memory inefficiency for large vocabularies.

9️⃣ What happens when vocabulary size increases in One-Hot Encoding?

The vectors become longer and more sparse, increasing computational and memory cost.

🔟 Can One-Hot Encoding be used for sentences?

Not directly. One-hot encodes **words**, not **sentences**.

1️⃣1️⃣ What does a One-Hot vector look like?

A binary vector with only one `1` and all other values as `0`.

1️⃣2️⃣ How is One-Hot Encoding implemented in Python manually?

Create a zero vector of vocab size and set the index corresponding to the word to 1.

1️⃣3️⃣ Which libraries can be used for One-Hot Encoding in Python?

`scikit-learn`, `keras`, `pandas`, `numpy`.

1️⃣4️⃣ Is One-Hot Encoding suitable for large corpora?

No, due to memory inefficiency. Use embeddings like Word2Vec or BERT instead.

1️⃣5️⃣ Does One-Hot Encoding consider word order?

❌ No. It treats each word independently.

1️⃣6️⃣ Is One-Hot Encoding same as Binary Bag of Words?

❌ No. One-Hot encodes single words; Binary BoW encodes full sentences based on word presence.

1️⃣7️⃣ Can One-Hot Encoding be used with deep learning models?

Yes, but not ideal. Word embeddings are preferred due to dense representation.

1️⃣8️⃣ What is the main reason for avoiding One-Hot Encoding in real-world NLP?

It doesn't capture semantic similarity between words.

1️⃣9️⃣ What is a sparse vector in context of One-Hot Encoding?

A vector mostly filled with 0s and only one 1.

2️⃣0️⃣ What is the difference between One-Hot Encoding and Word2Vec?

One-Hot: binary, sparse, no meaning. → Word2Vec: dense, low-dimensional, captures semantic meaning.