

1) Which of the following is/are not correct about Hamming Distance?

- A. Hamming distance tells about the number of positions at which two strings are different.
- B. Hamming distance works for strings of different lengths.
- C. Normalized Hamming distance is the ratio of hamming distance to the length of the string
- D. A higher value of Normalized Hamming distance means two strings are more similar.

2) Normalized Hamming Distance gives the percentage to which string are dissimilar"

- A. TRUE
- B. False

3) ----- is the task of finding relevant information to the user's query from the text documents. Fill in the blanks.

- a. String Similarity
- b. Information Extraction
- c. Information Retrieval
- d. Text Cleaning

4) Which of the following is/are not an application of Information Retrieval?

- A. Extracting information from a text document
- B. A search made for the world's fastest car on Google
- c. Searching of a smartphone on Amazon
- D. Checking how similar two words are

5) Which of the following is/are not an application of Information Retrieval?

- A. Extracting information from a text document
- B. A search made for the world's fastest car on Google
- c. Searching of a smartphone on Amazon
- D. Checking how similar two words are

6) Which of the following processed queries will give results according to the user'sn

- A) Extracting information from a text document
- B) A search made for the world's fastest car on google
- C) Searching of a smart phone on amazon
- D) Checking how similar two worlds are

7) what does the posting list contain?

- A) Frequency of a term
- B) list of documents in which a term is present
- C) list of Boolean values
- D) List of words present in the query

8) The merge algorithm used for processing the query creates a list of documents in which both the term exist. It is equivalent to which of the following operations(s)?

- A. Set difference
- B. Merge Sort
- C. Set interaction
- D. Concatenation

9) Precision the percentage of true positive from all the samples labelled as positive

- A. True
- B. False

10. -----is the measure of how many relevant documents we retrieved from the relevant documents.

- A. Precision
- B. Recall
- C. Accuracy
- D. Root Mean Square

11). Which of the following is/are correct about Information Retrieval and Ranked Retrieval?

- A. Documents returned by Information Retrieval are not ordered.
- B. In Ranked Retrieval, documents are returned in the descending order of relevance.
- C. Documents returned by Ranked Retrieval are ranked in order.
- D. All of the above

12) In ranked retrieval, queries are in the form of free text."

- A. True
- B. False

13) A lower value of the Jaccard Coefficient means the document is more relevant to the query. Therefore, documents are returned in the increasing order of value of the Jaccard Coefficient."

- A. True
- B. False

14) Term Frequency gives

- A. Frequency of a term in a document
- B. Frequency of a term in a corpus
- C. Frequency of a term in the query
- D. Frequency of all the query terms in the corpus

15) "Term Frequency, when used as a score for ranking documents, can return longer documents as more relevant than shorter documents."

- A. True
- B. False

Ans: Term frequency is the frequency of a term in document in longer documents a term can appear more often can appear more therefore, term frequency when used as a score for ranking document can return longer document as relevant

16) The value of inverse document frequency is higher for common terms

- A. True
- B. False

17) For the precision-recall curve, the value of precision and recall are calculated at each rank.

A.True

B.False

18) Which of the following is/are not correct about Mean Average Precision?

A. $P(k)$ is calculated for the first ten ranks.

B. Average Precision doesn't take precision values of the ranks at which a non-relevant document is present.

C. Average Precision for a query is the average of $p(k)$ for the first ten ranks.

D. Mean Average Precision is the mean of Average Precision for all queries

19) Which of the following techniques is used for predicting the next word in a sequence?

A. Information Extraction

B. Language Modeling

C. Language Translation

D. Information Retrieval

20) Which of the following is/are incorrect about language modelling?

A. The task of assigning using previous to a sentence

B. Works without using previous words

C. The task of predicting the probability of upcoming words

D. Predicts upcoming words based on the previous terms

21) Which of the following is/are the application of Language Modeling?

A. Search Autocomplete

B. Language Translation

C. Spell Correction

D. All of the above

22) An N-gram is a sequence of N tokens.

A.True

B. False

23) We required Markov assumption because as the length of the sequence of previous terms increases, the probability of its occurrence decreases".

A.True

B. False

24) Which of the following is/are correct about Perplexity?

A. Perplexity is lower for sentences with a higher probability

B. Perplexity is higher for sentences with a lower probability

C. Perplexity is higher for sentences with a lower probability

D. Perplexity is lower for sentences with a lower probability

25) Extrinsic Evaluation takes more time than Intrinsic Evaluation

A.True

B.False

26) In extrinsic evaluation, we evaluate our language model using some external sources."

A. True

B.False

27) Creating a new feature for a count of characters is a feature engineering technique for text data

A.True

B.False

28) What is the word count for the following sentence: "You can 10 new features"?

A. 5

B.6

C 7

D.8

29) Can you create features out of the dependency tags of the words in a document?

- A. YES
- B. NO

30) Which of the following is not an example of text classification?

- A. Email categorization
- B. Sentiment analysis
- C. Document summarization
- D. Auto – tagging customer queries

31) The Hand-Coded Text classification method has a set of predetermined rules that are the same

for all the problem statements. True or False?

FALSE

32) The Machine learning method for text classification also requires generating rules. True or

False?

FALSE

33) Can classifying text on the basis of its word count be considered a hand-coded rule for text

classification?

YES

34) Hand-coded rules are easy to create without domain knowledge. True or False?

FALSE

35) An exhaustive list of hand-coded rules can be generated for text classification problems. True

or False?

FALSE

36) The machine learning model is trained and tested on the same data. True or False?

FALSE

37) Naive Bayes algorithm considers all features to be independent of each other. True or False?

TRUE