

Revision2

1. _____ is the process of performing data mining on the web.
 - a. Data mining
 - b. File mining
 - c. Web mining
 - d. None of them
2. Web data includes
 - a. web documents
 - b. hyperlinks between documents
 - c. usage logs of web sites
 - d. All the above
3. Web data sets could be very large
 - a. True
 - b. False
4. Which of the following is a web content mining Agent-Based Approach?
 - a. Multilevel-Databases
 - b. Information-Categorization
 - c. Web-Query Systems
 - d. None of the above
5. The research at the hyperlink level is called Hyperlink analysis.
 - a. True
 - b. False
6. Measuring the completeness of Web sites is one of the Web Usage Mining applications?
 - c. True
 - d. False

7. _____ is the discovery of meaningful patterns from data generated by client-server transactions (or) from Web server logs.

- a. Web Content Mining
- b. Web Structure Mining
- c. Web Usage Mining
- d. none of them

8. The Web Usage Mining Data Preparation phase includes:

- a. Data cleaning
- b. User identification
- c. Transaction identification
- d. All the above

9. Clustering and Classification are one of the Pattern Discovery Tasks?

- a. True
- b. False

10. Validation and Interpretation are the Pattern Analysis Tasks

- a. True
- b. False

11. Which of the following techniques can be used for the purpose of keyword normalization, the process of converting a keyword into its base form?

- a. Lemmatization
- b. Levenshtein
- c. Stemming
- d. Both a&c

12. Which of the following features can be used for accuracy improvement of a classification model?

- a. Frequency count of terms
- b. Vector Notation of sentence
- c. Part of Speech Tag
- d. Dependency Grammar
- e. All the above

13. A crude heuristic process that chops off the ends of the words to reduce inflectional forms of words and reduce the size of the vocabulary is called:

- a. Lemmatization
- b. Case Folding
- c. True casing
- d. Stemming

14. The purpose of the inverse document frequency is to increase the weight of terms with high collection frequency.

- a. True
- b. False

15. The number of documents within a collection that contain a particular term is the collection frequency of the term.

- a. True
- b. False

16. refers to the use of a vocabulary and morphological analysis of words

- a. Lemmatization
- b. Case Folding
- c. True casing
- d. Stemming

17. A metric derived by taking the log of N divided by the document frequency where N is the total number of documents in a collection is called:

- a. document frequency
- b. tf-idf weight
- c. collection frequency
- d. inverse document frequency

18. 1. The truth values of traditional set theory is _____ and that of fuzzy set is _____

- a. Either 0 or 1, between 0 & 1
- b. Between 0 & 1, either 0 or 1
- c. Between 0 & 1, between 0 & 1
- d. Either 0 or 1, either 0 or 1

19. The Statement “when I say the temperature is 45° Celsius, the viewer converts the crisp input value into a linguistic variable like ‘hot or cold’” indicates which process?

- a. Fuzzification.
- b. Defuzzification.
- c. Both B & C.
- d. None of the above.

20- The room temperature is hot. Here the hot (use of linguistic variable is used) can be represented by _____.

- a. Fuzzy set.
- b. Crisp set.

21- The first step in fuzzy logic processing is called _____.

- a. Fuzzification.
- b. Rule valuation.
- c. Defuzzification.
- d. Aggregation of all outputs.

22. Which of the following steps use center of gravity (COG) and maximum methods?

- a. Fuzzification
- b. Rule valuation.
- c. Defuzzification
- d. Aggregation of all outputs.

23. Traditional set theory is also known as Crisp Set theory.

a. True

b. False

24. A single neuron consists of

- a. Weights
- b. Inputs
- c. Outputs
- d. All of the above

25. can help to automate complex decision making

- a. Neural Networks
- b. Artificial Neural Networks
- c. Artificial Neurons
- d. None of the above

26. Neural Networks learn from past experience and improve their performance levels.

a. True

b. False

27. NNs generalize the problem from examples to be able to respond to unseen data.

a. True

b. False

28. is the process of modifying the weights in order to produce a network that performs some function.

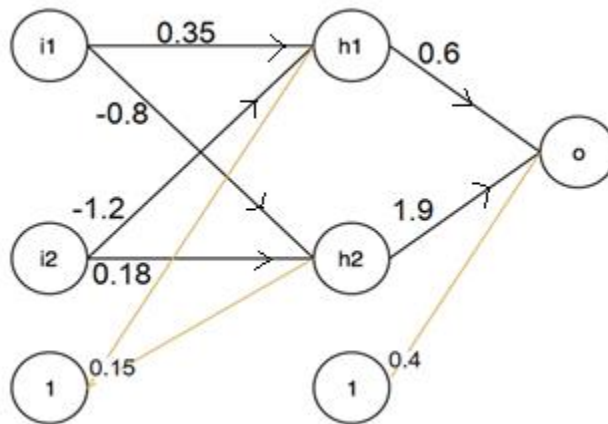
- a. Testing
- b. Learning
- c. Construction
- d. Updating

29. Long training time is considered as a

- a. Weakness
- b. Strength
- c. Not a correct statement

30. Is one of the simplest models of Neural Networks.

- a. single -layer feed-forward
- b. Perceptron
- c. Recurrent Neural Network
- d. Multi-layer Neural Network



Given

$$i_1 = 1.9, i_2 = 0.6$$

use the sigmoid function as an activation function to ans

Answer the following questions

31. $h_1 = \dots\dots\dots$

- a. 0.095
- b. 0.4763
- c. 0.5237

d. 0.3512

32. $h_2 = \dots\dots\dots$

a. -1.262

b. 0.2206

c. 0.7794

d. 1.2356

33. $O = \dots\dots\dots$

a. 0.75646

b. 1.13336

c. 0.24354

d. 0.4256

34. $\dots\dots\dots$ is a network with feedback, where some of its inputs are connected to some of its outputs (discrete time).

a. Perceptron.

b. multi-layer Networks.

c. Recurrent Neural Networks

d. none of the above

35. Perceptrons are used in linear problems such as Boolean problem (AND,OR).

a. True.

b. False

36. Sentiment Analysis is an application of Recurrent Neural Networks

a. True.

b. False

37. web data are usually semi-structured or unstructured

a. True.

b. False

38. Which of the following is a TF-IDF formula:

a. $(N/df(t))$

b. $TF(t,d) \times IDF(t)$

c. $TF(t,d) * \log(N/df(t))$

d. Both b & c

39. Which of the following approaches used for determining the terms' weights

a. Binary weights

b. Term Frequency (TF)

c. Inverse Document Frequency (IDF)

d. TF-IDF

e. All the above

40. Which of the following is a Pattern Discovery Task?

a. Clustering and Classification

b. Interpretation

c. Association Rules

d. Both a&c