Question Paper – for DAI & DBDA course

Total Questions: 40 Duration: 60 Min.

---

1. Consider a CNN with an input layer of size 64x64x3, followed by a convolutional layer with 16 filters of size 5x5, zero padding and stride 2, a batch normalization layer, and a ReLU activation layer. What is the size of the output feature map after the convolutional layer?

A. 30x30x16 B. 31x31x16 C. 32x32x16 D. 33x33x16

2. Which of the following neural embedding method is used in word2vec?

A. Continuous Bowl of Words B. Skip Gram

C. Term frequency-inverse document frequency D. Both A and B

3. Determine the value of k, so that vectors u and v are orthogonal where:

u = (2, 3k, -4, 1, 5) and v = (6, -1, 3, 7, 2k)

A. k = 2 B. k = -2 C. k = -1 D. k = 3

4. Which of the following algorithm will not overfit?

A. Random Forest B. GBM C. XGBoost D. SVM

5. Let us take y as the actual value of a target variable and f(x) as the predicted value of y. The values of y are: 0.5, 1.2, 2 and 5. The values of f(x) are 0.6, 1.4, 1.5 and 1.8. What is the average absolute L1 Loss of the above predictions?

A. 1.317 B. 4.0 C. 0 D. 1.0

6. The number of hidden layers in a Recurrent Neural Network is 10. What happens if we increase the number of hidden layers to 14? Assume bias and variance is denoted by B and V respectively.

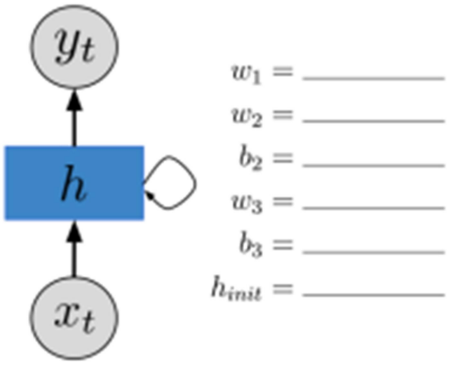
A. B will decrease, V will increase

B. B will increase, V will decrease

C. Both B and V will decrease

D. Both B and V will increase

7. Consider the following RNN architecture and the corresponding details.



where the model is defined as follows:

$$h_t = f(w_1 x_t + w_2 h_{t-1} + b_2)$$
$$y_t = g(w_3 h_t + b_3)$$

where the activation functions f and g are given as

$$f(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$
$$g(x) = x.$$

This RNN is supposed to output 0 as long as no 1 occurs in the sequence. As soon as a 1 shows up in the sequence, it is supposed to only output 1s. For example, if the input sequence is 00011001, the output sequence should be 00011111. What are the values for the weights, biases, and the initial value of the hidden state that produce this output?

A. w1 = 1, w2 = 1, b2 = -0.1, w3 = 1, b3 = 0, hinit = 0

B. w1 = 1, w2 = 1, b2 = 0, w3 = 1, b3 = 0, hinit = 0

C. w1 = 1, w2 = 1, b2 = -0.1, w3 = 1, b3 = 0.1, hinit = 0

D. w1 = 1, w2 = 1, b2 = -0.1, w3 = 1, b3 = -0.1, hinit = 0

8. A decision tree uses ordered numerical attributes. What is the number of splits if an attribute has 4 different values?

A. 7 splits or one 4-way split
B. 3 splits or one 4-way split
C. 4 splits or two 4-way split
D. 3 different binary splits or one 4-way split

9. The weather in Bangalore is notoriously unstable. We consider two cases: rainy and sunny, and we assume that the weather changes every day. It follows the following transition probabilities.

p1: If it rains, the probability of being sunny in the next day is 0.6

p2: When it is sunny, the probability of raining in the next day is 0.3

Assume that the following observations are made over a ten day period.

Day1 - Sunny, Day5 - Rainy, Day7 - Rainy, Day10 - Sunny

What is the probability of being sunny on Day6?

A. 9/17 B. 8/17 C. 3/5 D. 1/5

10. Which of the following statements is false in case of factor analysis?

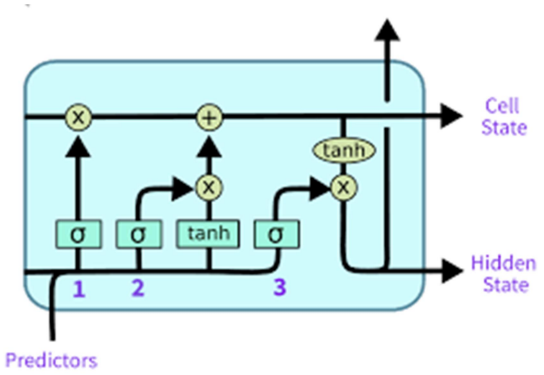
A: It helps to understand whether a number of variables of interest Y1, Y2,…Yn, are linearly related to a smaller number of unobservable factors F1, F2, : ::, Fk.

B: It can distinguish between dependent and independent variables.

C: It can be used to reduce the information in model by reducing the dimensions of the observations.

D: It is used to identify the latent variables.

11. The representation of a LSTM unit is given below. What is the mapping of 3 gates of LSTM unit?



A. 1- Input Gate, 2-Forget Gate and 3-Output Gate

B. 1- Forget Gate, 2-Input Gate and 3-Output Gate

C. 1- Update Gate, 2-Forget Gate and 3-Output Gate

D. 1- Input Gate, 2- Update Gate and 3-Output Gate

12. In transformers, what is the role of the attention mask?

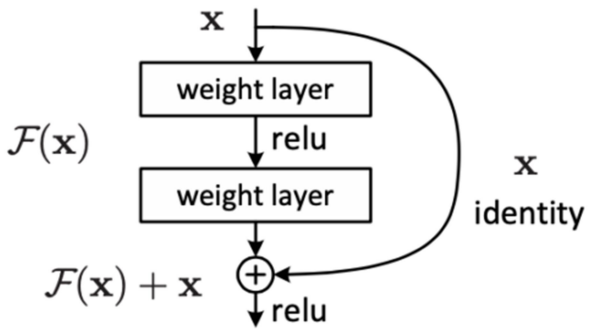
A. To indicate which words in the input sequence should be ignored when computing attention scores

B. To prevent the model from attending to future words in a sequence during training

C. To regularize the attention mechanism and prevent overfitting

D. None of the above

13. The ResNet architecture introduces the simple concept of adding an intermediate input to the output of a series of convolution blocks. The building block of a residual network is given below. Which of the following statements is FALSE?



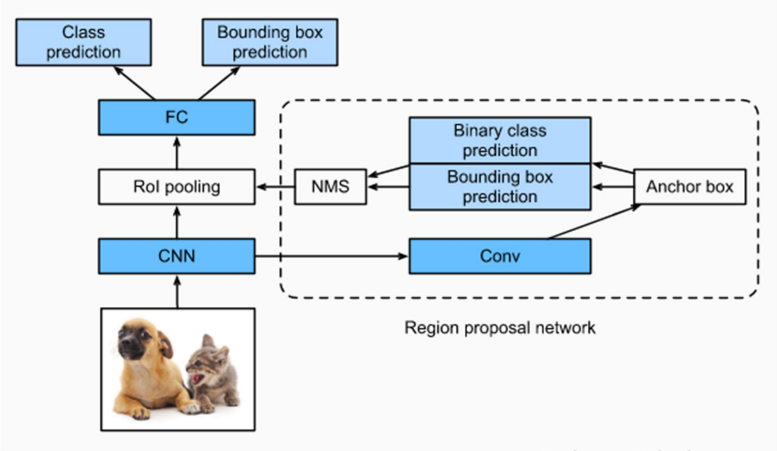
A. Skip connections will not add additional computational cost during training the network

B. Skip connections will increase the impact of vanishing gradient problem

C. ResNet stacks multiple identity mappings, skips those layers

D. None of the Above

14. What is the name of the network architecture that represents the following diagram?



A. RCNN

B. Mask RCNN

C. Faster RCNN

D. Fast RCNN

15. The following are the notations used in error decomposition in machine learning. Which of the following statements is TRUE?

H – Hypothesis Space

I – Number of train data samples

P – Tolerance of the optimizer

A. If H increases, I and P are fixed then the approximation error decreases

B. If H increases, I and P are fixed then the estimation error increases

C. If H and P are fixed, I increases then the estimation error increases

D. If H and P are fixed, P increases then optimization error increases

16. What is the output of the following lines of code?

```
def divide(x, y):
    try:
        result = x // y
    except Exception:
       print("ERROR: Incorrect operation!")
    else:
        print("Your answer is :", result)

divide(3, 2)
```

A. Your answer is: 1

B. ERROR: Incorrect operation!

C. Syntax Error at line 3

D. Syntax Error at line 6

17. What is the output of the following python code snippet?

```
import numpy as np
x = [1, 2, 3, 4, 5, 6, 7]
x1, x2, x3 = np.split(x, [2, 4])
print(x1, x2, x3)
```

A. [1 2] [3 4] [5 6 7]

B. [1 2] [3 4 5 6] [7]

C. [1 2 3 4] [5 6] [7]

D. [1 2] [1 2 3 4] [5 6 7]

18. What is the output of the following python program?

```
import pandas as pd
d = pd.DataFrame({"A": [2, 3, 4]})
(d.assign(A=lambda x: x['A'] + 1).assign(B=lambda x: x['A'] + 2)).iloc[1:2,1]
```

A. 1 6

B. 2 3

C. 3 4

D. 3

19. Which of the following is a transformation supported by Apache Spark?

A. flatMap B. countByKey C. collect D. count

20. What is the output of the following python code?

```
import numpy as np
z = np.array([1,2,3])
nz = 2
z0 = np.zeros(len(z)+ (len(z)-1)*nz)
z0[::nz+1] = z
print(z0)
```

A. [1. 0. 0. 2. 0. 0. 3.]

B. Syntax Error

C. [1 0 0 0 0 0 2]

D. [1.0, 0.0, 2.0, 0.0, 3.0]

21. What is the output of the following python code snippet?

```
s = (1, 3, 5, 7, 9)
print( s[ : -2 : ]
```

A. (7) B. (9, 7) C. (1, 3, 5) D. (1, 3)

22. What is the role of a Kafka producer when it connected to a data source?

A. Read data from a Kafka topic and process it

B. Manage the partitioning and replicate the of a Kafka topic

C. Write data to a Kafka topic

D. To manage the consumption of data from a Kafka topic

23. What is the purpose of magic methods in Python?

A: Can be used for operator overloading

B: Can be used for method overloading

C: Can be used for performing object swapping

D: All of the Above

24. What is the output of the following Python program?

```
class A:
    def __init__(self, x):
        print("I am A")
class B(A):
    def __init__(self, x, y):
        A.__init__(self, x)
        print("I am B")
I = B("I am X", "I am Y")
```

A. I am X
   I am Y

B. I am X I am A
   I am Y I am B

C. I am A
   I am B

D. I am A

25. What is the output of the following python code snippet?

```
f1 = lambda x: x**2
f2 = lambda F, m: lambda x: F(x)*m
f2(f1, 4)(3)
```

A. 12

B. 144

C. 48

D. 36

26. Which of the following statement is FALSE in Python3.x?

A. An “except” clause may name multiple exceptions as a parenthesized tuple.

B. The “try ... except” statement has an optional “else” clause, which, when present, must follow all except clauses.

C. “raise” statement allows us to force a specified exception to occur.

D. An “except” clause cannot specify a variable after the exception name.

27. Which of the following statistical measure cannot be a negative number?

A. Arithmetic Mean

B. Harmonic Mean

C. Geometric Mean

D. Median

28. Following are the results of 10 tosses of a coin:

H, T, T, T, T, H, T, H, T, T

What is the estimate of the standard error of estimate of the probability of head for this coin? H represents Heads and T represents Tails.

A. 0.7

B. 0.21

C. 0.14

D. 0.42

29. Which of the following are valid Database models?

A. Relational

B. E-R

C. O-O

D. All of the above

30. Database Normalization is____.

A. a design technique to minimize data redundancy and maintain data integrity.

B. the process of adding redundant data to speed up complex queries.

C. a process of reducing the number of tables by deleting non-relevant tables.

D. All of the above

31. In a 1:N relationship, the foreign key is placed in____.

A. either table without specifying parent and child tables.

B. the parent table.

C. the child table.

D. either the parent table or the child table.

32. Which of the following statements is TRUE w.r.t a Foreign Key?

A. helps connect tables.

B. references a primary key.

C. cannot be null.

D. All of the above

33. Oracle executables are installed in the location pointed to by___.

A. ORACLE_BASE

B. ORACLE_EXEC

C. ORACLE_HOME

D. None of the above

34. Which of the following databases leverages in-memory computing?

A. Oracle

B. MongoDB

C. HBase

D. MySQL

35. PL/SQL triggers____.

A. are named PL/SQL blocks stored in a database.

B. can fire before or after a triggering event.

C. can be Application Trigger or Database Trigger.

D. All of the above

36. Which of the following terms most appropriately explains the distribution of data across servers?

A. Schema free

B. Auto-sharding

C. Distributed query support

D. Integrated caching

37. Which of the following is INCORRECT about MongodDB?

A. _id field is reserved for use as a primary key.

B. If document inserted without the _id field, system will throw an error.

C. The maximum BSON document size is 16 MB.

D. The field names cannot start with "$" and cannot contain ".".

38. Which of the following NoSQL databases has version as one of the key features for data storage?

A. HBase

B. MongoDB

C. Neo4j

D. Redis

39. What is the correct syntax for creating a table named 'websitedata' under the namespace 'data' with a column family 'website1'

A. create table 'data:websitedata', 'website1'

B. create table data.websitedata, column_family: website1

C. create 'data:websitedata','website1'

D. create 'data.websitedata', {column => 'website1'}

40. Which of the following is correct SQL query to display details of employee with IDs from 150 to 160?

A. SELECT * FROM EMPLOYEES WHERE EMPLOYEE_ID FROM 150 TO 160

B. SELECT ALL COLS FROM EMPLOYEES WHERE EMPLOYEE_ID FROM 150 TO 160

C. SELECT * FROM EMPLOYEES WHERE EMPLOYEE_ID BETWEEN 150 AND 160

D. SELECT * FROM EMPLOYEES WHERE EMPLOYEE_ID IN RANGE (150, 160)