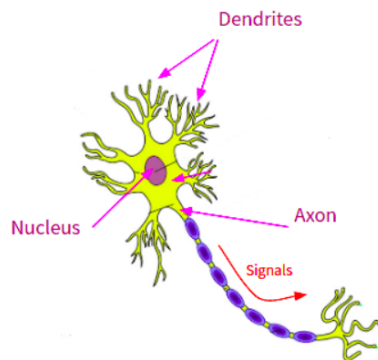


## Answer The Following

1. A biological neuron is illustrated by the following image:

A biological neuron is illustrated by the following image:



Assume that an artificial neuron is designed by  $y = f(x) = g(z)$  (where  $z = \phi(x)$ ) where  $g$  is the sigmoid activation function.

### Complete the following:

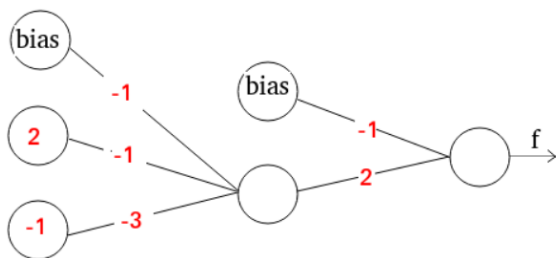
- i.  $x$  is similar to .....
- ii.  $g$  is similar to .....
- iii.  $y$  is similar to .....

2. When learning parameters (connection weights) for a neuron, if a parameter is equal to 0, then ...

- ☐ ... this is not acceptable because a connection weight must not be 0
- ☐ ... it means that the input corresponding to that weight is not taken into account i.e. the neuron's output does not depend on that input
- ☐ ... the neuron's output is always 0.5 because the activation would always be:  $g(0) = 1/(1+e^{-0}) = 0.5$

3. Given the following neural network: Apply forward propagation, to calculate the value of the network's output:

Given the following neural network:



4. Which of the following statements is NOT true?

- a. ☐ Deep Learning is normally Neural Network with many layers
- b. ☐ A neural network can have only one or more than one output units, depending on your problem
- c. ☐ Complex neural network (many layers) can model very complex nonlinear functions
- d. ☐ Adding more layers to a Neural Network is always better

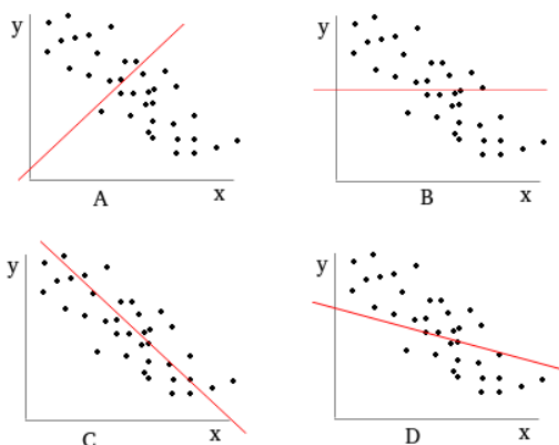
5. Which statement should be correct about classification using Neural Network?

- a. ☐ The number of output units (= the number of classes) must not be greater than the number of inputs (excluding the bias unit)
- b. ☐ If there are only 2 classes (binary classification), we just need only one output unit
- c. ☐ If there are K classes ( $K > 2$ ), then the output y should be a (K+1)-dimensional vector (including the bias unit)
- d. ☐ If there are K classes ( $K > 2$ ), the total value of all elements in the output vector y must be equal to 1

6. Given an input variable x and an output variable y, suppose that we have found a linear relationship between x and y, represented by a function  $f(x)$ . Which of the following could possibly be  $f(x)$ ?

- A)  $f(x) = \sqrt{x} + 2$
- B)  $f(x) = x^2 - 3$
- C)  $f(x) = \frac{1}{x} + x + 1$
- D)  $f(x) = 2 - x$

7. Which of the following linear functions best model the linear relationship in the data set?



8. For a training set of M data points, which of the following should be the cost function for simple linear regression?

A)  $C(\phi_0, \phi_1) = \frac{M}{2} \sum_{i=1}^M (\hat{y}_i - y_i)^2$   
B)  $C(\phi_0, \phi_1) = \frac{1}{2M} \sum_{i=1}^M (y_i - \hat{y}_i)^2$   
C)  $C(\phi_0, \phi_1) = \frac{M}{2} \sum_{i=1}^M (\hat{y}_i - y_i)$   
D)  $C(\phi_0, \phi_1) = \frac{1}{2M} \sum_{i=1}^M (\hat{y}_i - y_i)$

9. True or False:

- i. ☐ Simple Linear Regression: 1 input variable  
Multiple Linear Regression: more than 1 input variables
- ii. ☐ Simple Linear Regression: use only one linear function for modeling  
Multiple Linear Regression: use more than one linear functions for modeling
- iii. ☐ Simple Linear Regression: 1 output variable  
Multiple Linear Regression: more than 1 output variables

10. Given that the learning rate equals 0.2. Which of the following is the correct update step in Gradient Descent algorithm?

A)  $\phi_0 := \phi_0 - 0.2 \frac{\partial C}{\partial \phi_0}$   
 $\phi_1 := \phi_1 - 0.2 \frac{\partial C}{\partial \phi_1}$   
B)  $\phi_0 := 0.2\phi_0 - \frac{\partial C}{\partial \phi_0}$   
 $\phi_1 := 0.2\phi_1 - \frac{\partial C}{\partial \phi_1}$   
C)  $\phi_0 := \phi_0 - \frac{\partial C}{\partial \phi_0}$   
 $\phi_1 := \phi_1 - 0.2 \frac{\partial C}{\partial \phi_1}$   
D)  $\phi_0 := \phi_0 - \frac{1}{0.2} \frac{\partial C}{\partial \phi_0}$   
 $\phi_1 := \phi_1 - 0.2 \frac{\partial C}{\partial \phi_1}$

11. Which of the following is NOT correct about the learning rate of Gradient Descent?

- i. ☐ Learning rate is too small --> very slow to converge
- ii. ☐ Learning rate is too large --> may not converge may take very long time to converge
- iii. ☐ Learning rate is sufficiently small --> converge in an appropriate duration of time
- iv. ☐ Learning rate is sufficiently large --> converge in an appropriate duration of time

## 12. True or False:

- i. ☐ Binary classification is the task of classifying binary inputs (0, 1) into N groups ( $N \geq 2$ )
- ii. ☐ Binary classification is the task of classifying input data into two groups
- iii. ☐ Binary classification is the task of classifying binary inputs (0, 1) into only two groups

13. Which of the following is the correct sigmoid activation function in logistic regression?

- A)  $g(\vec{\phi} \cdot \vec{x}) = \frac{1}{1 - e^{-\vec{\phi} \cdot \vec{x}}}$
- B)  $g(\vec{\phi} \cdot \vec{x}) = \frac{1}{1 + e^{-\vec{\phi} \cdot \vec{x}}}$
- C)  $g(\vec{\phi} \cdot \vec{x}) = \frac{1}{1 - e^{\vec{\phi} \cdot \vec{x}}}$
- D)  $g(\vec{\phi} \cdot \vec{x}) = \frac{1}{1 + e^{\vec{\phi} \cdot \vec{x}}}$

14. What should be true about Multinomial classification?

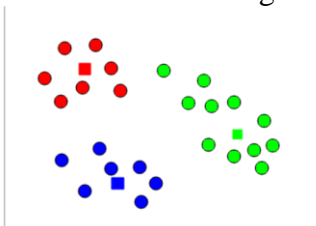
- i. ☐ In Multinomial classification, there are more than two output classes.
- ii. ☐ In Multinomial classification, there are more than two input variables
- iii. ☐ In Multinomial classification, the number of output classes must not be greater than the number of input variables

15. For "One-versus-one method" (OvO), then choose the correct conclusion from the following statements.

- i. ☐ In OvO, we learn a classifier for each pair of classes, then choose the class having the highest number of votes.
- ii. ☐ OvO is actually binary classifier for binary classification (only two output classes)
- iii. ☐ OvO is not for classification.

16. Given two data points:  $x_i$  (3, -1, 2) and  $x_j$  (-2, 0, 9).

17. After running K-means algorithm for a number of iterations, the newly recomputed centroids (square-shaped) are positioned as in the following figure. Can we terminate the clustering process now? (Has the final clustering been found?)



18. A practical use of K-means algorithm is to run it multiple times then select the clustering that has the lowest cost value

19. One of the advantages of GA is that solution time is highly predictable. Explain why.

20. Consider the following problem:

Find two non-negative integers  $x$  and  $y$  such that:  $3x + 4y = 20$

***Which of the following should be the best choice for fitness function?***

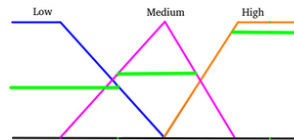
21. True or False: Crossover (recombination) will create two offspring which may not be different from their parent.

22.

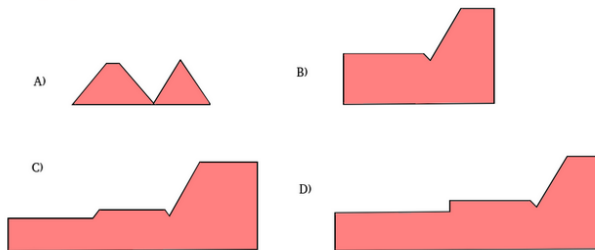
Choose the correct decision region:

Suppose that in a Fuzzy system, there are three membership functions for the output as illustrated in this figure:

Then, after using inference rules, we suppose that the truth values are represented by the green lines.



Which of the following is the correct decision region in which we can use the center-of-gravity defuzzification?



23. State PSO Algorithm Parameters