**QUESTION 1**

1. A decision tree of depth 1 is known as a stump. Suppose you have a dataset containing 4 Boolean attributes and a Boolean output variable. How many possible stumps can you create?

|  |  |  |
| --- | --- | --- |
|  |  | 4 |

### QUESTION 2

1. A decision tree of depth 1 is known as a stump. Suppose you have a dataset containing 4 Boolean attributes and a Boolean output variable. In how many ways can you label each stump?

4

**QUESTION 3**

1. Suppose you have a deck of **fair** cards. The class label of each card is its suit. Remember that there are 4 suits in a standard deck of cards - diamond, clubs, hearts, and spades.  
   https://en.wikipedia.org/wiki/Suit\_(cards)  
   What is the entropy of this deck of cards:

|  |  |  |
| --- | --- | --- |
|  |  | 2 |

### QUESTION 4

1. Suppose you have a deck of **unfair** cards, which has the following suit distribution:  
   Hearts: 15  
   Diamonds: 15  
   Clubs: 11  
   Spades: 11

The class label of each card is its suit. Remember that there are 4 suits in a standard deck of cards - diamond, clubs, hearts, and spades. https://en.wikipedia.org/wiki/Suit\_(cards) and the total number of cards in a deck is 52.  
Which of the following is true about this deck of cards:

* It will have lower entropy than the deck of fair cards

### QUESTION 5

1. Suppose that you want to send a secret message that will be encoded using 4 characters - A, B, C, and D.  
   Which of the following would be true?   
   Hint: See this: https://en.wikipedia.org/wiki/Entropy\_(information\_theory)#Example

->

|  |  |  |
| --- | --- | --- |
|  |  | If the probability distribution is skewed and a particular character, say A, is more likely than others, then we can devise a scheme in which the number of bits needed per character would be less than 2 |

* If all the 4 characters are equally likely, then we will need at least 2 bits to encode each character.

### QUESTION 6

1. Suppose you have the following dataset (represented in tabular and graphical form below):

Chart

Description automatically generated

Which attribute out of {color, shape, and size} would give the best information gain?

|  |  |  |
| --- | --- | --- |
|  |  | Color |

### QUESTION 7

1. Suppose you have the following dataset (represented in tabular and graphical form below):
2. Which of the following is false?

|  |  |  |
| --- | --- | --- |
|  |  | The best attribute to use for the root node would be shape |

### QUESTION 8

1. Consider the following dataset:  
   Table

   Description automatically generated

Which of the following is false?

-> The best decision tree that would correctly classify this dataset can have less than 4 leaf nodes

### QUESTION 9

1. Suppose you want to represent the following Boolean function:  
   Y = A V [B ^ C]  
   where A, B, and C are all Booleans.  
   In the shortest possible tree, which of the following would be true?

You can count root node as an internal  node.

|  |
| --- |
| * There would be 4 leaf nodes |
|  |  | There would be 3 internal nodes |
| QUESTION 10  1. Suppose you have the dataset shown below:  Which of the following is/are true?  |  |  |  | | --- | --- | --- | |  |  | The information gain of attribute Z is 0 | |  |  | The dataset can be correctly classified using a stump. | |  |  | The information gain of attribute Y is 1. | |  |  | The information gain of attribute X is between 0 and 1 |   Quiz 1: |  |  |

Quiz1

1. H = - sum of i pi log2 pi

2.

. Minimum value of entropy is 0 and maximum value is 1

. If all data points have the same class, entropy is 0

. If the data points have equal class distribution, then entropy is 1

3. Convex

4.

. Information gain between children nodes and parent node should be maximized.

5.

. There are no more attributes left to split on

. All the data items in a node belong to the same class

Quiz 3 Perceptron

**Question 1**

10 out of 10 points

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | | | | | |
|  | Which of  the following functions can be correctly represented by a single perceptron using sign activation function? | | |  |  |  |
|  | |  |  |  |  |
| Selected Answers: | | Correct  AND |
|  | | Correct  OR |
|  | | Correct  NOT |

A neuron with 4 inputs has the weight vector w = [1, 2, 3, 4]T and a bias θ = 0 (zero). The activation function is linear, where the constant of proportionality equals 2 — that is, the activation function is given by f(net)=2×net. If the input vector is x=[4,8,5,6]T then the output of the neuron will be

|  |  |
| --- | --- |
| Answers: |  |
|  |  |
|  | Correct  118 |

]

**Question 3**

10 out of 10 points

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | | | |
|  | Suppose you would like to represent the Boolean function A ^ NOT B using a perceptron. The function is represented below:  You can assume the representation of the perceptron would be f(x) = sign(w0 + w1 \* A + w2 \* B) where w1 = 1.0 and w2 = -1.0 Which of the following would be acceptable values for w0 |  |  |  |
| |  |  | | --- | --- | | Selected Answer: | Correct  w0 = -0.5 | |  |  |  |

**Question 4**

10 out of 10 points

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | | | |
|  | Suppose you would like to represent the Boolean function NOT (A ^ B) using a perceptron. The function is represented below:  You can assume the representation of the perceptron would be f(x) = sign(w0 + w1 \* A + w2 \* B) where w1 = -1.0 and w2 = -1.0 Which of the following would be acceptable values for w0 |  |  |  |
| |  |  | | --- | --- | | Selected Answer: | Correct  w0 = 1.5 | |  |  |  |

**Question 5**

10 out of 10 points

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | | | |
|  | When training a perceptron with the sign activation function, suppose the hypothesis is: h = sign(wTx).The actual value of the class label is . For any point to be correctly classified which of the following should be true: |  |  |  |
| |  |  | | --- | --- | | Selected Answer: | Correct  y \* h(x) > 0 | | Answers: | Correct  y \* h(x) > 0 | |  | **y > 0 and h(x) > 0** | |  | **y < 0 and  h(x) < 0** | |  | y = 0 and h(x) = 0 | |  |  |  |

**Question 6**

10 out of 10 points

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | | | |
|  | You have learned the following weights vector for a two attribute (x1 and x2, where  and ) Boolean classification problem:  Using the sign activation, which of the following functions can these weights represent? |  |  |  |
| |  |  | | --- | --- | | Selected Answer: | Correct  x2 | |  |  |  |

\_\_\_\_\_\_\_\_\_\_-------------

**Quiz 4**

**Question 1**

5 out of 10 points

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | | | |
|  | You studied the famous sigmoid function in last class and you learned the value of its derivative. If S(x) represents the sigmoid function and S'(x) represents the derivative of the sigmoid function. Which of the following are true? |  |  |  |
| |  |  | | --- | --- | | Selected Answers: | Correct  Plot of S'(x) between -1 and 1 is below: | |  |  |  |

Chart

Description automatically generated

Plot of S(x) between -100 and 100 is below:

**Chart, histogram

Description automatically generated**

**Question 2**

**10 out of 10 points**

**Consider the simple ANN shown below. Note that in this network the flow of data is from bottom (input) to top (output)**

**You can assume that the output of each of the neurons is the same as the input i.e. oi = neti.**

**The weights of the network are shown below. Assume that the input (from left to right) is 0.5 and 1.0.**

**Which of the following is/are true?**

**Diagram

Description automatically generated**

|  |  |
| --- | --- |
| Selected Answers: | Correct  The output from the final layer's right node is 3.0 |
|  | Correct  The output from the hidden layer's leftmost node is -0.5 |
|  | Correct  The output from the hidden layer's middle node is 2.0 |

|  |  |  |  |
| --- | --- | --- | --- |
| Consider the neural network below that simulates the XNOR function. The weights are indicated on the edges.    The activation function for each of the hidden layer and output layer neurons is:  Suppose the input vector is  0,1 , what will be the value if output? | | | |
| |  |  | | --- | --- | | Selected Answer: | Correct 0 | | | | |
| |  | | --- | | Consider the network shown below:    The functions represented are:  q = x + y f = q \* z  What will be the derivative of f with respect to x, y, and z  for the scenario shown above: | | |  |  | | --- | --- | | Selected Answer: | * (-4, -4, 3) | |  |  | | | | | |
|  | | | |
|  | If   represents the sigmoid function, then its derivative with respect to x is: |  |  |
| |  |  | | --- | --- | | Selected Answer: | Correct | |  |  |
| The ReLu activation function is defined as follows:  What would be the value of ReLu(0.5) | | | |
| |  |  | | --- | --- | | Selected Answer: | Correct  0.5 | | | | |
| The tanh function is defined as:    Which of  the following would be true about this function? | | | |
| |  |  | | --- | --- | | Selected Answers: | Correct  For x = 0, the function would be 0 | |  | Correct  It is a continuous function | |  | Correct  For x ->  , the value would be 1.0 | |  | Correct  For x -> - , the value would be -1.0 | | | | |

|  |
| --- |
| Which of the following are continuous and differentiable functions? |
| |  |  | | --- | --- | | Selected Answers: | Correct  Tanh Activation function defined as: | |  | Correct  Sigmoid Activation defined as: | |