:::DATA::: MIDTERM1

POSITIVE

( 1 , 0 , 0 , 0 )  
( 1 , 0 , 1 , 0 )  
( 1 , 1 , 1 , 0 )  
( 1 , 1 , 0 , 1 )  
( 1 , 0 , 1 , 1 )  
( 1 , 1 , 1 , 1 )

NEGATIVE

( 0 , 0 , 1 , 0 )  
( 0 , 1 , 1 , 0 )  
( 1 , 0 , 0 , 1 )  
( 0 , 1 , 0 , 1 )  
( 0 , 0 , 1 , 1 )

Calculate to 3 decimal places

Answers

1. Between 0.989 and 0.999
2. Between -0.005 and 0.005
3. Between 0.587 and 0.597
4. Between 0.613 and 0.623
5. Between 0.995 and 1.005
6. Between 0.966 and 0.976
7. Between 0.002 and 0.012
8. X1
9. Between 0.491 and 0.501
10. Between -0.005 and 0.005
11. Between 0.24 and 0.25
12. Between 0.335 and 0.345
13. Between 0.495 and 0.505
14. Between 0.475 and 0.485
15. Between 0 and 0.01
16. X1
17. Between 460.5 and 461.5
18. Between 111.5 and 112.5
19. Between 0.008 and 0.028
20. Between 0.402 and 0.502
21. Between 0.08 and 0.28
22. Between 4.02 and 5.02
23. Between 1.2 and 1.3
24. Between 0.05 and 0.15
25. Between 1.5 and 2.5
26. Calculate the entropy of the following set of examples to 3 decimals places:
27. Given the following set of examples of points (x1, x2, x3, x4), if we split the data on the x1 attribute, what is the entropy of the set where x1=0.
28. Given the following set of examples of points (x1, x2, x3, x4), if we split the data on the x1 attribute, what is the entropy of the set where x1=1.
29. Given the following set of examples of points (x1, x2, x3, x4), if we split the data on the x1 attribute, what is the information gain using entropy.
30. Given the following set of examples of points (x1, x2, x3, x4), if we split the data on the x2 attribute, what is the entropy of the set where x2=0.
31. Given the following set of examples of points (x1, x2, x3, x4), if we split the data on the x2 attribute, what is the entropy of the set where x2=1.
32. Given the following set of examples of points (x1, x2, x3, x4), if we split the data on the x2 attribute, what is the information gain using entropy.
33. Given the following set of examples of points (x1, x2, x3, x4), which is the better attribute to use first in our Decision Tree, x1 or x2?
34. Calculate the Gini of the following set of examples of points (x1, x2, x3, x4) to 3 decimals places:
35. Given the following set of examples of points (x1, x2, x3, x4), if we split the data on the x1 attribute, what is the Gini of the set where x1=0.
36. Given the following set of examples of points (x1, x2, x3, x4), if we split the data on the x1 attribute, what is the Gini of the set where x1=1.
37. Given the following set of examples of points (x1, x2, x3, x4), if we split the data on the x1 attribute, what is the information gain using Gini.
38. Given the following set of examples of points (x1, x2, x3, x4), if we split the data on the x2 attribute, what is the Gini of the set where x2=0.
39. Given the following set of examples of points (x1, x2, x3, x4), if we split the data on the x2 attribute, what is the Gini of the set where x2=1.
40. Given the following set of examples of points (x1, x2, x3, x4), if we split the data on the x2 attribute, what is the information gain using Gini.
41. Given the following set of examples of points (x1, x2, x3, x4), which is the better attribute to use first in our Decision Tree, x1 or x2?

Given the task of learning to predict the output of a function given a particular input, you are given the following training examples of inputs for the function together with the correct output (x, F(x)):

(-7, -13),  (-2, -3),  (1, 3),  (3, 7),  (7, 15)

We decide to model this as a linear function: F(x) = w0 + w1\*x

17) Initially choosing (w0, w1) = (0, 0)

What is the Squared Error for this model over the set of training examples? (express to 2 decimal place accuracy)

18) Initially choosing (w0, w1) = (1, 1)

What is the Squared Loss for this model over the set of training examples? (express to 2 decimal place accuracy)

19) Initially choosing (w0, w1) = (0, 0)

We decide to use Gradient Descent to find the best weights.  Using all of the training example, what would the new weight be for w0 after the first update from algorithm using a learning rate of 0.01? (Express to 2 decimal places)

20) Initially choosing (w0, w1) = (0, 0)

We decide to use Gradient Descent to find the best weights.  Using all of the training example, what would the new weight be for w1 after the first update from algorithm using a learning rate of 0.01? (Express to 2 decimal places)

21) Initially choosing (w0, w1) = (0, 0)

We decide to use Gradient Descent to find the best weights.  Using all of the training example, what would the new weight be for w0 after the first update from algorithm using a learning rate of 0.1? (Express to 2 decimal places)

22) Initially choosing (w0, w1) = (0, 0)

We decide to use Gradient Descent to find the best weights.  Using all of the training example, what would the new weight be for w1 after the first update from algorithm using a learning rate of 0.1? (Express to 2 decimal places)

Given the task of learning to determine between positive and negative examples of a class where each example has a single feature x.  Given these negative examples:

(-7),  (-2)

and these positive examples:

(1),  (3),  (7)

We decide to use Logistic Regression where we model positive as 1 and negative as 0 and guess positive using the model : Logistic(w0 + w1\*x)

23) Initially choosing (w0, w1) = (0, 0)

What is the Squared Loss for this model over the set of training examples? (express to 2 decimal place accuracy)

24) Initially choosing (w0, w1) = (0, 0)

We decide to use Gradient Descent to find the best weights.  Using all of the training example, what would the new weight be for w0 after the first update from algorithm using a learning rate of 1.0? (Express to 2 decimal places)

25) Initially choosing (w0, w1) = (0, 0)

We decide to use Gradient Descent to find the best weights.  Using all of the training example, what would the new weight be for w1 after the first update from algorithm using a learning rate of 1.0? (Express to 2 decimal places)