Logistic regression is what type of machine learning? Correct classification

In logistic regression, the probability of success i.e. P(Y|X) vs attribute follows a \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ curve, while the logit of success vs attribute follows a \_\_\_\_\_\_\_\_\_\_\_\_ curve Correct sigmoid, linear

|  |
| --- |
| You are given a job that comprises 3 tasks - flipping a coin, rolling a dice, and picking a card from a standard deck. Success in each task is defined as: Flipping a coin - Getting heads Rolling a dice - Getting 6 Picking a card - Getting any spades. You will win if you succeed in any of the tasks. What are the odds of winning? |
| |  |  | | --- | --- | | Selected Answer: | Correct 11:5 | |

A picture containing text, document, screenshot

Description automatically generatedA picture containing text, document, receipt

Description automatically generated

To find the maxima of a \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ function, you would use gradient ascent, whereas to find the minima of a \_\_\_\_\_\_\_\_\_\_\_\_\_\_ function, you would use gradient descent Correct concave, convex

The cost function in classification is defined as   What would be the form of this function (remember h is the independent variable and y is the true class label that is given to you): Correct convex

Which one of the following are main difference between naive Bayes (NB) and Logistic Regression (LR)?

|  |  |
| --- | --- |
| Selected Answers: | Correct  NB makes conditional independence assumption, while LR makes no such assumption |
|  | Correct  NB calculates the conditional probability of the class in an indirect way, while LR models it directly as a function of attributes |
|  | Correct  LR creates a model while NB performs probability computations |

How is overfitting controlled in Logistic Regression? Correct By adding a term for regularization that penalizes large weights

. Suppose we are developing a logistic regression model which is expressed as:

Text

Description automatically generated with low confidence

Note that Delta is the weights vector. Suppose for two-attribute case, you obtain the weights vector as .   
Which one of the following would be true for the decision boundary in the attribute plane?

Answer:
X₂1
Y=0
6
y=1
XL
Explanation - In
this figure, we transition
from negative to positive when (x2) goes from
aboveDiagram

Description automatically generated

Model Evaluation Quiz

Confusion matrix is as shown

Table

Description automatically generated

What is the value of accuracy in percentage? (Only 1 decimal point) -> 83.3

p class=No
20
bo
Predicted class
class=ye
Actual class=Yes
100
class
12
class=No
EL
40
Total
140
Accuracy
= All correct
All
1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | | | | |
|  | For the table shown in previous table, the cost table is shown below:    What is the cost of misclassification. Write a number with no decimal points -> 1940 | |  |  |  |
|  |  |  |  |  |

For the confusion matrix shown below, compute the value of precision?

Table

Description automatically generated

Write answer as a percent with 1 decimal points? [x] % -> 71.4

Graphical user interface, text, application

Description automatically generated

For the confusion matrix shown below, compute the value of recall?

Table

Description automatically generated

Write answer as a percent with 1 decimal points? [x] %

Ans:
Guven Confusion matrix
Predicted I class
Yes
do (EN)
NO
yes
Actual
class
CTP) 100
NO
Чо
200
(FP)
(TN)
Recall
True Positi

For the confusion matrix shown below, compute the value of F-measure?

Table

Description automatically generated

Write answer as a percent with 1 decimal points? [x] % -> 76.9

given that
The confuslon matrix is.
predicted class o
class=yes
class=NO
20 SFP
Actual
class=yes
( 0 ) = TP
40 =EN
120
240
20

If a dataset contains 120 points, and you are  using 10 fold cross-validation, how many data points will be in the test set each time?  
Write your answer as a number with no decimal points -> 12

Dataset contains points p=120

Total fold cross-validation k=10

Data points in the test set each time is p/k

So, p/k = 120/10 =12

So, 12 data points will be in the test set at each time.

Receiver Operating Characteristics (ROC) curve is a plot of \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ vs \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.Correct TPR, FPR

Instance Based Learning

In k-nearest algorithm, which of the following phases requires computation? Correct Test

Which of the following is true about the value k in k-nearest neighbors algorithm? Correct It is a hyperparameter

For the following dataset, what will be the value of error for 1-NN algorithm if we use 14-fold cross validation?Hint: 14-fold cross validation means divide the data into 14 parts (also known as folds), and each time put 1 fold in test and remaining in training. So, the algorithm will need to be run 14 times

Chart, scatter chart

Description automatically generated

Correct 10/14

Text

Description automatically generated

For the following dataset, which of the following are true? Assume that we are doing 14-fold cross validation

|  |
| --- |
| Correct We can design a decision tree such that the error is 0. |
|  |

If I wanted to reduce the effect of noise in a dataset, what should I do with the value of hyperparameter k?

  Increase value of k



Which of the following are true about the time complexity of k-NN using the brute force method?  
Assume following:  
Number of data items: N  
Dimension of each data item: d  
You can assume that k << N

|  |
| --- |
| Correct The total time complexity:O(Nd) + O(N log N) |
| CorrectThe value of k has no effect in time complexity |  |

If you use the scikit learn package, what is the default value of k in k-NN classifier?  
Hint: see <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.NearestNeighbors.html#sklearn.neighbors.NearestNeighbors> -> 5

The time complexity for 1-NN search using a KD tree (i.e. balanced binary tree) is in what range?  
Assume there are N points in the dataset  
Hint: Watch this video <https://www.coursera.org/lecture/ml-clustering-and-retrieval/complexity-of-nn-search-with-kd-trees-BkZTg> Correct O(log N) - O(N)

|  |
| --- |
| Which of the following are valid differences between lazy and eager methods? |
| |  |  | | --- | --- | | Answers: | Correct  Lazy methods wait for test data to arrive before local model creation | |  | Correct  Eager method starts model creation after receiving the training data | |  | Correct  Neural net is an example of eager method | |  | Locally weight regression is an example of an eager method | |
| Which of the following are true about locally weighted regression? Hint: See https://en.wikipedia.org/wiki/Local\_regression |
| |  |  | | --- | --- | | Answers: | Correct  Closer points are given more emphasis | |  | Correct  It is an example of weighted regression where weight is the distance between points. | |  | Correct  A low level polynomial is used for local fitting | |  | Correct  Generally, it provides lower error that global model | |

Ensemble Methods

Suppose n candidates have been called for a job and have been ranked 1,2,3,...,n.  
Let X = the rank of a randomly selected candidate, so that X has pmf

p(x) =  1/n if x = 1, 2, 3, ...,n  
           0     otherwise

Compute E(X) and var(X) for this scenario.

Text

Description automatically generated

Option ③
z1
12
L
ད
1***་•) ：
[=uc-ita, ]
[སྒོ - ] ཡ - ཟླ ) ༥
(་མང - (40)- 6.ངོས་) •() 3 - (༦)༣ -( །
（༤༩X+0) ཙ) 《 - (04 *《-(43

Let X be a continuous, exponentially distributed random variable with parameter . Then the density function of X is 

Compute the value of E(X) and var(X). You can assume that the data is distributed from 0 to infinity.

Hint: The value of the following integrals would help:

Text, letter

Description automatically generated

ANSWER:
Civen that Exponantial distribution
with p.dif
f(2)= teta, och
etada
mean = E(X) =
then
Stafectan
-
[[-teteata); )
**

|  |
| --- |
| For random variables X and Y and constants a, b, and c, which of the following are false? |
| |  |  | | --- | --- | | Selected Answers: | Correct  var(aX + c) = a2 var(X) + c | |  | Correct  var(X + c) = var(X) + c | |

|  |
| --- |
| Customers at a gas station pay with a credit card (A), debit card (B), or cash (C). Assume that successive customers make independent choices, with P(A) = 0.5, P(B) = 0.2, and P(C )= 0.3.  Among the next 100 customers, what are the mean and variance of the number who pay with a debit card? |
| |  |  | | --- | --- | | Selected Answer: | Correct  Mean = 20 Variance = 16 | |

Graphical user interface, application

Description automatically generated

|  |
| --- |
| Suppose that N people throw their hats into the center of a room. The hats are mixed up, and each person randomly selects one hat. Find the expected number of people that select their own hat.  Hint: Formulate the problem as follows: Let X be the number of people that select their own hat, it can be decomposed as:  X = X1 + X2 + ... + XN where each Xi could be 0 or 1 and is a Bernoulli. Applying the E operator,  E(X) = E(X1) + E(X2) + ... + E(XN)  and get the final answer. |
| |  |  | | --- | --- | | Selected Answer: | Correct 1 | |

Text

Description automatically generated

|  |
| --- |
| The mean squared error can be decomposed into 3 parts. Identify them from the list below: |
| |  |  | | --- | --- | | Selected Answers: | Correct  Square of bias | |  | Correct  Noise | |  | Correct  Variance | |

|  |
| --- |
| In case of Artificial Neural Networks, if you increase the number of hidden layers, which of the following would happen: |
| |  |  | | --- | --- | | Selected Answers: | Correct  bias would go down | |  | Correct  variance would go up | |
| Pruning a decision tree would lead to \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ in variance. Hint: Think carefully. The general idea is that if you reduce model complexity, variance goes down and bias goes up. Also, look at this link to understand the concept: https://en.wikipedia.org/wiki/Bias%E2%80%93variance\_tradeoff#Approaches |
| |  |  | | --- | --- | | Selected Answer: | Correct  reduction | |
| In case of k-Nearest Neighbor classifier, the expected value of error can be decomposed as:  E[Error] = Bias2 + Variance + Noise Source: https://en.wikipedia.org/wiki/Bias%E2%80%93variance\_tradeoff#K-nearest\_neighbors Which of the following is/are true? |
| |  |  | | --- | --- | | Selected Answers: | Correct  As k decreases, the value of the variance term goes up. | |  | Correct  As k decreases, the value of the bias term goes down. | |  | Correct  As k increases, the value of the second term (variance) goes down. | |  | Correct  As k increases, the value of the first term (bias) goes up. | |

Suppose there are n independent and identically distributed (iid) random variables X1, X2, ..., Xn  
You can assume that each random variable has the following properties:  
Text

Description automatically generated

Which of the following are aims of PCA analysis?

|  |  |
| --- | --- |
| Answers: | Correct  To convert a set of correlated attributes of a dataset into features that are orthonormal to each other |
|  | Correct  To try to keep as much variance of the original dataset as possible |
|  | Correct  It discovers principal components that are most important for the dataset |

|  |
| --- |
| Which of the following is true about principal components? |
| |  |  | | --- | --- | | Selected Answer: | Correct  Principal components are linear combination of attributes of data | |

|  |
| --- |
| Which of  the following is true about the first principal component? |
| |  |  | | --- | --- | | Selected Answer: | Correct  It is orthogonal to all other principal components and explains the largest percentage of variance in the dataset. | |

|  |
| --- |
| Which of the following is true about the covariance matrix |
| |  |  | | --- | --- | | Selected Answers: | Correct  It is a symmetric matrix | |  | Correct  It is a square matrix | |

Which of the following would have orthonormal eigenvectors?

A picture containing text, game

Description automatically generated

QUESTION 5
10 points
Save Answer
In human genome, there are 4 output states {A, C, T, G} and two hidden states {H, L}. H
stan

Text, letter

Description automatically generated

QUESTION 1
Hidden Markov Models (HMM) are applied to what type of data?
labeled data
unlabeled data
time series or sequential

QUESTION 4
Forward algorithm for HMM uses which of the following programming models?
C++ programming
Functional Programming
D