Incorrect answer!

Review your last 10 questions

3 days, 2 hours remaining until quiz ends.

Pause Quiz | Show Leaderboard | Continue with Quiz

Your Stats	
Not attempted questions(Timed out)	2
Skipped questions	0
Correct questions	3
Incorrect questions	5
Marked Ambiguous	0
Total Questions	160

1/10

Kate wants to build a binary classifier, using either Brute force implementation of Naïve Bayes or KNN. The dataset, she wants to work on, has exactly a million datapoints. Further it is given that the dimensionality of the data is very small compared to the number of datapoints. She is critical about saving space rather than the time taken to classify points.

Which algorithm among the two should be chosen Kate?

- KNN
- Naïve Bayes
- Any of the two, since both have comparable space complexities in this case
- Insufficient information

Incorrect

2/10

Mike has started working on random Random Forest Classifier without much knowledge of the model. He has imported the model using the Python Code given below

```
from sklearn.ensemble import RandomForestClassifier
```

Online Chat

Which of the following Random Forest Classifier model created by Mike would execute the fastest?

- classifier = RandomForestClassifier(n_estimators = 50, random_state = 0)
- classifier = RandomForestClassifier(n_estimators = 10, random_state = 997)
- classifier = RandomForestClassifier(n_estimators = 1000, random_state =10)
- classifier = RandomForestClassifier(n_estimators = 200, random_state = 20)

Incorrect

3/10

A data scientist received a dataset and wanted to fit a regression model on it, but he was challenged by his colleague to not use gradient descent to minimize the cost function. So, what other method can he use instead of gradient descent?

- Newton's method
- quasi-Newton's method
- Can use 1 or 2
- No other method can be used

Correct

4/10

Which of the following is more prone to outliers?

- Logistic regression
- Linear regression
- Both a and b are equally prone

Correct

5/10

Two students A and B were fighting on the use of regularization in regression model, A says that it is used to prevent model to overfit on the data and B says that it is used to prevent model to underfit on the data. Which of the two is\are correct?

- A is correct
- B is correct
- Both of them are correct
- None of them are correct

Online Chat

Incorrect

6/10

Complete the code to print all proper nouns (both plural and singular) from a string s:

```
s = "I love Analytics Vidhya."
import nltk
pos = nltk.?1?(nltk.?2?(s))
for each in pos:
    if each?3? == 'NNP' or each?3? == 'NNPS':
        print(each?4?)
```

- 1 - pos_tag
  2 - word_tokenizer
  3 - [1]
  4 - [0]
- 1 - word_tokenize
  2 - pos_tag
  3 - [1]
  4 - [0]
- 1 - word_tokenizer
  2 - pos_tag
  3 - [0]
  4 - [1]
- 1 - pos_tag
  2 - word_tokenize
  3 - [1]
  4 - [0]

Not attempted

7/10

Given a sample, which of the following tests can be used to determine whether the sample mean is statistically significant when population standard deviation is unknown

- Z-test
- t-test
- F-test
- a & b

Correct

8/10

John, a student, is given the task of binary classification of a given dataset, consisting of 1 million datapoints. Further, after feature engineering, the dimensionality of this dataset has been decreased from 750 to 200. He has been told to use KNN approach only. Further, his priority is that the algorithm should be as fast as possible.

Online Chat

Assume that, using LSH, he tried to classify an unseen point. It is given that he used a total of 6 hyperplanes to do so. Further, it is given that he constructed hashtables for a total of 10 times. If the calculation of a 6 dimensional hashtable key for a datapoint takes 0.1 seconds, then what is the minimum time taken to display the class label of a point(it is given that the number of points per region, in which the hyperplanes divide the 200 dimensional space, is neither large nor small)?

- 12000 seconds
- 1200 seconds
- 6 seconds
- None of the above

Not attempted

9/10

Following is a dataframe *df* with description of MOOCs and whether or not the course is an NLP course:

	Description	NLP
0	This course will help you get started with nat...	1
1	This course will cover all the advance natural...	1
2	In this course we will cover the most basic da...	0

b) A data scientist has created a bag words vector which he created using CountVectorizer class. She has a CountVectorizer object(countvectorizer) and a bag of words vector(bow_vector as a scipy.sparse.csr.csr_matrix).

Which of the following options will help her get the features(unique words after removing stopwords) and their count?

- `countvectorizer.get_feature_names()`
  `bow_vector.toarray()`
- `countvectorizer.toarray()`
  `bow_vector.get_feature_names()`
- `bow_vector.to_array()`
  `countvectorizer.feature_names()`
- `countvectorizer.to_array()`
  `bow_vector.feature_names()`

Incorrect

10/10

You work as a dataset intern in Analytics Vidhya and you received a dataset for analysis. There is one problem, the current dataset has 220 dimensions and you want to reduce them to perform better analysis.

For such high dimensional data, which of the following method is best suited.

- Correlation threshold method
- Variance threshold method

Online Chat

- Genetic algorithm method
- None of the above

Incorrect

Suggested reading

How to create a poet / writer using Deep Learning (Text Generation using Python)?

Pause Quiz Show Leaderboard Continue with Quiz