MLP

Section Id :	64065364126
Section Number :	13
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	34
Number of Questions to be attempted :	34
Section Marks :	100
Display Number Panel :	Yes
Section Negative Marks :	0
Group All Questions :	No
Enable Mark as Answered Mark for Review and Clear Response :	No
Maximum Instruction Time :	0
Sub-Section Number :	1
Sub-Section Id :	640653134056
Question Shuffling Allowed :	No

Question Number : 357 Question Id : 640653903932 Question Type : MCQ Calculator : Yes Correct Marks : 0

Question Label : Multiple Choice Question

THIS IS QUESTION PAPER FOR THE SUBJECT "DIPLOMA LEVEL : MACHINE LEARNING PRACTICE (COMPUTER BASED EXAM)"

ARE YOU SURE YOU HAVE TO WRITE EXAM FOR THIS SUBJECT?
CROSS CHECK YOUR HALL TICKET TO CONFIRM THE SUBJECTS TO BE WRITTEN.

(IF IT IS NOT THE CORRECT SUBJECT, PLS CHECK THE SECTION AT THE TOP FOR THE SUBJECTS REGISTERED BY YOU)

Options :

6406533043842. ✔ YES

6406533043843. ✖ NO

Sub-Section Number :	2
Sub-Section Id :	640653134057
Question Shuffling Allowed :	No

Question Id : 640653903933 Question Type : COMPREHENSION Sub Question Shuffling Allowed : No Group Comprehension Questions : No Question Pattern Type : NonMatrix Calculator : None

Question Numbers : (358 to 362)

Question Label : Comprehension

Consider the following common data and answer the subquestion:

```
df=pd.DataFrame(data={"Name":['Akash','Brajesh','Charu','Deepak'],
                      "Maths":[34,43,56,77],
                      "History":[23,45,67,82],
                      "Hindi":[53,35,np.nan,"bye"],},
                index = range(21,25))
```

Sub questions

Question Number : 358 Question Id : 640653903934 Question Type : MSQ Calculator : Yes Correct Marks : 2 Max. Selectable Options : 0

Question Label : Multiple Select Question

Consider the following code snippet:

```
selection = df.loc[23:24, 'Maths':'History']
```

Which of the following will be equivalent to the above code snippet?

Options :

6406533043844. ✔ `selection = df.iloc[[2,3], [1,2]]`

6406533043845. ✖ `selection = df.iloc[[1,3], [1,2]]`

6406533043846. ✔ `selection = df.iloc[2:4, 1:3]`

6406533043847. ✖ `selection = df.iloc[1:3, 1:2]`

6406533043848. ✖ None of these

Question Number : 359 Question Id : 640653903935 Question Type : SA Calculator : None Correct Marks : 2

Question Label : Short Answer Question

What is the output of the following code snippet:

```
df.loc[23, 'History'] + df.iloc[0,3]
```

Enter -1, if you think the above statement will generate an error.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText
Possible Answers :
120

Question Number : 360 Question Id : 640653903936 Question Type : SA Calculator : None Correct Marks : 2
Question Label : Short Answer Question
What is the output of the following code snippet:

```
print(df.Hindi.apply(type).nunique())
```

Enter -1, if you think the above statement will generate an error.
Response Type : Numeric
Evaluation Required For SA : Yes
Show Word Count : Yes
Answers Type : Equal
Text Areas : PlainText
Possible Answers :
3

Question Number : 361 Question Id : 640653903937 Question Type : SA Calculator : None Correct Marks : 2
Question Label : Short Answer Question
What is the output of the following code snippet:

```
def getPassing(aRow):
  if aRow['Maths']>=40 and aRow['History']>=40:
    return True
  return False

df.apply(getPassing).sum()
```

Enter -1, if you think the above statement will generate an error.
Response Type : Numeric
Evaluation Required For SA : Yes
Show Word Count : Yes
Answers Type : Equal
Text Areas : PlainText
Possible Answers :
-1

Question Number : 362 Question Id : 640653903938 Question Type : MCQ Calculator : Yes Correct Marks : 2

Question Label : Multiple Choice Question

What is the output of the following code snippet:

```
df=df.dropna()
```

Choose correct options from following:

Options :

6406533043852. ✔ The number of rows will decrease in the dataset.

6406533043853. ✖ The number of columns will decrease in the dataset.

6406533043854. ✖ The number of columns and number of rows, both, will decrease in the dataset.

6406533043855. ✖ There will be no change.

6406533043856. ✖ Insufficient information.

Sub-Section Number : 3

Sub-Section Id : 640653134058

Question Shuffling Allowed : Yes

Question Number : 363 Question Id : 640653903939 Question Type : MCQ Calculator : Yes Correct Marks : 2

Question Label : Multiple Choice Question

To load datasets from openml.org, which method will be appropriate?

Options :

6406533043857. ✖ load_openml()

6406533043858. ✖ read_openl()

6406533043859. ✖ read_data()

6406533043860. ✔ fetch_openml()

6406533043861. ✖ load_data()

6406533043862. ✖ load_csv()

6406533043863. ✖ fetch_csv()

Question Number : 364 Question Id : 640653903941 Question Type : MCQ Calculator : Yes Correct Marks : 2

Question Label : Multiple Choice Question

In which of the option given below data preprocessing is required?

Options :

6406533043868. ✖ In some columns which has values between 0 and 1.

6406533043869. ✔ A column contains entity names such as ‘Chennai’, ‘CHENNAI’ and ‘MADRAS’.

6406533043870. ✖ The data has only numbers in all the columns.

Question Number : 365 Question Id : 640653903944 Question Type : MCQ Calculator : Yes Correct Marks : 2

Question Label : Multiple Choice Question

Consider the following code block:

```
X = ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i']
from sklearn.model_selection import KFold
kf = KFold(n_splits = 3)
for train, test in kf.split(X):
  print(train, test)
```

Which of the following may be the correct output of the above code?

Options :

6406533043876. ✖

```
[3 4 5 6 7 8] [0 3 4]
[0 1 2 6 7 8] [2 4 6]
[0 1 2 3 4 5] [0 5 8]
```

6406533043877. ✔

```
[3 4 5 6 7 8] [0 1 2]
[0 1 2 6 7 8] [3 4 5]
[0 1 2 3 4 5] [6 7 8]
```

6406533043878. ✖

```
[3 4 5] [6 7 8] [0 1 2]
[0 1 2] [6 7 8] [3 4 5]
[0 1 2] [3 4 5] [6 7 8]
```

6406533043879. ✖

```
[0 1 2] [3 4 5 6 7 8]
[3 4 5] [0 1 2 6 7 8]
[6 7 8] [0 1 2 3 4 5]
```

Question Number : 366 Question Id : 640653903945 Question Type : MCQ Calculator : Yes Correct Marks : 2

Question Label : Multiple Choice Question

Which of the following scikit-learn classes is best suited for examining how the number of samples impacts training and testing errors?

Options :

6406533043880. ✖ `sklearn.model_selection.train_test_split`

6406533043881. ✖ `sklearn.model_selection.cross_validate`

6406533043882. ✖

sklearn.model_selection.cross_val_score

6406533043883. ✔ sklearn.model_selection.learning_curve

Question Number : 367 Question Id : 640653903952 Question Type : MCQ Calculator : Yes Correct Marks : 2

Question Label : Multiple Choice Question

Consider following two statements:
Statement 1: The GaussianNB classifier can incrementally learn using partial_fit.
Statement 2: GaussianNB performance suffers when features are dependent, as it assumes independence in calculating conditional probabilities.

Options :

6406533043910. ✔ Both statements are True

6406533043911. ✖ Only statement 1 is True

6406533043912. ✖ Only statement 2 is True

6406533043913. ✖ Both statements are False

Question Number : 368 Question Id : 640653903955 Question Type : MCQ Calculator : Yes Correct Marks : 2

Question Label : Multiple Choice Question

Given below code to load a huge file name as filename.csv and this file is not loading at once in the system which parameter should be added to pd.read_csv to load this file ?

```
import pandas as pd
from sklearn.linear_model import SGDRegressor
for train_df in pd.read_csv("filename.csv", __________=1024):
    X = train_df.iloc[:, :-1]
    y = train_df.iloc[:, -1]
    model = SGDRegressor()
    model.partial_fit(prep_X,y)
```

Options :

6406533043923. ✖ max_depth

6406533043924. ✖ C

6406533043925. ✔ chunksize

6406533043926. ✖ warm_start

Question Number : 369 Question Id : 640653903956 Question Type : MCQ Calculator : Yes Correct Marks : 2

Question Label : Multiple Choice Question

Which of the following is true for a hard margin SVM algorithm ?

Options :

6406533043927. ✖ It does not create hyperplanes as a decision boundary

6406533043928. ✔ It will correctly classify all the data point if the data is linearly separable

6406533043929. ✖ It is robust to outliers

6406533043930. ✖ It is mostly used for clustering the data

Question Number : 370 Question Id : 640653903959 Question Type : MCQ Calculator : Yes Correct Marks : 2

Question Label : Multiple Choice Question

What could be the output for below code

```
from sklearn.feature_extraction.text import CountVectorizer
corpus = [ 'This is the first document.',
           'This document is the second document.']
vectorizer = CountVectorizer()
vectorizer.fit_transform(corpus)
print(vectorizer.vocabulary_)
```

Options :

6406533043939. ✖ `['document' 'first' 'is' 'second' 'the' 'this']`

6406533043940. ✔ `{'this': 5, 'is': 2, 'the': 4, 'first': 1, 'document': 0, 'second': 3}`

6406533043941. ✖ `[5, 2, 4, 1, 0, 3]`

6406533043942. ✖
```
[[1, 1, 1, 0, 1, 1],
 [2, 0, 1, 1, 1, 1]]
```

Question Number : 371 Question Id : 640653903961 Question Type : MCQ Calculator : Yes Correct Marks : 2

Question Label : Multiple Choice Question

Decision Trees are prone to:

Options :

6406533043947. ✖ Low bias, low variance

6406533043948. ✖ High bias, low variance

6406533043949. ✔ Low bias, high variance

6406533043950. ✖ High bias, high variance

Sub-Section Number : 4

Sub-Section Id : 640653134059

Question Shuffling Allowed : Yes

Question Number : 372 Question Id : 640653903946 Question Type : MCQ Calculator : Yes Correct Marks : 3

Question Label : Multiple Choice Question

Following is the code to tune the degree parameter of a polynomial regression model.

```
from sklearn.model_selection import GridSearchCV
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import PolynomialFeatures
from sklearn.linear_model import SGDRegressor

param_grid = [{_________: [2, 3, 4, 5, 6, 7, 8, 9]}]
pipeline = Pipeline(steps=[('poly', PolynomialFeatures()),
                           ('sgd', SGDRegressor())])
grid_search = GridSearchCV(pipeline, param_grid, cv=5,
scoring='neg_mean_squared_error',
return_train_score=True)
grid_search.fit(X_train, y_train)
```

What should the blank space contain?

Options :

6406533043884. ✖ 'degree'

6406533043885. ✖ 'sgd_degree'

6406533043886. ✖ 'sgd__degree'

6406533043887. ✖ 'poly_degree'

6406533043888. ✔ 'poly__degree'

Question Number : 373 Question Id : 640653903948 Question Type : MCQ Calculator : Yes Correct Marks : 3

Question Label : Multiple Choice Question

Consider the following code block:

```
from sklearn.datasets import make_regression
X, y = make_regression(n_samples = 1000,
                       n_features = 4,
                       n_informative = 2,
                       random_state=42)

from sklearn.linear_model import SGDRegressor
sgd1 = SGDRegressor(alpha=1e-3,
                    random_state=42
                    penalty='_________________', )
sgd1.fit(X, y)
print(sgd1.coef_)

sgd2 = SGDRegressor(alpha=1e-3,
                    random_state=42,
                    penalty='___________________')
sgd2.fit(X, y)
print(sgd2.coef_)
```

What are the most suitable values to be filled in the two blank spaces (in that order) in the code to expect the following output?:

```
[48.50064306,  0    ,  0     , 8.53931469]
[4.84494867e+01, -7.58443057e-04, -2.09844306e-03,  8.53220113e+00]
```

Options :

6406533043893. ✔ 'l1', 'l2'

6406533043894. ✖ 'l1', None

6406533043895. ✖ 'l2', 'l1'

6406533043896. ✖ 'l2', None

Question Number : 374 Question Id : 640653903950 Question Type : MCQ Calculator : Yes Correct Marks : 3

Question Label : Multiple Choice Question

Consider following code:

```
estimator = SGDClassifier(loss='log_loss',
                          penalty='l2',
                          max_iter=1,
                          warm_start="_____",
                          eta0=0.01,
                          alpha=0,
                          learning_rate='constant',
                          )
pipe_sgd= make_pipeline(MinMaxScaler(), estimator)
```

Which of the following is a suitable choice for warm_start if you wanted to plot learning curve with decreasing loss when trained for 100 epochs? Make necessary assumptions.

Options :

6406533043901. ✔ True

6406533043902. ✖ False

6406533043903. ✖ Yes

6406533043904. ✖ No

6406533043905. ✖ None of these

Question Number : 375 Question Id : 640653903964 Question Type : MCQ Calculator : Yes Correct Marks : 3

Question Label : Multiple Choice Question

How does the `learning_rate` parameter affect the AdaBoost algorithm?

Options :

6406533043960. ✖ It controls the number of weak learners.

6406533043961. ✔ It shrinks the contribution of each classifier.

6406533043962. ✖ It sets the weight of each weak learner.

6406533043963. ✖ It adjusts the speed at which the model learns.

Sub-Section Number : 5

Sub-Section Id : 640653134060

Question Shuffling Allowed : Yes

Question Number : 376 Question Id : 640653903960 Question Type : MCQ Calculator : Yes Correct Marks : 4

Question Label : Multiple Choice Question

The following code produces an output of 0.9125. How is the output expected to change if we increase the max_depth value?:

```
from sklearn.datasets import load_wine
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
X,y = load_wine(as_frame = True, return_X_y = True)

X_train,X_test,y_train,y_test = train_test_split(X,
                                                 y,
                                                 test_size = 0.10,
                                                 random_state = 12)

clf = DecisionTreeClassifier(max_depth = 2,
                             min_samples_split = 2,
                             min_samples_leaf=3,
                             random_state = 81)

clf.fit(X_train, y_train)
print(clf.score(X_train, y_train))
```

Options :

6406533043943. ✔ Output score is likely to increase.

6406533043944. ✖ Output score is likely to decrease.

6406533043945. ✖ Output score may increase or decrease.

6406533043946. ✖ Output score will remain the same.

Question Number : 377 Question Id : 640653903962 Question Type : MCQ Calculator : Yes Correct Marks : 4

Question Label : Multiple Choice Question

Consider the following code. How many different combinations of DecisionTreeClassifier models will be trained internally?

```
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import GridSearchCV
param_grid = [{'max_depth':range(1, 10, 2),
               'min_samples_split': range(1, 10, 3)}]
gs = GridSearchCV(DecisionTreeClassifier(), param_grid, cv = 5)
gs.fit(X,y)
```

Options :

6406533043951. ✖ 20

6406533043952. ✖ 75

6406533043953. ✖ 8

6406533043954. ✔ 15

6406533043955. ✖ 40

Question Number : 378 Question Id : 640653903965 Question Type : MCQ Calculator : Yes Correct Marks : 4

Question Label : Multiple Choice Question

Consider the following code for a VotingClassifier. What is the effect of setting voting='soft'?

```
from sklearn.ensemble import VotingClassifier
clf1 = LogisticRegression()
clf2 = RandomForestClassifier()
clf3 = SVC(probability=True)
eclf = VotingClassifier(estimators=[('lr', clf1),
                                    ('rf', clf2),
                                    ('svc', clf3)],
                                    voting='soft')
eclf.fit(X_train, y_train)
```

Options :

6406533043964. ✖ The final predictions are based on the majority vote.

6406533043965. ✔ The final predictions are based on the average of probabilities predicted by each classifier.

6406533043966. ✖ The final predictions are based on the weighted sum of the predictions.

6406533043967. ✖ The final predictions are based on the classifier with the highest accuracy.

Question Number : 379 Question Id : 640653903966 Question Type : MCQ Calculator : Yes Correct Marks : 4

Question Label : Multiple Choice Question

Given the following code, what will be the output of print(scores.mean())?

```
from sklearn.ensemble import BaggingClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import cross_val_score
clf = BaggingClassifier(DecisionTreeClassifier(),n_estimators=50)
scores = cross_val_score(clf, X, y, cv=5)
print(scores.mean())
```

Options :

6406533043968. ✔ The mean accuracy across all cross-validation folds.

6406533043969. ✖ The mean precision across all cross-validation folds.

6406533043970. ✖ The mean recall across all cross-validation folds.

6406533043971. ✖ The mean F1 score across all cross-validation folds.

Question Number : 380 Question Id : 640653903968 Question Type : MCQ Calculator : Yes

Correct Marks : 4

Question Label : Multiple Choice Question

Which of the following approach(es) is(are) helpful to find a good value for **n_clusters** in **KMeans** clustering algorithm?

Options :

6406533043973. ✔ Plotting the sum of the squared distances of samples to their closest cluster center

6406533043974. ✖ by Changing the distance metric for **KMeans**

6406533043975. ✖ By initialising clustering centroids very far from each other

6406533043976. ✖ All of these

Sub-Section Number : 6

Sub-Section Id : 640653134061

Question Shuffling Allowed : Yes

Question Number : 381 Question Id : 640653903940 Question Type : MSQ Calculator : Yes

Correct Marks : 2 Max. Selectable Options : 0

Question Label : Multiple Select Question

Consider the following code:

```
from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split
X, y = load_iris(return_X_y = True)
```

The sizes of X and y are (150, 4) and (150,) respectively. Which of the following would be the correct code snippet to split X and y into training and test data such that test data has exactly 30 samples?

Options :

6406533043864. ✔

```
X_train, X_test, y_train, y_test =
            train_test_split(X,y,test_size=30)
```

6406533043865. ✖

```
X_train, X_test, y_train, y_test =
            train_test_split(X, y,train_size="20%")
```

6406533043866. ✖

```
X_train, X_test, y_train, y_test =
            train_test_split(X, y, test_size=0.3)
```

6406533043867. ✔

```
X_train, X_test, y_train, y_test =
            train_test_split(X, y, train_size=0.8, )
```

Question Number : 382 Question Id : 640653903942 Question Type : MSQ Calculator : Yes Correct Marks : 2 Max. Selectable Options : 0

Question Label : Multiple Select Question

Consider the following code block:

```
from sklearn.linear_model import linear_regression
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import ShuffleSplit
lin_reg = linear_regression()
shuffle_split = ShuffleSplit(n_splits=5, test_size=0.2, random_state=42)
score = cross_val_score(lin_reg, X, y, cv=shuffle_split,
scoring='------------------------')
```

Which of the following may be appropriate to be filled in the blank space?

Options :

6406533043871. ✔ r2

6406533043872. ✖ neg_r2

6406533043873. ✖ mean_absolute_error

6406533043874. ✔ neg_mean_absolute_error

Question Number : 383 Question Id : 640653903947 Question Type : MSQ Calculator : Yes Correct Marks : 2 Max. Selectable Options : 0

Question Label : Multiple Select Question

Which of the following is a hyper parameter?

Options :

6406533043889. ✖ 'intercept_' in LinearRegression

6406533043890. ✔ 'degree' in 'PolynomialFeatures

6406533043891. ✔ 'n_neighbors' in KNeighborsClassifier

6406533043892. ✔ 'max_depth' of tree in a DecisionTreeClassifier

Question Number : 384 Question Id : 640653903951 Question Type : MSQ Calculator : Yes Correct Marks : 2 Max. Selectable Options : 0

Question Label : Multiple Select Question

Which of the following models are inherently multiclass models?

Options :

6406533043906. ✖ Perceptron

6406533043907. ✔ DecisionTreeClassifier

6406533043908. ✔ KNeighborClassifier

6406533043909. ✖ LogisticRegression

Question Number : 385 Question Id : 640653903969 Question Type : MSQ Calculator : Yes Correct Marks : 2 Max. Selectable Options : 0

Question Label : Multiple Select Question

Which of the following class(es) is/are used to instantiate a neural network in Sklearn.

Options :

6406533043977. ✖ SGDClassifier()

6406533043978. ✔ MLPClassifier()

6406533043979. ✖ NNClassifier()

6406533043980. ✔ MLPRegressor()

Sub-Section Number : 7

Sub-Section Id : 640653134062

Question Shuffling Allowed : Yes

Question Number : 386 Question Id : 640653903949 Question Type : MSQ Calculator : Yes Correct Marks : 3 Max. Selectable Options : 0

Question Label : Multiple Select Question

Which of the following is correct?

Options :

6406533043897. ✔ SGDClassifier(loss="perceptron") is stochastic version of a perceptron model.

6406533043898. ✖ SGDClassifier(loss="hinge") is stochastic version of a DecisionTree classifier model.

6406533043899. ✖ SGDClassifier(loss="multinomial") is stochastic version of a softmax classifier model.

6406533043900. ✔ SGDClassifier(loss="log_loss") is stochastic version of a logistic classifier model.

Question Number : 387 Question Id : 640653903953 Question Type : MSQ Calculator : Yes

Correct Marks : 3 Max. Selectable Options : 0

Question Label : Multiple Select Question

Which of the following processes can be done if we have a dataset with imbalanced target class distribution?

Options :

6406533043914. ✖ Remove all the minority classes

6406533043915. ✔ Up-sample the minority classes

6406533043916. ✖ Remove all the majority classes

6406533043917. ✔ Down-sample the majority classes

6406533043918. ✔ create synthetic samples to balance the classes

Question Number : 388 Question Id : 640653903954 Question Type : MSQ Calculator : Yes

Correct Marks : 3 Max. Selectable Options : 0

Question Label : Multiple Select Question

Which of the following is true about Naive Bayes algorithm ?

Options :

6406533043919. ✖ It is primarily used for regression problems

6406533043920. ✔ It is primarily used for classification problems

6406533043921. ✖ Hyperparameter tuning is required

6406533043922. ✔ Hyperparameter tuning is not required

Question Number : 389 Question Id : 640653903957 Question Type : MSQ Calculator : Yes

Correct Marks : 3 Max. Selectable Options : 0

Question Label : Multiple Select Question

Which of the following is/are correct regarding RadiusNeighborsClassifier

Options :

6406533043931. ✖ Only 5 neighbours in the range of some radius are used to compute the label of a sample.

6406533043932. ✔ All the neighbours in the range of some radius are used to compute the label of a sample.

6406533043933. ✖ It is sensitive to outliers.

6406533043934. ✔ It is not sensitive to outliers.

Question Number : 390 Question Id : 640653903958 Question Type : MSQ Calculator : Yes

Correct Marks : 3 Max. Selectable Options : 0

Question Label : Multiple Select Question

In which of the following cases we should use `partial_fit` instead of fit method ?

Options :

6406533043935. ✔ Data is continuously being generated

6406533043936. ✔ Data is generated every month

6406533043937. ✔ Whole data is generated and its in a huge file size

6406533043938. ✖ For very small dataset

Sub-Section Number : 8

Sub-Section Id : 640653134063

Question Shuffling Allowed : Yes

Question Number : 391 Question Id : 640653903970 Question Type : MSQ Calculator : Yes Correct Marks : 4 Max. Selectable Options : 0

Question Label : Multiple Select Question

The following line of code create a neural network (make necessary assumptions)

```
from sklearn.neural_network import MLPRegressor
regr = MLPRegressor(hidden_layer_size= (5,3),max_iter=5)
regt.fit(X_train, y_train)
```

Select the correct statements from the following list of statements

Options :

6406533043981. ✖ The neural network contains 3 hidden layers with 5 neurons in each hidden layer

6406533043982. ✖ The neural network contains 5 hidden layers with 3 neurons in each hidden layer

6406533043983. ✔ The neural network contains 2 hidden layers with 3 neurons in the second hidden layer

6406533043984. ✔ The neural network contains 2 hidden layers with 5 neurons in the first hidden layer

6406533043985. ✖ All of the given options are correct

Sub-Section Number : 9

Sub-Section Id : 640653134064

Question Shuffling Allowed : Yes

Question Number : 392 Question Id : 640653903963 Question Type : MSQ Calculator : Yes Correct Marks : 5 Max. Selectable Options : 0

Question Label : Multiple Select Question

Consider the following block of code:

```
from sklearn.datasets import load_breast_cancer
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
X,y = load_breast_cancer(as_frame = True, return_X_y = True)
X_train,X_test,y_train,y_test = train_test_split(X,y,
                                                 test_size = 0.2,
                                                 random_state=42)
clf = DecisionTreeClassifier(min_samples_split = 6,
                             min_samples_leaf = 4,
                             random_state = 5)
clf.fit(X_train, y_train)
print(clf.score(X_test, y_test))
```

In which of the following scenarios, the split will can happen at node N?

Options :

6406533043956. ✔ Number of samples at node N = 15. If it is split, it will result in 9 samples in the left child node and 6 sample in the right child node.

6406533043957. ✖ Number of samples at node N = 5. If it is split, it will result in 4 samples in the left child node and 2 samples in the right child node.

6406533043958. ✖ Number of samples at node N = 12. If it is split, it will result in 3 samples in the left child node and 9 samples in the right child node.

6406533043959. ✔ Number of samples at node N = 7. If it is split, it will result in 4 samples in the left child node and 3 samples in the right child node.

Sub-Section Number : 10

Sub-Section Id : 640653134065

Question Shuffling Allowed : Yes

Question Number : 393 Question Id : 640653903943 Question Type : SA Calculator : None Correct Marks : 2

Question Label : Short Answer Question

What will be the output of the following code?: (upto 3 decimals)

```
from sklearn.metrics import mean_absolute_error
y_true = [0.5, 0.2, 0.7, 1]
y_pred = [0.4, 0.2, 1, 0.1]
mean_absolute_error(y_true, y_pred)
```

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

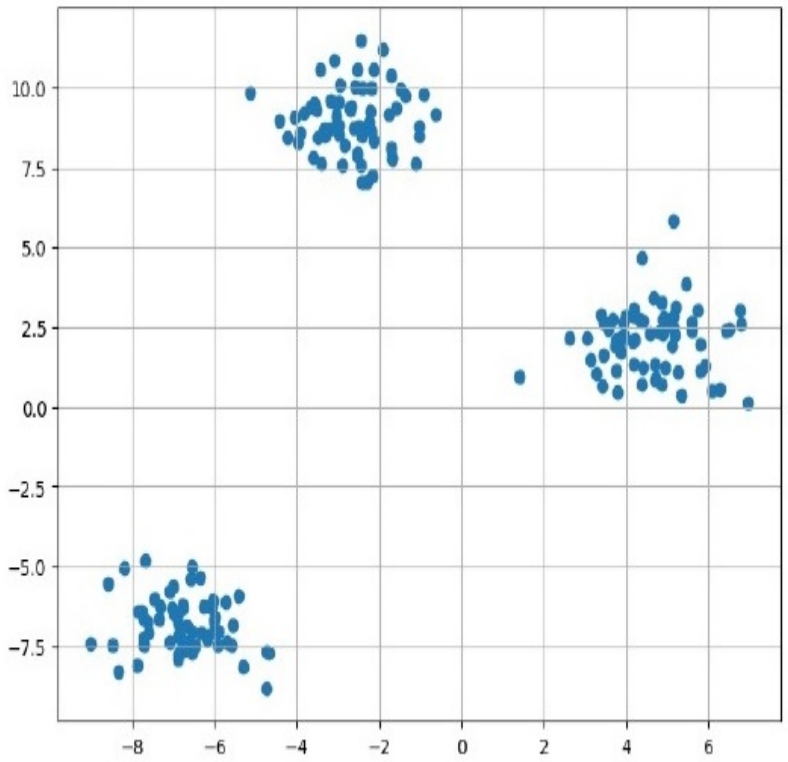
0.319 to 0.328

Sub-Section Number :	11
Sub-Section Id :	640653134066
Question Shuffling Allowed :	Yes

Question Number : 394 Question Id : 640653903967 Question Type : SA Calculator : None Correct Marks : 4

Question Label : Short Answer Question

The plot below displays a distribution of 180 samples with 2 features in Euclidean space.
If we apply the KMeans clustering algorithm to group these data points, what is the value of n_clusters for which the inertia (sum of squared errors) will be exactly zero? If you conclude,



Response Type : Numeric
Evaluation Required For SA : Yes
Show Word Count : Yes
Answers Type : Equal
Text Areas : PlainText
Possible Answers :

180

System commands