Delete all lines with ':' character in the file `file.txt`

6406531885239. ✔ Delete all empty lines before the first non-empty line in the file `file.txt`

6406531885240. ✖ Delete all empty lines after the last non-empty line in the file `file.txt`

MLP

Section Id :	64065338327
Section Number :	10
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	37
Number of Questions to be attempted :	37
Section Marks :	100
Display Number Panel :	Yes
Group All Questions :	No
Enable Mark as Answered Mark for Review and Clear Response :	Yes
Maximum Instruction Time :	0
Sub-Section Number :	1
Sub-Section Id :	64065380405
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Number : 254 Question Id : 640653563971 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

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 <u>TOP</u> FOR THE SUBJECTS REGISTERED BY YOU)

Options :

6406531885261. ✔ YES

6406531885262. ✖ NO

Sub-Section Number :	2
Sub-Section Id :	64065380406
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 255 Question Id : 640653563972 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

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

Options :

6406531885263. ✖ load_openml()

6406531885264. ✖ read_openml()

6406531885265. ✖ read_data()

6406531885266. ✔ fetch_openml()

6406531885267. ✖ load_data()

6406531885268. ✖ load_csv()

6406531885269. ✖ fetch_csv()

Question Number : 256 Question Id : 640653563974 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Why is data preprocessing necessary?

Options :

6406531885274. ✖ Some columns have values only between 0 and 1.

6406531885275. ✔ A column has same entities with different names, e.g. "India", "In", "IN".

6406531885276. ✖ The data has only numbers in all the columns.

Question Number : 257 Question Id : 640653563978 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Consider the following code block:

```python
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 :

6406531885291. ✖
$$[3\ 4\ 5\ 6\ 7\ 8]\ [0\ 3\ 4]$$
$$[0\ 1\ 2\ 6\ 7\ 8]\ [2\ 4\ 6]$$
$$[0\ 1\ 2\ 3\ 4\ 5]\ [0\ 5\ 8]$$

6406531885292. ✖
$$[3\ 4\ 5]\ [6\ 7\ 8]\ [0\ 1\ 2]$$
$$[0\ 1\ 2]\ [6\ 7\ 8]\ [3\ 4\ 5]$$
$$[0\ 1\ 2]\ [3\ 4\ 5]\ [6\ 7\ 8]$$

6406531885293. ✖
$$[0\ 1\ 2]\ [3\ 4\ 5\ 6\ 7\ 8]$$
$$[3\ 4\ 5]\ [0\ 1\ 2\ 6\ 7\ 8]$$
$$[6\ 7\ 8]\ [0\ 1\ 2\ 3\ 4\ 5]$$

6406531885294. ✔
$$[3\ 4\ 5\ 6\ 7\ 8]\ [0\ 1\ 2]$$
$$[0\ 1\ 2\ 6\ 7\ 8]\ [3\ 4\ 5]$$
$$[0\ 1\ 2\ 3\ 4\ 5]\ [6\ 7\ 8]$$

Question Number : 258 Question Id : 640653563979 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Which of the following sklearn classes will be the most suitable for examining the effect of the number of samples on the training and testing errors?

Options :

6406531885295. ✖ `sklearn.model_selection.train_test_split`

6406531885296. ✖ `sklearn.model_selection.cross_validate`

6406531885297. ✖ `sklearn.model_selection.cross_val_score`

6406531885298. ✔ `sklearn.model_selection.learning_curve`

Question Number : 259 Question Id : 640653563983 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Which of the following is not a hyper parameter?

Options :

6406531885308. ✖ degree in Polynomial Regression

6406531885309. ✖ k in KNN

6406531885310. ✔ intercept value in linear regression

6406531885311. ✖ depth of tree in a decision tree

Question Number : 260 Question Id : 640653563987 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction

Correct Marks : 2

Question Label : Multiple Choice Question

Consider following two statements:

Statement 1: The multinomial Naive Bayes classifier is suitable for classification with discrete features

Statement 2: Two dependent features impact GaussianNB performance because internally it calculates the conditional probability.

Options :

6406531885325. ✔ Both statements are True

6406531885326. ✖ Only statement 1 is True

6406531885327. ✖ Only statement 2 is True

6406531885328. ✖ Both statements are False

Question Number : 261 Question Id : 640653563990 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

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 ?

```python
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(X,y)
```

Options :

6406531885337. ✖ max_depth

6406531885338. ✖ C

6406531885339. ✔ chunksize

6406531885340. ✖ warm_start

Question Number : 262 Question Id : 640653563991 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

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

Options :

6406531885341. ✖ It does not create hyperplanes as decision boundary

6406531885342. ✔ It can only work for linearly separable data

6406531885343. ✖ It is robust to outliers

6406531885344. ✖ It is mostly used for clustering the data

Question Number : 263 Question Id : 640653563994 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

What will 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.get_feature_names_out())
```

Options :

6406531885353. ✔ ['document' 'first' 'is' 'second' 'the' 'this']

6406531885354. ✖ {'this': 5, 'is': 2, 'the': 4, 'first': 1, 'document': 0, 'second': 3}

6406531885355. ✖ [3,1,2,1,2,2]

6406531885356. ✖ [0,1,2,3,4,5]

Sub-Section Number :	3
Sub-Section Id :	64065380407
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 264 Question Id : 640653563975 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3

Question Label : Multiple Choice Question

Which of the following APIs can be used to construct an ML pipeline that has numerical and categorical features both? Choose the most suitable answer.

Options :

6406531885277. ✖ Pipeline alone

6406531885278. ✖ ColumnTransformer with pipeline

6406531885279. ✔ FeatureUnion with ColumnTransformer and/or pipeline

6406531885280. ✖ All of these

Question Number : 265 Question Id : 640653563976 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3

Question Label : Multiple Choice Question

Consider the following preprocessing steps:

1. Read the data from a file (named 'dataset.csv'). It has 4 columns, with following column names 'city', 'title', 'expert_rating' and 'user_rating' in this order.

2. Drop rows with missing values.

3. Apply 'CountVectorizer' on 'title'.

4. Apply one hot encoding to 'city'.

5. Drop the remaining columns.

6. Fit and transform the data and print it.

Which of the following code snippets correctly accomplishes the above task? Assume necessary imports.

Options :

6406531885281. ✔

```
data = pd.read_csv('dataset.csv')
data.dropna()
column_trans = ColumnTransformer(
    [('categories', OneHotEncoder(dtype='int'), ['city']),
     ('title_bow', CountVectorizer(), 'title')],
    remainder='drop')

print(column_trans.fit_transform(data))
```

6406531885282. ✖

6406531885283.

```
data = pd.read_csv('dataset.csv')
data = data.dropna()
column_trans = ColumnTransformer(
    [('categories', OneHotEncoder(dtype='int'), ['Title']),
     ('title_bow', CountVectorizer(), 'City')],
    remainder='drop')

print(column_trans.fit_transform(data))
```

```
data = pd.read_csv('dataset.csv')
data = data.dropna()
column_trans = ColumnTransformer(
    [('categories', OneHotEncoder(dtype='int'), ['city']),
     ('title_bow', CountVectorizer(), 'title')],
    remainder='passthrough')

print(column_trans.fit_transform(data))
```
6406531885284. ✖

```
data = pd.read_csv('dataset.csv')
data = data.dropna()
column_trans = ColumnTransformer(
    [('categories', OneHotEncoder(dtype='int'), ['city']),
     ('title_bow', CountVectorizer(), 'title')],
    remainder='drop')

print(column_trans.fit(data))
```
6406531885285. ✖

6406531885286. ✖ None of these

Question Number : 266 Question Id : 640653563981 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3

Question Label : Multiple Choice Question

Consider the following code block:

```python
from sklearn.datasets import make_regression
X, y = make_regression(n_samples = 1000,
                       n_features = 5,
                       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?:

[1.68059576e+01, 1.89752021e+01, 7.49212536e-04, -6.53455275e-04, 3.01471918e-04]

[16.82258106, 18.99248887, 0., 0., 0.]

Options :

6406531885300. ✖ 'l1', 'l2'

6406531885301. ✖ 'l1', None

6406531885302. ✔ 'l2', 'l1'

6406531885303. ✖ 'l2', None

Question Number : 267 Question Id : 640653563982 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3

Question Label : Multiple Choice Question

What will be the output of the following code?:

```python
from sklearn.datasets import make_regression
X, y = make_regression(n_samples = 10, n_features = 2)
from sklearn.preprocessing import PolynomialFeatures
poly_transform = PolynomialFeatures(degree=2, interaction_only=True)

X_trans = poly_transform.fit_transform(X)
print(X_trans.shape)
```

Options :

6406531885304. ✷ (10, 2)

6406531885305. ✔ (10, 4)

6406531885306. ✷ (10, 5)

6406531885307. ✷ (10, 6)

Question Number : 268 Question Id : 640653563985 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3

Question Label : Multiple Choice Question

Consider following model:

```python
estimator = SGDClassifier(loss='log',
                          penalty='l2',
                          max_iter=1,
                          warm_start=True,
                          eta0=0.01,
                          alpha=0,
                          learning_rate='constant',
                          random_state=1729)
pipe_sgd= make_pipeline(MinMaxScaler(), estimator)
```

Which of the following code snippets will plot the learning curve for training for 100 epochs? Assume necessary imports and the variable names suggest the data they hold/-point to.

Options :

6406531885316. ✔
```python
Loss=[]
iterations= 100
for i in range(iterations):
    pipe_sgd.fit(x_train,y_train)
    y_pred = pipe_sgd.predict_proba(x_train)
    Loss.append(log_loss(y_train,y_pred))
plt.plot(np.arange(iterations),Loss)
```

6406531885317. ✖
```python
Loss=[]
iterations= 100
for i in range(iterations):
    pipe_sgd.fit(x_train,y_train)
    y_pred = pipe_sgd.predict(x_train)
    Loss.append(log_loss(y_train,y_pred))
plt.plot(np.arange(iterations),Loss)
```

6406531885318. ✖
```python
Loss=[]
iterations= 100
for i in range(iterations):

    pipe_sgd.fit(x_train,y_train)
    y_pred = pipe_sgd.predict(x_test)
    Loss.append(log_loss(y_train,y_pred))
plt.plot(np.arange(iterations),Loss)
```

6406531885319. ✖

```
Loss=[]
iterations= 100
for i in range(iterations):
    pipe_sgd.fit(x_train,y_train)
    y_pred = pipe_sgd.predict(x_test)
    Loss.append(log_loss(y_test,y_pred))
plt.plot(np.arange(iterations),Loss)
```

6406531885320. ✖ None of these

Question Number : 269 Question Id : 640653563998 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3

Question Label : Multiple Choice Question

If we need quick results during the testing phase, then which classification techniques may not be appropriate?

Options :

6406531885369. ✖ sklearn.linear_model.LogisticRegression

6406531885370. ✔ sklearn.neighbors.KNeighborsClassifier

6406531885371. ✖ sklearn.tree.DecisionTreeClassifier

6406531885372. ✖ sklearn.naive_bayes.GaussianNB

Question Number : 270 Question Id : 640653564002 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3

Question Label : Multiple Choice Question

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

Options :

6406531885386. ✖ 'degree'

6406531885387. ✖ 'sgd_degree'

6406531885388. ✖ 'sgd __degree'

6406531885389. ✖ 'poly_degree'

6406531885390. ✔ 'poly__degree'

Sub-Section Number :	4
Sub-Section Id :	64065380408
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 271 Question Id : 640653563997 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

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 :

6406531885365. ✔ Output score is likely to increase.

6406531885366. ✖ Output score is likely to decrease.

6406531885367. ✖ Output score may increase or decrease.

6406531885368. ✖ Output score will remain the same.

Question Number : 272 Question Id : 640653563999 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 4

Question Label : Multiple Choice Question

Consider the following code. How many 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 :

6406531885373. ✻ 20

6406531885374. ✔ 75

6406531885375. ✻ 8

6406531885376. ✻ 15

6406531885377. ✻ 40

Question Number : 273 Question Id : 640653564001 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 4

Question Label : Multiple Choice Question

Consider two classifiers as shown in the following block of code:

```python
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 = 0.2,
                                        random_state = 1)
clf1 = DecisionTreeClassifier(min_samples_split = 7,
                         min_samples_leaf = 4,
                         random_state = 5)
clf1.fit(X_train, y_train)

clf2 = DecisionTreeClassifier(min_samples_split = 4,
                         min_samples_leaf = 2,
                         random_state = 5)
clf2.fit(X_train, y_train)
```

What can we say about the depths of the classifiers clf1 and clf2?

Options :

6406531885382. ✱ depth(clf1) ≥ depth(clf2)

6406531885383. ✔ depth(clf1) ≤ depth(clf2)

6406531885384. ✱ depth(clf1) = depth(clf2)

6406531885385. ✱ Insufficient Information

Sub-Section Number :   5

Sub-Section Id :   64065380409

Question Shuffling Allowed :   Yes

Is Section Default? :   null

Question Number : 274 Question Id : 640653563973 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Selectable Option : 0

Question Label : Multiple Select Question

Consider the following code:

```python
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 45 samples?

Options :

6406531885270. ✔
```python
train_X, test_X, train_y, test_y = train_test_split(X, y,
test_size=45, random_state=42)
```

6406531885271. ✔
```python
train_X, test_X, train_y, test_y = train_test_split(X, y,
train_size=105, random_state=42)
```

6406531885272. ✔
```python
train_X, test_X, train_y, test_y = train_test_split(X, y,
test_size=0.3, random_state=42)
```

```
                    train_X, test_X, train_y, test_y = train_test_split(X, y,
6406531885273. ✔   train_size=0.7, random_state=42)
```

Question Number : 275 Question Id : 640653563977 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Selectable Option : 0

Question Label : Multiple Select Question

Consider the following code block with respect to some dataset contained in X and y.

```
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 :

6406531885287. ✔ r2

6406531885288. ✖ neg_r2

6406531885289. ✖ mean_absolute_error

6406531885290. ✔ neg_mean_absolute_error

Question Number : 276 Question Id : 640653563986 Question Type : MSQ Is Question

Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Selectable Option : 0

Question Label : Multiple Select Question

Which of the following models are inherently multiclass models?

Options :

6406531885321. ✔ KNN

6406531885322. ✔ Decision trees

6406531885323. ✖ Perceptron

6406531885324. ✖ Logistic regression

Question Number : 277 Question Id : 640653563992 Question Type : MSQ Is Question

Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Selectable Option : 0

Question Label : Multiple Select Question

Which of the following is/are correct regarding Radius Neighbors Classifier

Options :

6406531885345. ✖ Only 5 neighbours in the range of some radius are used to compute the label of a sample.

6406531885346. ✔ All the neighbours in the range of some radius are used to compute the label of a sample.

6406531885347. ✖ It is sensitive to outliers.

6406531885348. ✔ It is not sensitive to outliers.

Question Number : 278 Question Id : 640653563996 Question Type : MSQ Is Question

Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Selectable Option : 0

Question Label : Multiple Select Question

Which options are correct for large datasets that don't fit into the system's main memory?

Options :

6406531885361. ✻ No data preprocessing can be performed.

6406531885362. ✔ One hot encoding can not be applied directly or by iteratively learning one hot encoder's parameters and then applying one hot encoding in batches.

6406531885363. ✔ Standard scaling parameters can be learnt iteratively then standard scaling can be applied in batches.

6406531885364. ✻ One hot encoding and standard scaling can be applied directly or iteratively.

Question Number : 279 Question Id : 640653564006 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Selectable Option : 0

Question Label : Multiple Select Question

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

Options :

6406531885400. ✻ SGDClassifier()

6406531885401. ✔ MLPClassifier()

6406531885402. ✻ NNClassifier()

6406531885403. ✔ MLPRegressor()

Sub-Section Number :	6
Sub-Section Id :	64065380410
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 280 Question Id : 640653563984 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3 Selectable Option : 0

Question Label : Multiple Select Question

Which of the following is correct?

Options :

6406531885312. ✔ SGDClassifier(loss="perceptron") is stochastic version of a perceptron model

6406531885313. ✘ SGDClassifier(loss="percept") is stochastic version of a perceptron model

6406531885314. ✔ SGDClassifier(loss="log_loss") is stochastic version of a logistic classifier model

6406531885315. ✔ SGDClassifier() is stochastic version of a SVM model

Question Number : 281 Question Id : 640653563988 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3 Selectable Option : 0

Question Label : Multiple Select Question

Which of the following processes should be done if we get an imbalanced dataset? Data is imbalanced when the target class has an uneven distribution of label values.

Options :

6406531885329. ✘ Remove all the minority classes

6406531885330. ✘ Remove all the majority classes

6406531885331. ✔ Up-sample the minority classes

6406531885332. ✔ down-sample the majority classes

Question Number : 282 Question Id : 640653563989 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3 Selectable Option : 0

Question Label : Multiple Select Question

Which of the following is true about Naive Bayes algorithm ?

Options :

6406531885333. ✖ It is primarily used for regression problems

6406531885334. ✔ It is primarily used for classification problems

6406531885335. ✖ Hyperparameter tuning is required

6406531885336. ✔ Hyperparameter tuning is not required

Question Number : 283 Question Id : 640653563993 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3 Selectable Option : 0

Question Label : Multiple Select Question

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

Options :

6406531885349. ✔ Data is continuously being generated

6406531885350. ✔ Data is generated every month

6406531885351. ✔ Whole data is generated and its in a huge file size

6406531885352. ✖ For very small dataset

Question Number : 284 Question Id : 640653563995 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3 Selectable Option : 0

Question Label : Multiple Select Question

Which options are correct for Support Vectors in SVM ?

Options :

6406531885357. ✔ Support vectors are the data points nearest to the hyperplane

6406531885358. ✔ Using these support vectors, we maximize the margin of the classifier.

6406531885359. ✖ Using these support vectors, we minimize the margin of the classifier.

6406531885360. ✖ None of these

Sub-Section Number :	7
Sub-Section Id :	64065380411
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 285 Question Id : 640653564000 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 4 Selectable Option : 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 = 1)

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 be done at node N?

Options :

6406531885378. ✔ Number of samples at node N = 15. If it is split, it will result in 9 nodes in the left child and 6 nodes in the right child.

6406531885379. ✖ Number of samples at node N = 5. If it is split, it will result in 4 nodes in the left child and 2 nodes in the right child.

6406531885380. ✔ Number of samples at node N = 7. If it is split, it will result in 4 nodes in the left child and 3 nodes in the right child.

6406531885381. ✖ Number of samples at node N = 12. If it is split, it will result in 3 nodes in the left child and 9 nodes in the right child.

Question Number : 286 Question Id : 640653564004 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 4 Selectable Option : 0

Question Label : Multiple Select Question

Suppose we want to choose the best value of k in $k-$means clustering algorithm using Silhouette Coefficient values. Which of the following are required to compute the coefficient value? Note: Labels (cluster number) of the samples are random. There is no functional relationship between a sample and its label

Options :

6406531885392. ✔ All the samples in the dataset

6406531885393. ✖ A few randomly selected samples in the dataset

6406531885394. ✔ The number of clusters k, such that $k \geq 2$

6406531885395. ✖ The number of clusters k, such that $k \geq 1$

Question Number : 287 Question Id : 640653564005 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 4 Selectable Option : 0

Question Label : Multiple Select Question

Which of the following approach(es) is(are) helpful to find a good value for k in k−means clustering algorithm?

Options :

6406531885396. ✔ Plotting an Elbow curve

6406531885397. ✖ Using GridSearchCV or RandomizedSearchCV for various values of k

6406531885398. ✔ Plotting Silhouette coefficient for various values of k

6406531885399. ✖ Using k-fold cross validation

Question Number : 288 Question Id : 640653564007 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 4 Selectable Option : 0

Question Label : Multiple Select Question

The following line of code create a neural network (assume necessary imports)

```
regr = MLPRegressor(hidden_layer_size= (5,3),max_iter=5).fit(X_train, y_train)
```

Select the correct statements from the following list of statements

Options :

6406531885404. ✱ The neural network contains 3 hidden layers with 5 neurons in each hidden layer

6406531885405. ✱ The neural network contains 5 hidden layers with 3 neurons in each hidden layer

6406531885406. ✔ The neural network contains 2 hidden layers with 3 neurons in the second hidden layer

6406531885407. ✔ The neural network contains 2 hidden layers with 5 neurons in the first hidden layer

6406531885408. ✱ None of the given options are correct

Sub-Section Number :	8
Sub-Section Id :	64065380412
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 289 Question Id : 640653563980 Question Type : SA Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Short Answer Question

What will be the output of the following code?:

```
from sklearn.metrics import max_error
y_true = [0.5, 0.2, 0.7, 1]
y_pred = [0.4, 0.2, 1, 0.1]
max_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 :

Sub-Section Number :   9

Sub-Section Id :   64065380413

Question Shuffling Allowed :   Yes

**Is Section Default? :**   null

Question Number : 290 Question Id : 640653564003 Question Type : SA Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 4

Question Label : Short Answer Question

The plot below shows a distribution of 200 samples with 2 features in Euclidean space. Suppose we use the K-means clustering algorithm to group these data points into individual clusters. Enter the value of K for which the inertia (Sum Square Error) will be zero. If you conclude, there is no such value for K, then enter -1.
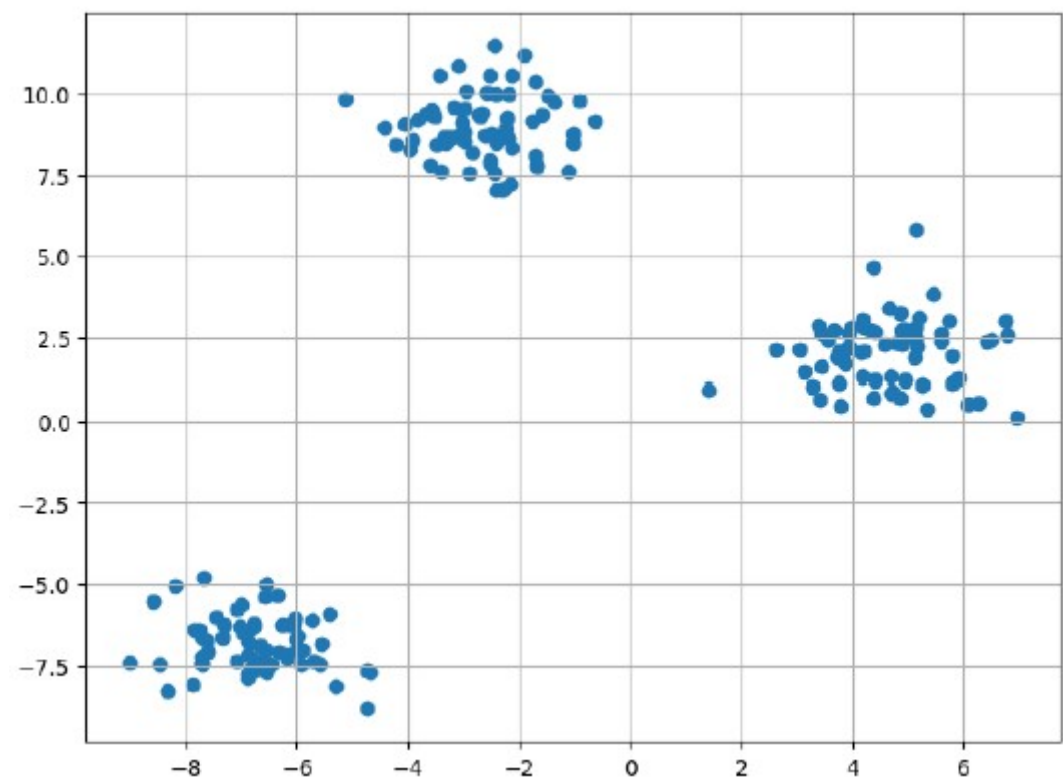
Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

200

PDSA

Section Id :	64065338328
Section Number :	11
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	28
Number of Questions to be attempted :	28
Section Marks :	100
Display Number Panel :	Yes
Group All Questions :	No
Enable Mark as Answered Mark for Review and Clear Response :	Yes
Maximum Instruction Time :	0
Sub-Section Number :	1
Sub-Section Id :	64065380414
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Number : 291 Question Id : 640653564008 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 0

Question Label : Multiple Choice Question

THIS IS QUESTION PAPER FOR THE SUBJECT "DIPLOMA LEVEL : PROGRAMMING, DATA STRUCTURES AND ALGORITHMS USING PYTHON (COMPUTER BASED EXAM)"