Correct Marks : 8

Question Label : Multiple Choice Question

What will be the output of the following command?

```
echo {a,b,c,c}{c,c,e,f} | tr ' ' '\n' | sort | uniq | awk '
{
    n += NR
}
END {
    print n
}
'
```

Options :

6406531892394. ✔ 45

6406531892395. ✖ 78

6406531892396. ✖ 136

6406531892397. ✖ 55

MLP

Section Id :	64065338407
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 :	64065380971
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Number : 255 Question Id : 640653566235 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 TOP FOR THE SUBJECTS REGISTERED BY YOU)

Options :

6406531892406. ✔ YES

6406531892407. ✖ NO

Sub-Section Number :	2
Sub-Section Id :	64065380972
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 256 Question Id : 640653566236 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

_____ generates a bunch of normally-distributed clusters of points with specific mean and standard deviations for each cluster.

Options :

6406531892408. ✖ sklearn.datasets.make_clusters()

6406531892409. ✖ sklearn.datasets.make_centers()

6406531892410. ✖ sklearn.datasets.make_normal_clusters()

6406531892411. ✔ sklearn.datasets.make_blobs()

Question Number : 257 Question Id : 640653566238 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 :

6406531892416. ✖ Some columns have values only between 0 and 1

6406531892417. ✔ The data is divided into multiple types of files i.e. html, csv, tsv, etc.

6406531892418. ✖ The data has only numbers in all the columns.

Question Number : 258 Question Id : 640653566241 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 correct with respect to R2-score?

Options :

6406531892429. ✖ R2 score is always positive and it may go up to infinity.

6406531892430. ✖ R2 score is always positive, but it ranges between 0 and 1 only.

6406531892431. ✔ R2 score can be negative. That happens if our model is worse than the mean model.

6406531892432. ✖ R2 score can be negative. That happens if the mean model is worse than our model.

Question Number : 259 Question Id : 640653566242 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 options represent the output of the following code block?

```
from sklearn.metrics import confusion_matrix
y_true = ["Rainy", "Sunny", "Sunny", "Rainy", "Sunny", "Rainy"]
y_pred = ["Sunny", "Sunny", "Rainy", "Rainy", "Sunny", "Sunny"]
confusion_matrix(y_true, y_pred, labels=["Rainy", "Sunny"])
```

Options :

6406531892433. ✖
```
array([[2, 1],
       [2, 1]])
```

6406531892434. ✖
```
array([[2, 1],
       [1, 2]])
```

6406531892435. ✔
```
array([[1, 2],
       [1, 2]])
```

6406531892436. ✖

```
array([[1, 2],
       [2, 1]])
```

Question Number : 260 Question Id : 640653566243 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:

```
from sklearn.datasets import make_regression
X, y = make_regression(n_samples = 10, n_features = 3)

from sklearn.model_selection import LeavePOut
lpo = LeavePOut(p = 2)
count = 0
for train, test in lpo.split(X):
  print(train, test)
  count += 1
print(count)
```

What will be the value of the 'count'?

Options :

6406531892437. ✖ 20

6406531892438. ✖ 30

6406531892439. ✔ 45

6406531892440. ✖ 15

Question Number : 261 Question Id : 640653566244 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:

```
X = ['a', 'b', 'c', 'd', 'e', 'f']
from sklearn.model_selection import RepeatedKFold
rkf = RepeatedKFold(n_splits = 3, n_repeats = 2, random_state = 10)
for train, test in rkf.split(X):
  print(train, test)
```

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

Options :

6406531892441. ✖

```
[0 1] [3 4] [2 5]
[1 2] [4 5] [0 3]
[0 2] [3 5] [1 4]
[1 3] [4 5] [0 2]
[0 2] [3 4] [1 5]
[0 1] [2 5] [3 4]
```

6406531892442. ✖

```
[0 1 3 4] [3 5]
[1 2 4 5] [0 2]
[0 2 3 5] [1 5]
[1 3 4 5] [0 4]
[0 2 3 4] [1 0]
[0 1 2 5] [3 2]
```

6406531892443. ✔

```
[0 1 3 4] [2 5]
[1 2 4 5] [0 3]
[0 2 3 5] [1 4]
[1 3 4 5] [0 2]
[0 2 3 4] [1 5]
[0 1 2 5] [3 4]
```

6406531892444. ✖

[0 1 1 4] [2 5]
[1 2 5 5] [0 3]
[0 2 0 5] [1 4]
[1 3 1 5] [0 2]
[0 2 0 4] [1 5]
[0 1 5 5] [3 4]

6406531892445. ✖

[0 1] [3 4] [0 1]
[1 2] [1 2] [0 3]
[0 2] [3 5] [3 5]
[1 3] [4 5] [1 3]
[0 2] [3 4] [0 2]
[0 1] [0 1] [3 4]

Question Number : 262 Question Id : 640653566247 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 a hyperparameter?

Options :

6406531892452. ✖ L1-ratio in elasticnet

6406531892453. ✖ Pruning parameter in a decision tree

6406531892454. ✖ Learning rate in SGDRegressor

6406531892455. ✔ All of these

Question Number : 263 Question Id : 640653566255 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 estimators can implement the `partial_fit` method ?

Options :

6406531892484. ✖ DecisionTreeClassifier

6406531892485. ✖ RandomForestRegressor

6406531892486. ✖ LogisticRegressor

6406531892487. ✔ SGDRegressor

Question Number : 264 Question Id : 640653566256 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

```
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression

pipe = ([
    ('scaler', StandardScaler()),
    ('softmax', ____________ )
])
```

Which of the following options is True for blank space if I want to train the above pipeline as softmax regression ?

Options :

6406531892488. ✖ `LogisticRegression(solver = 'sag')`

6406531892489. ✖ `LogisticRegression(multi_class = 'ovr')`

6406531892490. ✖ LogisticRegression(solver = 'lbfgs')

6406531892491. ✔ LogisticRegression(multi_class = 'multinomial')

Question Number : 265 Question Id : 640653566258 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:

```
from sklearn.metrics import  f1_score
y_true = [1,1,0,1,0,0,1,0,1]
y_pred = [0,1,0,1,0,1,1,1,1]
print(f1_score(y_true,y_pred))
```

What will be the output?

Options :

6406531892496. ✖ 0.66

6406531892497. ✔ 0.72

6406531892498. ✖ 0.80

6406531892499. ✖ 1.00

Question Number : 266 Question Id : 640653566259 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 are correct about sklearn.svm.SVC:

```
from sklearn.svm import SVC
clf = SVC(C=1).fit(X_train, y_train)
print(clf.support_)
```

What will be the output?

Options :

6406531892500. ✖ It will print number of support vectors

6406531892501. ✔ It will print an array of support vectors

6406531892502. ✖ It will print an array of probabilities representing distance from decision boundary with each data point.

6406531892503. ✖ It will print indices of the support vectors from the training set.

Question Number : 267 Question Id : 640653566260 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 below code which of the following option is true for that

```
from sklearn.neighbors import NearestNeighbors
neigh = NearestNeighbors(n_neighbors=1)
neigh.fit(X_train)
print(neigh.kneighbors(X_test[1:2,:]))
```

Assume X_train and X_test are of type numpy.ndarray .

Options :

6406531892504. ✖ It will print nearest neighbours from the test point

6406531892505. ✔ It will print indices of and distances to the neighbouring points (in training set) from test point

6406531892506. ✖ It will print indices, distance and nearest training point from the test point

6406531892507. ✖ It will throw an error

Sub-Section Number :	3
Sub-Section Id :	64065380973
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 268 Question Id : 640653566239 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 only supports conjoining transformers and estimators in series (i.e. one after another)?

Options :

6406531892419. ✔ Pipeline

6406531892420. ✖ ColumnTransformer

6406531892421. ✖ FeatureUnion

6406531892422. ✖ All of these

Question Number : 269 Question Id : 640653566240 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 ML task/steps for a regression dataset:

1. Read the data from a file (named 'dataset.csv'). It has 7 columns. The last column is the target variable, all 6 features numerical.
2. Remove rows which have target values missing.
3. Fill the missing values in the features by KNN using 3 nearest neighbours.
4. Split the data into training and test sets. Take randomly the 70% of rows in the training set and

the rest of them into the test set.

5. Train a simple linear regression model, with intercept, on the training set.

6. Report R2 score on the test set.

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

Options :

6406531892423. ✔

```
data = pd.read_csv('dataset.csv')
data = data.dropna()
X, y = data[data.columns[:-1]],data[data.columns[-1]]
X = X[~y.isna()]
y = y.dropna()
X_train, X_test, y_train, y_test = train_test_split(X,y,
test_size=0.7)
pipe = Pipeline([('imputer', KNNImputer(n_neighbors  = 3)),
                 ('estimator', LinearRegression())])

pipe.fit(X_train,y_train)
print(pipe.score(X_test, y_test))
```

6406531892424. ✖

```
data = pd.read_csv('dataset.csv')
data = data.dropna()
X, y = data[data.columns[:-1]],data[data.columns[-1]]
X = X[y.isna()]
y = y.dropna()
X_train, X_test, y_train, y_test = train_test_split(X,y,
                                                    test=0.3)
pipe = Pipeline([('imputer', KNNImputer(n_neighbors  = 3)),
                 ('estimator', LinearRegression())])

pipe.fit(X_train,y_train)
print(pipe.score(X_test, y_test))
```

6406531892425. ✖

```
data = pd.read_csv('dataset.csv')
data = data.dropna()
X, y = data[data.columns[:-1]],data[data.columns[-1]]
X = X[~y.isna()]
y = y[y.isna()]
X_train, X_test, y_train, y_test = train_test_split(X,y,
                                                    train_size=0.8)
pipe = Pipeline([('imputer', KNNImputer(n_neighbors  = 3)),
                 ('estimator', LinearRegression())])

pipe.fit(X_train,y_train)
print(pipe.score(y_test, y_test))
```

6406531892426. ✖

```
data = pd.read_csv('dataset.csv')
data = data.dropna()
X, y = data[data.columns[:-1]],data[data.columns[-1]]
X = X[~y.isna()]
y = y.dropna()
X_train, X_test, y_train, y_test = train_test_split(X,y,
                                                    test=0.2)
pipe = Pipeline([('imputer', KNNImputer(n_neighbors  = 2)),
('estimator', LinearRegression(fit_intercept=False))])

pipe.fit(X_train,y_train)
print(pipe.score(X_test, X_test))
```

6406531892427. ✖

```
data = pd.read_csv('dataset.csv')
data = data.dropna()
X, y = data[data.columns[:-1]],data[data.columns[-1]]

X_train, X_test, y_train, y_test = train_test_split(X,y,
                                                    test_size=0.2)
pipe = Pipeline([('imputer', KNNImputer(n_neighbors  = 3)),
('estimator', LinearRegression())])

pipe.fit(X_train,y_train)
print(pipe.score(X_test, X_test))
```

6406531892428. ✖ None of these

Question Number : 270 Question Id : 640653566246 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?:

```
from sklearn.datasets import make_regression
from sklearn.preprocessing import PolynomialFeatures
X, y = make_regression(n_samples = 10, n_features = 3)
poly_transform = PolynomialFeatures(degree=2, interaction_only=True)
X_trans = poly_transform.fit_transform(X)
print(X_trans.shape)
```

Options :

6406531892447. ✖ (10, 3)

6406531892448. ✔ (10, 7)

6406531892449. ✖ (10, 8)

6406531892450. ✖ (10, 10)

6406531892451. ✖ (10, 11)

Question Number : 271 Question Id : 640653566249 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 statements about SGDClassifier:

1. It can be used to train a model on large dataset that doesn't fit in main memory
2. It can emulate a KNN model
3. It can emulate a decision tree model
4. It can emulate a perceptron

Choose the correction option(s)

Options :

6406531892460. ✔ 1 and 4

6406531892461. ✖ 1 and 2

6406531892462. ✖ 2 and 3

6406531892463. ✖ 3 and 4

Question Number : 272 Question Id : 640653566250 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 an image classification task: an image can have dogs, birds and trees. An image can have any combination of these three. The classifier is expected to report all these three for every sample. What kind of classification problem is this?

Options :

6406531892464. ✔ multi-label and multiclass problem.

6406531892465. ✖ multi-label and binary class problem.

6406531892466. ✖ multiclass problem.

6406531892467. ✖ binary class single label problem.

Question Number : 273 Question Id : 640653566254 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 code snippets, assuming necessary imports.

```
model1 = KNeighborsClassifier(n_neighbors=2)
model2 = KNeighborsClassifier(n_neighbors=5)

model1.fit(X_train,y_train)
model2.fit(X_train,y_train)
```

Choose the correct options:

Options :

6406531892480. ✔ `model1` will have smoother decision boundary compared to model `model2`

6406531892481. ✖ `model2` will have smoother decision boundary compared to model `model1`

6406531892482. ✖ `model1` will have same decision boundary compared as model `model2`

6406531892483. ✖ No comparison can be made between decision boundaries of `model1` and `model2`

Question Number : 274 Question Id : 640653566257 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 options is true for the gamma parameter in a non-linear soft margin SVM?

Options :

6406531892492. ✔ For high values of gamma, the points need to be very close to each other in order to be considered in the same class

6406531892493. ✖ For low values of gamma, the points need to be very close to each other in order to be considered in the same class

6406531892494. ✖ Gamma doesn't affect the SVM model at all

6406531892495. ✖ None of these

Question Number : 275 Question Id : 640653566261 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:

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

dtc1 = DecisionTreeClassifier(ccp_alpha = 0.0)
dtc1.fit(X, y)
dtc2 = DecisionTreeClassifier(ccp_alpha = 0.06)
dtc2.fit(X, y)

dtc3 = DecisionTreeClassifier(ccp_alpha = 0.1)
dtc3.fit(X, y)

dtc4 = DecisionTreeClassifier(ccp_alpha = 0.03)
dtc4.fit(X, y)
```

Which model is likely to overfit the most?

Options :

6406531892508. ✔ dtc1

6406531892509. ✖ dtc2

6406531892510. ✖ dtc3

6406531892511. ✖ dtc4

Question Number : 276 Question Id : 640653566265 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 n_estimators parameter of a Bagging Classifier model.

```
from sklearn.model_selection import GridSearchCV
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import BaggingClassifier

param_grid = [
{_____: [200, 300, 400, 500, 600]}
]
pipeline = Pipeline(steps=[('scaler', StandardScaler()),
('bc', BaggingClassifier())])
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 :

6406531892526. ✖ 'n_estimators'

6406531892527. ✖ 'bc_n_estimators'

6406531892528. ✔ 'bc__n_estimators'

6406531892529. ✖ 'bc___n_estimators'

6406531892530. ✖ 'bc.n_estimators'

Sub-Section Number :	4
Sub-Section Id :	64065380974
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 277 Question Id : 640653566262 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 = 10)
gs.fit(X,y)
```

Options :

6406531892512. ✖ 20

6406531892513. ✖ 200

6406531892514. ✖ 8

6406531892515. ✖ 150

6406531892516. ✖ 15

6406531892517. ✔ 80

Question Number : 278 Question Id : 640653566264 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:

```
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 = 3, min_samples_leaf = 2,
                              random_state = 5)
clf1.fit(X_train, y_train)

clf2 = DecisionTreeClassifier(min_samples_split = 6, min_samples_leaf = 4,
                              random_state = 5)
clf2.fit(X_train, y_train)
```

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

Options :

6406531892522. ✔ depth(clf1) ≥ depth(clf2)

6406531892523. ✖ depth(clf1) ≤ depth(clf2)

6406531892524. ✖ depth(clf1) = depth(clf2)

6406531892525. ✖ Insufficient Information

Question Number : 279 Question Id : 640653566266 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:

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

dtc1 = DecisionTreeClassifier(ccp_alpha = 0.0)
dtc1.fit(X, y)

dtc2 = DecisionTreeClassifier(ccp_alpha = 0.03)
dtc2.fit(X, y)

dtc3 = DecisionTreeClassifier(ccp_alpha = 0.06)
dtc3.fit(X, y)

dtc4 = DecisionTreeClassifier(ccp_alpha = 0.1)
dtc4.fit(X, y)

d1 = dtc1.get_depth()
d2 = dtc2.get_depth()
d3 = dtc3.get_depth()
d4 = dtc4.get_depth()
```

What can we say about d1, d2, d3 and d4?

Options :

6406531892531. ✖ d1 < d2 < d3 < d4

6406531892532. ✖ d1 ≤ d2 ≤ d3 ≤ d4

6406531892533. ✖ d1 > d2 > d3 > d4

6406531892534. ✔ d1 ≥ d2 ≥ d3 ≥ d4

Question Number : 280 Question Id : 640653566268 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

Suppose that we have 10000 samples in a dataset. Suppose further we use the K − means algorithm to find clusters in the dataset. Then, the statement that K-Means algorithm always converges with zero inertia (or zero Sums Square Error) for some value of K is

Options :

6406531892539. ✔ Always True

6406531892540. ✖ Always False

6406531892541. ✖ True, sometimes

Question Number : 281 Question Id : 640653566269 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

Suppose that we use k-means clustering for a dataset having 100 samples. The initial centroids for *k* clusters can be initialized in multiple ways. One such as way is shown below

```
km = KMeans(n_clusters=20,init='random', n_init=10,random_state=42)
  km.fit(X)
```

Choose the correct statements

Options :

6406531892542. ✔ 20 centroids are randomly initialized 10 times

6406531892543. ✖ 10 centroids are randomly initialized 20 times

6406531892544. ✖ 20 samples in the dataset are selected as initialization point such that they are at least 10 units away from each other

6406531892545. ✖ 10 samples in the dataset are selected as initialization point such that they are at least 20 units away from each other

Sub-Section Number :	5
Sub-Section Id :	64065380975
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 282 Question Id : 640653566237 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

We wish to load the wine dataset from sklearn. Which of the following will throw an error?

Options :

6406531892412. ✔
```
from sklearn.datasets import load_wine
X, y = load_wine(load_X_y = True)
```

6406531892413. ✔
```
from sklearn.datasets import load_wine
data = load_wine(load_X_y = True)
```

6406531892414. ✖
```
from sklearn.datasets import load_wine
data = load_wine(return_X_y = True)
```

6406531892415. ✖
```
from sklearn.datasets import load_wine
X, y = load_wine(return_X_y = True)
```

Question Number : 283 Question Id : 640653566248 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 correct?

Options :

6406531892456. ✖ SGDClassifier(loss="percept") is stochastic version of a perceptron model

6406531892457. ✔ SGDClassifier(loss="log_loss") is stochastic version of a logistic classifier model

6406531892458. ✖ SGDClassifier(loss="log_loss") is stochastic version of a SVM model

6406531892459. ✔ SGDClassifier(loss="hinge") is stochastic version of a SVM model

Question Number : 284 Question Id : 640653566251 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 algorithms may get impacted by feature scaling?

Options :

6406531892468. ✔ LinearRegression

6406531892469. ✖ DecisionTree

6406531892470. ✔ SVM

6406531892471. ✖ BinomialNaiveBayes

Question Number : 285 Question Id : 640653566270 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 :

6406531892546. ✖ SGDClassifier()

6406531892547. ✔ MLPClassifier()

6406531892548. ✖ NNClassifier()

6406531892549. ✔ MLPRegressor()

Sub-Section Number : 6

Sub-Section Id : 64065380976

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 286 Question Id : 640653566252 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

Following information about `X_train` is given:

- Shape of `X_train` is (100,6)
- 4 continuous features, 2 categorical features
- One categorical feature contains 3 categories/unique values
- Second categorical feature contains 4 categories/unique values

```
from sklearn.preprocessing import OneHotEncoder
Ohe  = OneHotEncoder()
Encoded_X_train =  ohe.fit_transform(X_train)
Encoded_X_train.shape
```

Which of the following is(are) correct option(s) for above information ?

Options :

6406531892472. ✔ `Encoded_X_train` will have more number of columns than `X_train`

6406531892473. ✖ `Encoded_X_train` will have 11 columns

6406531892474. ✔ `Encoded_X_train` will have more than 11 columns

6406531892475. ✖ Encoded_X_train will have more number of rows than X_train

Question Number : 287 Question Id : 640653566253 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 ways can help in feature selection?

Options :

6406531892476. ✔ Drop a Feature with many missing values

6406531892477. ✖ Drop a feature containing data with high standard deviation

6406531892478. ✔ Use SelectKBest or SelectKPercentile methods

6406531892479. ✖ Drop a feature which has high correlation with target variable

Sub-Section Number :	7
Sub-Section Id :	64065380977
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 288 Question Id : 640653566263 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 NOT be done at node N?

Options :

6406531892518. ✖ 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.

6406531892519. ✔ 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.

6406531892520. ✔ 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.

6406531892521. ✔ 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 : 289 Question Id : 640653566267 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 code given below attempts to find the clusters in the dataset, *X* using the K-means algorithm.

```
X,y = make_blobs(n_samples=7,n_features=2,centers=2,random_state=42)
km = KMeans(n_clusters=8,init='random',n_init=1,random_state=42)
km.fit(X)
```

Select the true statements about the code upon execution. Assume necessary imports.
Note: There are no typos in the code, the argument names passed to the function are all correct

Options :

6406531892535. ✖ The code attempts to find 2 clusters in the given dataset

6406531892536. ✔ The dataset *X* can be visualized in the Euclidean space

6406531892537. ✔ The code raises an error upon execution

6406531892538. ✖ The code gets executed without an error upon execution

Question Number : 290 Question Id : 640653566271 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 creates a neural network (assume necessary imports)

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

Select the correct statements from the following list of statements

Options :

6406531892550. ✖ The neural network contains 3 hidden layers with 5 neurons in each hidden layer

6406531892551. ✖ The neural network contains 5 hidden layers with 3 neurons in each hidden layer

6406531892552. ✔ The neural network contains 2 hidden layers with 5 neurons in the second hidden layer

6406531892553. ✔ The neural network contains 2 hidden layers with 3 neurons in the first hidden layer

6406531892554. ✖ None of the given options are correct

Sub-Section Number :	8
Sub-Section Id :	64065380978
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 291 Question Id : 640653566245 Question Type : SA Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3

Question Label : Short Answer Question

Consider the following code block:

```
from sklearn.model_selection import GridSearchCV
from sklearn.tree import DecisionTreeClassifier
from sklearn.datasets import make_classification

X, y = make_classification(n_samples = 100, n_features = 3,
n_informative = 2, n_redundant = 1)

param_grid = [{'max_depth': [2, 3, 4, 5, 6], 'min_samples_split': [2, 3, 4, 5, 6]},
              {'min_samples_leaf': [2, 3, 4, 5, 6]},
              {'min_impurity_decrease': [0.2, 0.3, 0.4, 0.5, 0.6],
              'ccp_alpha': [0.1, 0.2, 0.3, 0.4, 0.5, 0.6]}]

gscv = GridSearchCV(DecisionTreeClassifier(), param_grid, cv = 3)
gscv.fit(X, y)

print(gscv.best_params_)
```

How many parameter combinations will be tried by GridSearchCV?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

60

PDSA

Section Id :	64065338408
Section Number :	11
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	29
Number of Questions to be attempted :	29
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 :	64065380979
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Number : 292 Question Id : 640653566272 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)"