6406532240080. ✖ K^{-1}

6406532240081. ✖ X^{-1}

Question Number : 180 Question Id : 640653668623 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

Suppose that the 95% of the variance of the data is captured using $d - 1$ eigenvectors. This implies that

Options :

6406532240082. ✖ The features are highly correlated (or dependent)

6406532240083. ✔ The features have low correlation (or independent)

MLP

Section Id :	64065344906
Section Number :	12
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	23
Number of Questions to be attempted :	23
Section Marks :	50
Display Number Panel :	Yes
Section Negative Marks :	0

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 :	64065395196
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Number : 181 Question Id : 640653668624 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 :

6406532240084. ✔ YES

6406532240085. ✖ NO

Sub-Section Number :	2
Sub-Section Id :	64065395197
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 182 Question Id : 640653668625 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

You have a dataset of student grades in a Pandas DataFrame called `grades_df`.
The columns are: `StudentName`, `Subject`, and `Score`. Consider the following code:

```
average_scores = grades_df.groupby('StudentName')['Score'].mean()
```

After executing the above code, what will `average_scores` contain?

Options :

6406532240086. ✖ The highest score for each student.

6406532240087. ✖ A list of subjects sorted by their average scores.

6406532240088. ✔ The average score of each student across all subjects.

6406532240089. ✖ A DataFrame with the scores of all students for each subject.

Question Number : 183 Question Id : 640653668626 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

You are working on a machine learning project and have received a dataset containing numeric and categorical features. The dataset has some missing values and potential outliers. Given the following data cleaning steps:

1. Use One-Hot Encoding for categorical variables.
2. Impute missing values with feature's mean for numeric features.
3. Remove duplicates.
4. Standardize numeric features using Z-score normalization.
5. Identify and handle outliers using the IQR method.

Which of the following represents the MOST appropriate sequence for preparing the data for a machine learning model?

Options :

6406532240090. ✖ $1 \rightarrow 4 \rightarrow 2 \rightarrow 3 \rightarrow 5$

6406532240091. ✖ $3 \rightarrow 2 \rightarrow 1 \rightarrow 4 \rightarrow 5$

6406532240092. ✖ $2 \rightarrow 3 \rightarrow 1 \rightarrow 5 \rightarrow 4$

6406532240093. ✔ $3 \rightarrow 5 \rightarrow 2 \rightarrow 1 \rightarrow 4$

Question Number : 184 Question Id : 640653668629 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

You're working with a dataset that consists of training data ('train_data') and test data ('test_data'). The dataset contains both numerical and categorical features. You decide to employ a combination of 'StandardScaler' (for numerical columns) and 'OneHotEncoder' (for categorical columns) from 'scikit-learn' using the 'ColumnTransformer' utility. Which of the following actions is MOST likely to introduce data leakage or potential modeling issues?

Options :

6406532240099. ✖ You utilize 'fit_transform' on 'train_data' and then 'transform' on 'test_data' using the 'ColumnTransformer'.

6406532240100. ✖ After observing a new category in the test data that was not present in the training data, you set the 'handle_unknown' parameter to 'ignore' in 'OneHotEncoder'.

6406532240101. ✔ You first apply 'fit' on the 'test_data' and then 'transform' on 'train_data' using the 'ColumnTransformer'.

6406532240102. ✖ Before using 'ColumnTransformer', you independently apply 'fit_transform' to 'train_data' for both 'StandardScaler' and 'OneHotEncoder'.

Question Number : 185 Question Id : 640653668630 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

You are working on a machine learning project that aims to predict housing prices based on various features of the houses. As the first step, you decide to perform exploratory data analysis and visualize the data to understand its structure and relationships. Which of the following visualization techniques or principles is LEAST likely to provide meaningful insights for this kind of regression problem?

Options :

6406532240103. ✖ Plotting a heatmap of the correlation matrix to understand the linear relationship between the numeric features.

6406532240104. ✖ Using a scatter plot to visualize the relationship between the square footage of a house and its price.

6406532240105. ✔ Visualizing the distribution of housing prices using a pie chart.

6406532240106. ✖ Creating box plots for housing prices, grouped by the number of bedrooms, to detect outliers and understand the distribution across different categories.

Question Number : 186 Question Id : 640653668633 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 potentially incorrect steps were taken in the following code snippet?

```
from sklearn.datasets import load_diabetes
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler
from sklearn.linear_model import LinearRegression

X,y = load_diabetes(return_X_y=True)
X_train,X_test, y_train,y_test= train_test_split(X,y,test_size=0.2, random_state=42)

mms = MinMaxScaler()

X_train = mms.fit_transform(X_train)
X_test = mms.fit_transform(X_test)

lr = LinearRegression()
lr.fit(X_train,y_train)

print("linearRegression R2 Score :", lr.score(X_test,y_test))
```

```
linearRegression R2 Score : 0.413160707535728
```

Options :

6406532240109. ✖ Important parameter in MinMaxScaler was missing while transforming the data.

6406532240110. ✖ train_test_split shouldn't be done while setting random_state parameter.

6406532240111. ✔ X_test was transformed incorrectly.

6406532240112. ✖ All the steps are correct

Question Number : 187 Question Id : 640653668634 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

Imagine you've loaded a dataset with 1000 samples into a Pandas DataFrame, and each sample has 30 features. Unfortunately, some samples have missing values for a few features, and you want to remove samples with more than 3 null values present. Please select the method to

accomplish this task.?

Options :

6406532240113. ✖ drop(how= 27)

6406532240114. ✖ drop(columns=['all'])

6406532240115. ✔ dropna(thresh = 27)

6406532240116. ✖ dropna(how='any')

6406532240117. ✖ dropna(thresh=3)

Question Number : 188 Question Id : 640653668635 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

Choose the options with respect to the given statements:

Statement1 : To apply various sklearn methods from in series on a column we should use Pipeline.

Statement2 : To apply various sklearn methods on various columns in parallel we should use ColumnTransformer.

Options :

6406532240118. ✖ Statement 1 False, Statement 2 False

6406532240119. ✖ Statement 1 True, Statement 2 False

6406532240120. ✖ Statement 1 False, Statement 2 True

6406532240121. ✔ Both statements are True

Question Number : 189 Question Id : 640653668641 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:

```
import numpy as np
from sklearn.model_selection import ShuffleSplit

X = np.array([[24, 13],[19, 18],
              [25, 18],[27, 23],
              [11, 25],[22, 12],
              [27, 16],[17, 25]])
y = np.array([-1, 1, -1, 1, 1, -1, 1, -1])

ss = ShuffleSplit(n_splits=3, test_size=.25, random_state= 42)

for train_index, test_index in ss.split(X):
  print(train_index,test_index)
```

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

Options :

6406532240140. ✖

```
[[24 13] [17 25] [25 18] [11 25] [27 23] [27 16]]
[[24 13] [11 25] [22 12] [25 18] [19 18] [27 16]]
[[27 23] [19 18] [11 25] [22 12] [25 18] [17 25]]
```

6406532240141. ✔

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

6406532240142. ✖

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

6406532240143. ✖ Error in the code block

Question Number : 190 Question Id : 640653668643 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 likely to be the correct output of the code given below?

```
from sklearn import linear_model
clf = linear_model.Ridge(alpha=0.01)
X= [[1,0], [2, 1], [3, 2]]
y= [1, 2, 3]
clf.fit(X, y)
linear_model.Ridge(alpha=0.01,max_iter=1000, tol=0.0001,fit_intercept=True)
clf.score(X,y)
```

Options :

6406532240145. ✖ 5

6406532240146. ✖ 99

6406532240147. ✔ 0.999

6406532240148. ✖ No evaluation metrics is mentioned, hence it will produce error.

Question Number : 191 Question Id : 640653668644 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

You are working on optimizing a machine learning model for predicting the energy efficiency of buildings. To capture potential non-linear relationships between features like floor area, wall area, and roof area, you decide to introduce polynomial features. However, considering the risk of multicollinearity due to the introduction of these polynomial features, you also want to ensure the data is appropriately scaled. You construct the following pipeline:

```
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import PolynomialFeatures, MinMaxScaler

pipeline = Pipeline([
    ('poly', PolynomialFeatures(degree=2, interaction_only=True)),
    ('scaler', MinMaxScaler())
])
```

Given this setup, which of the following statements accurately describes the operation of this pipeline on the training data?

Options :

6406532240149. ✖ The pipeline will generate polynomial features (including squared terms) and then scale these features to a range between 0 and 1.

6406532240150. ✖ The transformed data will consist of the original features, their squares, and interaction terms, all scaled between 0 and 1.

6406532240151. ✖ The pipeline scales the original features between 0 and 1, then subsequently generates polynomial combinations including both square terms and interaction terms.

6406532240152. ✔ Only interaction terms between features are generated by the pipeline, which are then scaled between 0 and 1, without including the squared terms of individual features.

Question Number : 192 Question Id : 640653668646 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

You're developing a regression model for predicting house prices based on various attributes of a house. Given that some features might be redundant or irrelevant, you consider Lasso regression to help with feature selection. To determine the most appropriate regularization strength α, you decide to use `LassoCV` from `scikit-learn`. Here's a part of your implemented code:

```
from sklearn.datasets import make_regression
from sklearn.linear_model import LassoCV

X, y = make_regression(n_samples=400, n_features=25, noise=1.0,
↪  random_state=7)
lasso = LassoCV(cv=10)
lasso.fit(X, y)
```

Given the nature of Lasso regression and the purpose of the code, which potential benefit are you hoping to achieve?

Options :

6406532240157. ✔ Optimize the model's complexity by automatically determining the best α through cross-validation.

6406532240158. ✖ Reduce overfitting by incorporating 10-fold cross-validation during model

selection.

6406532240159. ✖ Make predictions using an ensemble of 10 different Lasso models trained on different subsets of the data.

6406532240160. ✖ Maximize the number of features retained in the model, ensuring a complex model representation.

Sub-Section Number : 3

Sub-Section Id : 64065395198

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 193 Question Id : 640653668647 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

You are working on a regression problem and decide to use the `SGDRegressor` from `scikit-learn`. You set up two different regressors with distinct parameter values and train them on the same dataset:

```
from sklearn.linear_model import SGDRegressor

# First SGDRegressor
sgd1 = SGDRegressor(max_iter=1000, tol= None, penalty='none')
sgd1.fit(X_train, y_train)

# Second SGDRegressor
sgd2 = SGDRegressor(max_iter=5, tol=None, penalty='none')
sgd2.fit(X_train, y_train)
```

Given the configurations above, which `SGDRegressor` is more likely to underfit the training data?

Options :

6406532240161. ✖ sgd1

6406532240162. ✔ sgd2

Sub-Section Number : 4

Sub-Section Id : 64065395199

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 194 Question Id : 640653668636 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Max. Selectable Options : 0

Question Label : Multiple Select Question

Which columns may not be included in the selected data within the code below?

```
from sklearn.feature_selection import VarianceThreshold

data =[[  95,  0.332,   112,    1,    0.56  ],
       [ 146,  0.332,   177,    1,    9.2   ],
       [ -96,  0.332,  -139,    1,   -0.82  ],
       [ 116,  0.332,   117,    1,    4.8   ],
       [ -87,  0.332,   -63,    1,   -1.1   ],
       [   5,  0.332,   139,    1,    1.40  ],
       [-142,  0.332,  -214,    1,   -1.31  ],
       [ 148,  0.332,     6,    1,   -8.6   ],
       [ 162,  0.332,    34,    1,   -6.5   ],
       [ -65,  0.332,  -120,    1,   -8.3   ],
       [ 197,  0.332,    44,    1,   -0.76  ]]

vf = VarianceThreshold(threshold=0)

selected_data = vf.fit_transform(data)
selected_data
```

Options :

6406532240122. ✖ Column indexed at 0

6406532240123. ✔ Column indexed at 1

6406532240124. ✖ Column indexed at 2

6406532240125. ✔ Column indexed at 3

6406532240126. ✖ Column indexed at 4

6406532240127. ✖ No columns

Question Number : 195 Question Id : 640653668640 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Max. Selectable Options : 0

Question Label : Multiple Select Question

Which of the following code blocks will correctly take the learning rate as 'optimal' ?

Options :

6406532240136. ✔
```
from sklearn.linear_model import SGDRegressor
linear_regressor = SGDRegressor(learning_rate='optimal', eta0=1e-3)
```

6406532240137. ✖
```
from sklearn.linear_model import SGDRegressor
linear_regressor = (SGDRegressor_learning_rate='adaptive',eta0=1e-2)
```

6406532240138. ✖
```
from sklearn.model_selection import SGDRegressor
SGD_regressor = LinearRegressor(learning_rate='optimal',eta0=1e-2)
```

6406532240139. ✖ None of these

Sub-Section Number :	5
Sub-Section Id :	64065395200
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 196 Question Id : 640653668627 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3 Max. Selectable Options : 0

Question Label : Multiple Select Question

While performing exploratory data analysis (EDA) on a dataset, you come across some columns with a high percentage of missing values. Along with that, a few categorical columns have a large number of unique categories. Which of the following actions would typically be a recommended initial approach during EDA? (Choose multiple correct options.)

Options :

6406532240094. ✔ Visualizing the data distribution of columns to understand their characteristics.

6406532240095. ✖ Using dimensionality reduction techniques, like PCA, to handle columns with many unique categories.

6406532240096. ✔ Visualizing the distribution of missing values across the dataset to ascertain any patterns or systematic missingness.

6406532240097. ✔ Removing columns that have more than 90% missing values without any context.

Question Number : 197 Question Id : 640653668645 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3 Max. Selectable Options : 0

Question Label : Multiple Select Question

Given the following code snippet involving GridSearchCV for hyperparameter tuning of a LinearRegression model:

```
from sklearn.datasets import make_regression
from sklearn.linear_model import SGDRegressor
from sklearn.model_selection import GridSearchCV

X, y = make_regression(n_samples=200,
                       n_features=15,
                       noise=0.5,
                       random_state=24)

params = {'penalty': ['l1', 'l2'], 'max_iter': [500, 1000]}
reg = GridSearchCV(estimator= SGDRegressor(),
                   param_grid= params,
                   scoring= 'neg_mean_squared_error',
                   refit= True)
reg.fit(X, y)
```

Select all statements that are TRUE given this code snippet:

Options :

6406532240153. ✖ The best model chosen by GridSearchCV will be the one that minimizes the negative mean squared error.

6406532240154. ✔ With the given param_grid, the GridSearchCV will evaluate the SGDRegressor model using a total of 4 combinations of hyperparameters to identify the best set for the regression task.

6406532240155. ✔ The optimal model will be refit on the entire data after determining the best hyperparameters using cross-validation.

6406532240156. ✖ The optimal model will be refit on only a subset of the data after determining the best hyperparameters using cross-validation.

Sub-Section Number :	6
Sub-Section Id :	64065395201
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 198 Question Id : 640653668628 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

Given the following code snippet, how many **unique values** will be present in the column Z of the resulting DataFrame df?

```
import pandas as pd

# Creating a DataFrame
data = { 'X': ['apple', 'orange', 'apple', 'banana', 'banana', 'orange'],
         'Y': [1, 2, 3, 3, 2, 2]}
df = pd.DataFrame(data)

df['Z'] = df['X'] + df['Y'].astype(str)
```

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

5

Question Number : 199 Question Id : 640653668631 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 ?

```
data = [['apple', 120],
        ['cherry', 130],
        ['apple', 122],
        ['apple', 125],
        ['grapes', 70]]

from sklearn.preprocessing import OneHotEncoder
ohe = OneHotEncoder(sparse_output=False)
ohe.fit(data)
print(ohe.transform(data).shape[1])
```

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

8

Question Number : 200 Question Id : 640653668632 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.preprocessing import MinMaxScaler, StandardScaler

data = [[0, 5],
        [8, 3],
        [3, 4],
        [7, 2],
        [7, 9]]

scaler = StandardScaler()
scaler.fit(data)
print(scaler.var_[0])
```

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

9.2

Question Number : 201 Question Id : 640653668648 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

You're using `GridSearchCV` to optimize a `Ridge` regression model from `scikit-learn`. Consider the following hyperparameter grid:

```
from sklearn.linear_model import Ridge
from sklearn.model_selection import GridSearchCV

param_grid = {
    'alpha': [0.001, 0.01, 0.1, 1, 10, 100],
    'fit_intercept': [True, False],
    'solver': ['auto', 'lsqr', 'sag']
}

grid_search = GridSearchCV(Ridge(), param_grid, cv=5)
```

How many combinations will `GridSearchCV` evaluate?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

36

Sub-Section Number :	7
Sub-Section Id :	64065395202
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 202 Question Id : 640653668642 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

For LinearRegression with equation $Y = W_0X_0 + W_1X_1 + W_2X2 + \epsilon$ and given that $W_2 = \frac{5}{7} * W_1$ and $\epsilon = 0$. What will be the value of the W_1 for the below code? (Write 3 digits after the decimal)

Where X_1 and X_2 are column1 and column2 respectively and W_1 and W_2 are weights associated to the respected columns while fitting

```
from sklearn.linear_model import LinearRegression
X_train = [[0,0], [2,1.43], [4,2.86], [6,4.29]]
y_train = [0,1,2,3]
reg = LinearRegression(fit_intercept=False) #intercept=0
reg.fit(X_train,y_train)
print(reg.coef_[0])
```

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

0.327 to 0.333

Sub-Section Number :	8
Sub-Section Id :	64065395203
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Id : 640653668637 Question Type : COMPREHENSION Sub Question Shuffling Allowed : No Group Comprehension Questions : No Question Pattern Type : NonMatrix Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Question Numbers : (203 to 204)

Question Label : Comprehension

Go through the code snippet given below and answer the given subquestions.

```
from sklearn.linear_model import SGDRegressor
from sklearn.pipeline import make_pipeline
n_samples, n_features = 18, 4
rng = np.random.RandomState(0)
y = rng.randn(n_samples)
X = rng.randn(n_samples, n_features)
reg = SGDRegressor(max_iter=1000,
                   tol=1e-3,
                   eta0= 0.04,
                   power_t=5,
                   n_iter_no_change=3,
                   validation_fraction=0.3 ,
                   random_state=42)
reg.fit(X, y)
print(reg.coef_)
```

Sub questions

Question Number : 203 Question Id : 640653668638 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 options will be the output of the given code?

Options :

6406532240128. ✔ [-0.02634908 0.01189399 0.0917284 0.08966849]

6406532240129. ✖ array([-0.22622766, -0.00582008, -0.1820344 , 0.03518086, -0.14490955])

6406532240130. ✖ array([-0.22622766, -0.00582008, -0.1820344])

6406532240131. ✖ Given code will return an error because the data set is not given.

Question Number : 204 Question Id : 640653668639 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 could be the possible output of print(reg.score())?

Options :

6406532240132. ✖ -0.528

6406532240133. ✖ 1

6406532240134. ✖ 0.528

6406532240135. ✔ Given code will return an error

Business Analytics

Section Id :	64065344907
Section Number :	13
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	8
Number of Questions to be attempted :	8
Section Marks :	20
Display Number Panel :	Yes
Section Negative Marks :	0
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 :	64065395204
Question Shuffling Allowed :	No
Is Section Default? :	null