NOTE: Enter your answer in two decimal places.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

0.32 to 0.36

MLP

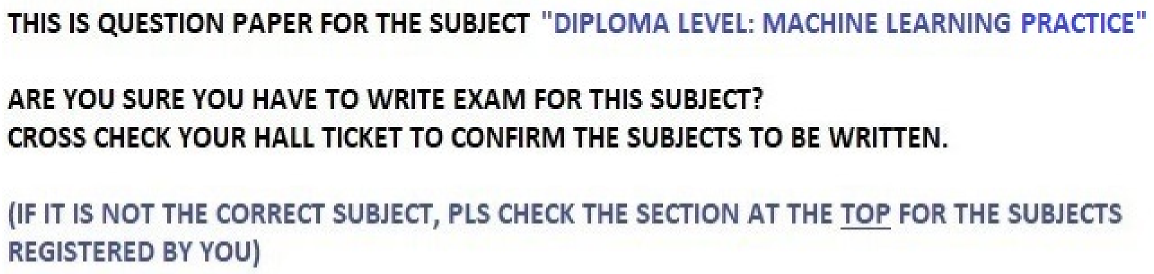
Number of Questions : 21

Section Marks : 50

Question Number : 184 Question Type : MCQ

Correct Marks : 0

Question Label : Multiple Choice Question



THIS IS QUESTION PAPER FOR THE SUBJECT "DIPLOMA LEVEL: MACHINE LEARNING PRACTICE"

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 :

A. ✓ Yes

B. ✖ No

Question Number : 185 Question Type : MCQ

Correct Marks : 1

Question Label : Multiple Choice Question

Which of the following is/are NOT a metric for regression?

Options :

A. ✖ Mean Absolute Error

B. ✖ R^2 score

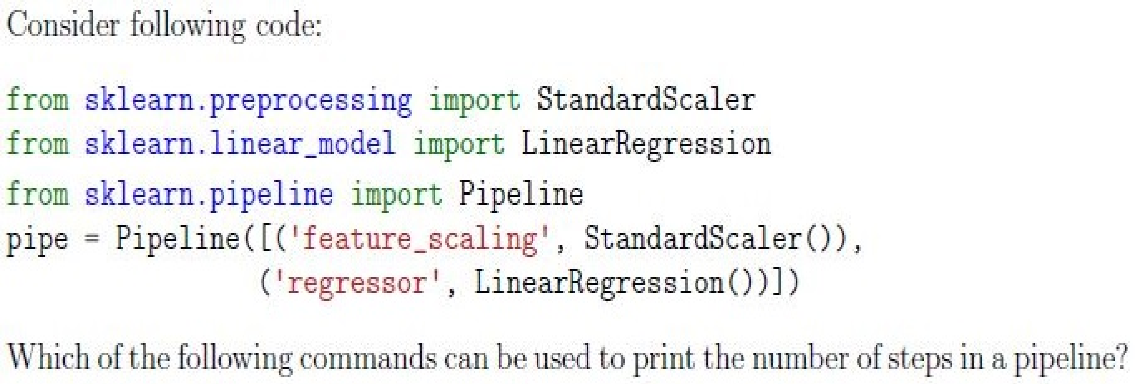
C. ✔ $F1$ score

D. ✖ Mean Squared Error

Question Number : 186 Question Type : MCQ

Correct Marks : 1

Question Label : Multiple Choice Question

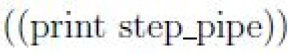


Consider following code:

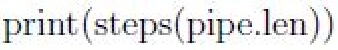
```
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LinearRegression
from sklearn.pipeline import Pipeline
pipe = Pipeline([('feature_scaling', StandardScaler()),
                 ('regressor', LinearRegression())])
```

Which of the following commands can be used to print the number of steps in a pipeline?

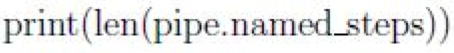
Options :

A. ✖ 

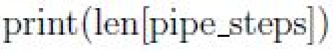
((print step_pipe))

B. ✖ 

print(steps(pipe.len))

C. ✔ 

print(len(pipe.named_steps))

D. ✖ 

print(len[pipe_steps])

Question Number : 187 Question Type : MCQ

Correct Marks : 2

Question Label : Multiple Choice Question

Consider the following code:

```
from sklearn.datasets import load_diabetes
data = load_diabetes()
```

The feature matrix containing the samples will be present in:

Options :

A. ✖ data.feature_matrix

B. ✖ data.features

C. ✖ data.matrix

D. ✓ data.data

Question Number : 188 Question Type : MCQ

Correct Marks : 2

Question Label : Multiple Choice Question

Consider following code snippet:

```
from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split
X, y = load_iris(return_X_y = True)
print(X.shape)
print(y.shape)
train_X, test_X, train_y, test_y = train_test_split(X, y,
                                                    test_size=0.2,
                                                    random_state=42)
```

The output of the above code is the following:

```
(150, 4)
(150,)
```

What will be the output of the following code?

```
print(train_X.shape)
print(test_X.shape)
print(train_y.shape)
print(test_y.shape)
```

Options :

A. ✗ (130, 3) (20, 1) (130, 3) (20, 1)

B. ✓ (120, 4) (30, 4) (120,) (30,)

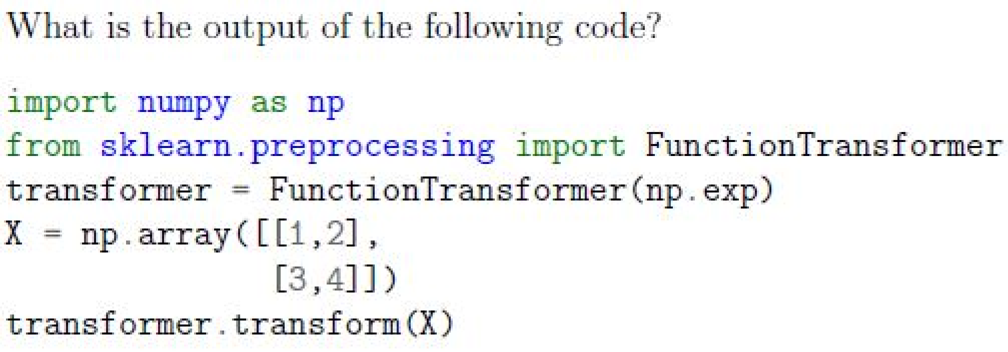
C. ✗ (130, 4) (20, 4) (130,) (20,)

D. ✗ (120, 3) (30, 1) (120, 3) (30, 1)

Question Number : 189 Question Type : MCQ

Correct Marks : 2

Question Label : Multiple Choice Question

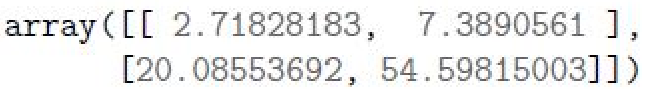


What is the output of the following code?

```
import numpy as np
from sklearn.preprocessing import FunctionTransformer
transformer = FunctionTransformer(np.exp)
X = np.array([[1,2],
              [3,4]])
transformer.transform(X)
```

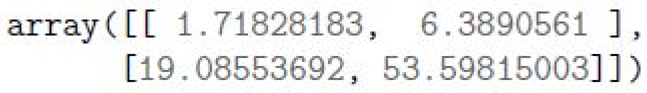
Options :

A. ✓



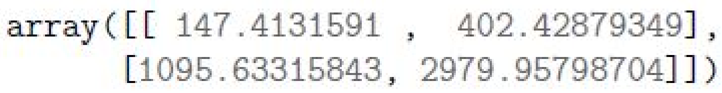
```
array([[ 2.71828183,  7.3890561 ],
       [20.08553692, 54.59815003]])
```

B. ✗



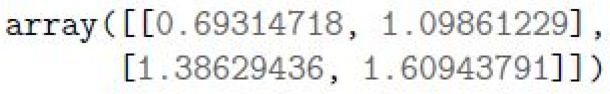
```
array([[ 1.71828183,  6.3890561 ],
       [19.08553692, 53.59815003]])
```

C. ✗



```
array([[ 147.4131591 ,  402.42879349],
       [1095.63315843, 2979.95798704]])
```

D. ✗



```
array([[0.69314718, 1.09861229],
       [1.38629436, 1.60943791]])
```

Question Number : 190 Question Type : MCQ

Correct Marks : 2

Question Label : Multiple Choice Question

Based on the following statements, select the correct option.

I. For a very large dataset, it is better to employ LinearRegressor over SGDRegressor.

II. SGDRegressor solves the normal equation to calculate the value of unknown parameters.

Options :

A. ✖ Statement I is true, Statement II is true, Statement II is a correct explanation for statement I.

B. ✖ Statement I is true, Statement II is true, Statement II is not a correct explanation for statement I.

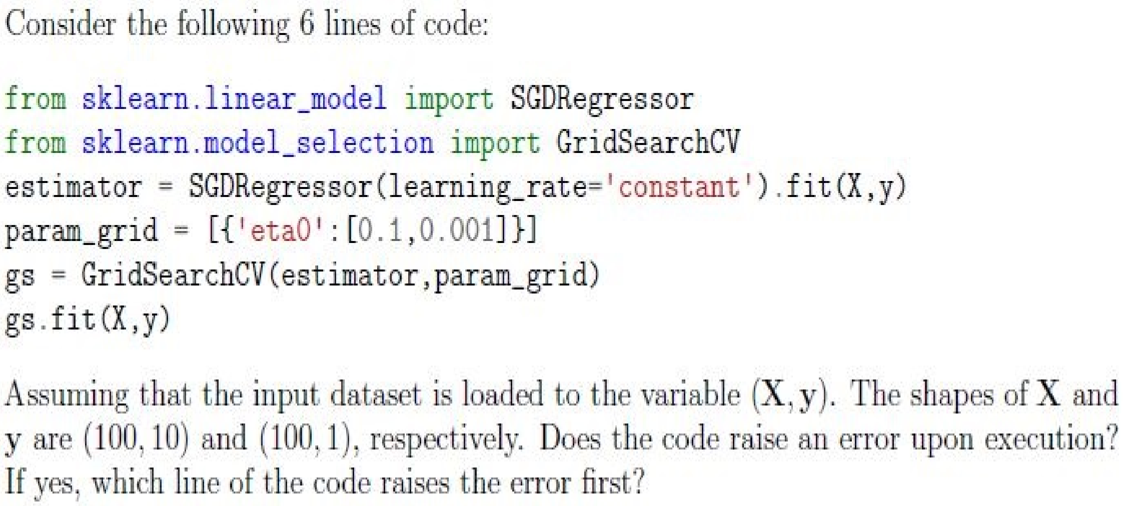
C. ✖ Statement I is true, Statement II is false.

D. ✔ Statement I and II are false.

Question Number : 191 Question Type : MCQ

Correct Marks : 2

Question Label : Multiple Choice Question



Consider the following 6 lines of code:

```
from sklearn.linear_model import SGDRegressor
from sklearn.model_selection import GridSearchCV
estimator = SGDRegressor(learning_rate='constant').fit(X,y)
param_grid = [{'eta0':[0.1,0.001]}]
gs = GridSearchCV(estimator,param_grid)
gs.fit(X,y)
```

Assuming that the input dataset is loaded to the variable $(\mathbf{X}, \mathbf{y})$. The shapes of $\mathbf{X}$ and $\mathbf{y}$ are $(100, 10)$ and $(100, 1)$, respectively. Does the code raise an error upon execution? If yes, which line of the code raises the error first?

Options :

A. ✖ Error in the third line of code

B. ✖ Error in the fourth line of code

C. ✖ Error in the fifth line of code

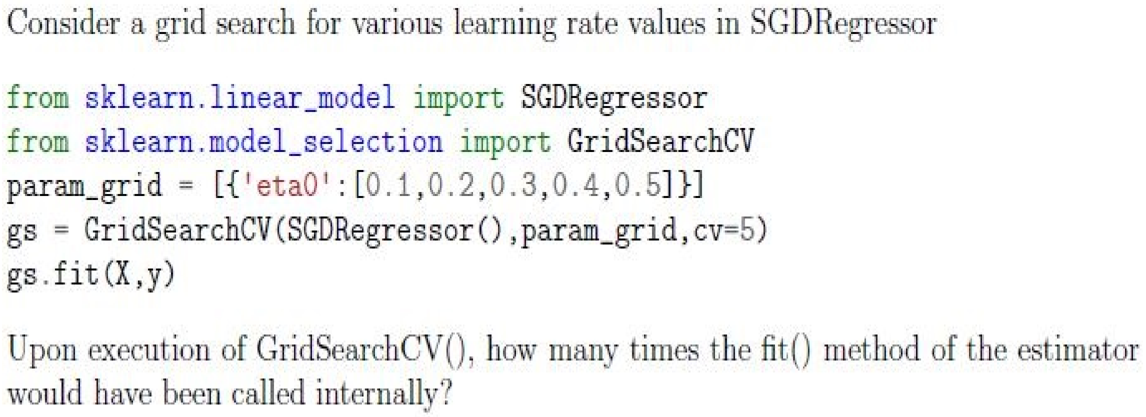
D. ✖ Error in the sixth line of code

E. ✔ It won’t raise any error

Question Number : 192 Question Type : MCQ

Correct Marks : 2

Question Label : Multiple Choice Question



Consider a grid search for various learning rate values in SGDRegressor

```
from sklearn.linear_model import SGDRegressor
from sklearn.model_selection import GridSearchCV
param_grid = [{'eta0':[0.1,0.2,0.3,0.4,0.5]}]
gs = GridSearchCV(SGDRegressor(),param_grid,cv=5)
gs.fit(X,y)
```

Upon execution of GridSearchCV(), how many times the fit() method of the estimator would have been called internally?

Options :

A. ✖ 1

B. ✓ 25

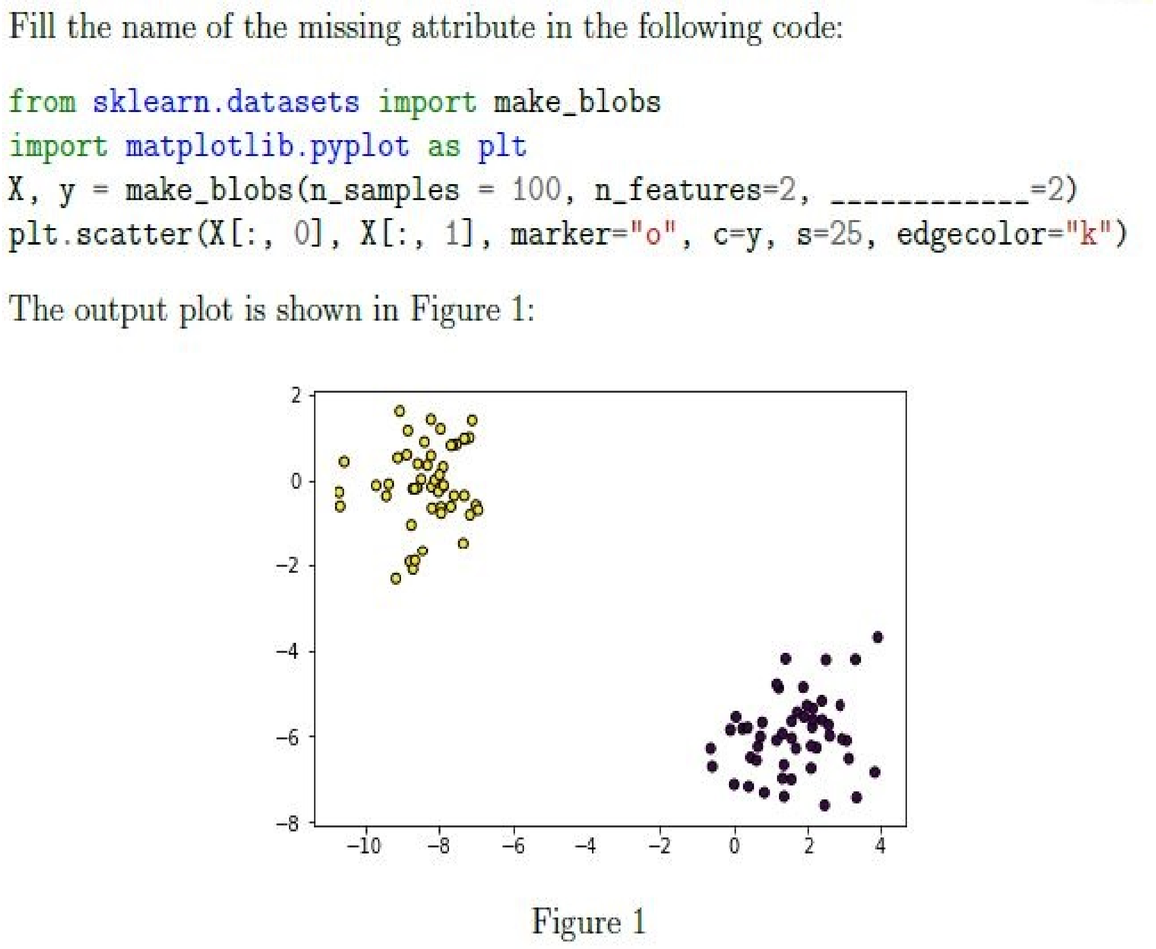
C. ✖ 5

D. ✖ 125

Question Number : 193 Question Type : MCQ

Correct Marks : 3

Question Label : Multiple Choice Question



Fill the name of the missing attribute in the following code:

```
from sklearn.datasets import make_blobs
import matplotlib.pyplot as plt
X, y = make_blobs(n_samples = 100, n_features=2, ____________=2)
plt.scatter(X[:, 0], X[:, 1], marker="o", c=y, s=25, edgecolor="k")
```

The output plot is shown in Figure 1:

Figure 1

Options :

A. ✖ clusters

B. ✖ labels

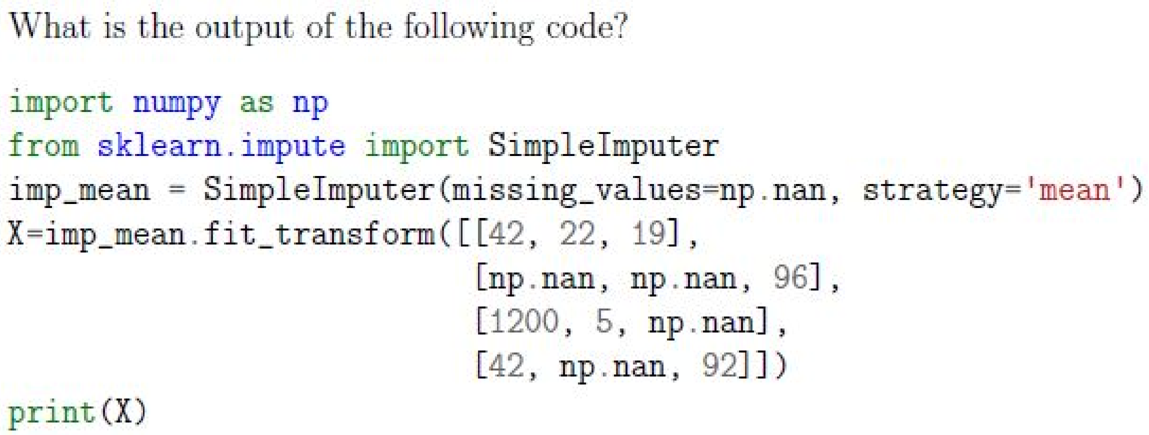
C. ✓ centers

D. ✖ classes

Question Number : 194 Question Type : MCQ

Correct Marks : 3

Question Label : Multiple Choice Question



What is the output of the following code?

```
import numpy as np
from sklearn.impute import SimpleImputer
imp_mean = SimpleImputer(missing_values=np.nan, strategy='mean')
X=imp_mean.fit_transform([[42, 22, 19],
                          [np.nan, np.nan, 96],
                          [1200, 5, np.nan],
                          [42, np.nan, 92]])
print(X)
```

Options :

A. ✓ $\begin{bmatrix} 42. & 22. & 19. \\ 428. & 13.5 & 96. \\ 1200. & 5. & 69. \\ 42. & 13.5 & 92. \end{bmatrix}$

B. ✖ $\begin{bmatrix} 42 & 22 & 19 \\ 2 & 3 & 96 \\ 10 & 100 & 69 \\ 42 & 13.5 & 92 \end{bmatrix}$

C. ✖ $\begin{bmatrix} 42 & 22 & 19 \\ 3 & np.nan & np.nan \\ 10 & 100 & 92 \\ 1200 & 5 & 42 \end{bmatrix}$

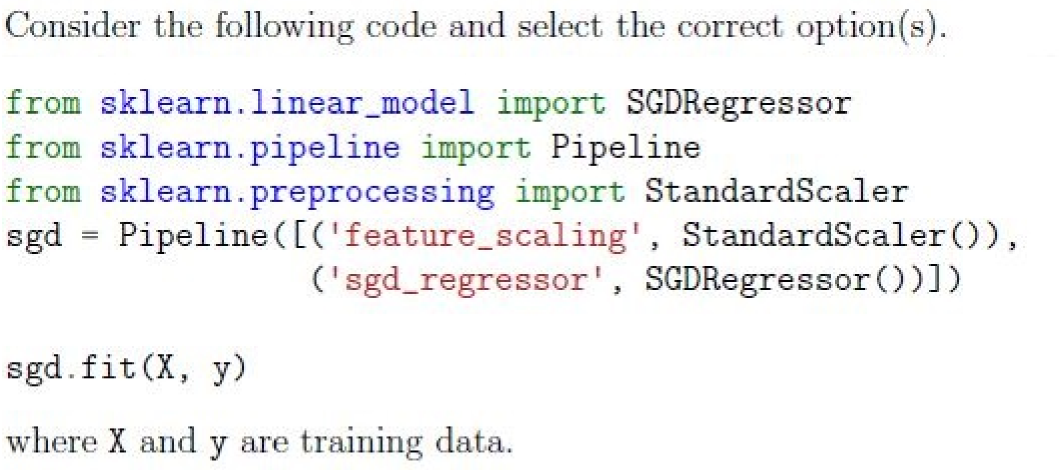
D. ✖

$$\begin{bmatrix} 42. & 13.5 & 91. \\ 428. & 22 & 96. \\ 1200. & 13.5. & 96. \\ 42. & 236 & 92. \end{bmatrix}$$

Question Number : 195 Question Type : MCQ

Correct Marks : 4

Question Label : Multiple Choice Question



Consider the following code and select the correct option(s).

```
from sklearn.linear_model import SGDRegressor
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import StandardScaler
sgd = Pipeline([('feature_scaling', StandardScaler()),
                ('sgd_regressor', SGDRegressor())])

sgd.fit(X, y)
```

where X and y are training data.

Options :

A. ✖ It transforms the original feature vector X into a new feature vector so that all values fall within range [0, 1].

B. ✖ It transforms the original feature vector X into a new feature vector so that all values fall within range [−1, 1].

C. ✖ It will fit the linear regression model on the transformed data using normal equation.

D. ✓ It will fit the linear regression model on the transformed data using iterative optimization.

Question Number : 196 Question Type : MSQ

Correct Marks : 2

Question Label : Multiple Select Question

Which of the following list of functions are used for hyper-parameter tuning in sklearn?

Options :

A. ✓ GridSearchCV()

B. ✓ RandomizedSearchCV()

C. ✖ MinMaxScaler()

D. ✖ StandardScaler()

Question Number : 197 Question Type : MSQ

Correct Marks : 2

Question Label : Multiple Select Question

Which of the following statements are true about hyper-parameter tuning?

Options :

A. ✖ It is only applicable for applications that require gradient calculations.

B. ✖ The tuning is done by using a portion of the test data set.

C. ✓ It is used to select the best model’s various settings of hyper-parameters.

D. ✓ More than one hyper-parameter of a model can be tuned simultaneously in sklearn.

Question Number : 198 Question Type : MSQ

Correct Marks : 3

Question Label : Multiple Select Question

Which of the following statements is/are correct?

Options :

A. ✓ Oversampling refers to replicating some points from the minority class in order to increase the cardinality of the minority class.

B. ✓ Oversampling may consist of either replicating or generating synthetic data for the minority class.

C. ✖ SMOTE is a popular technique for undersampling.

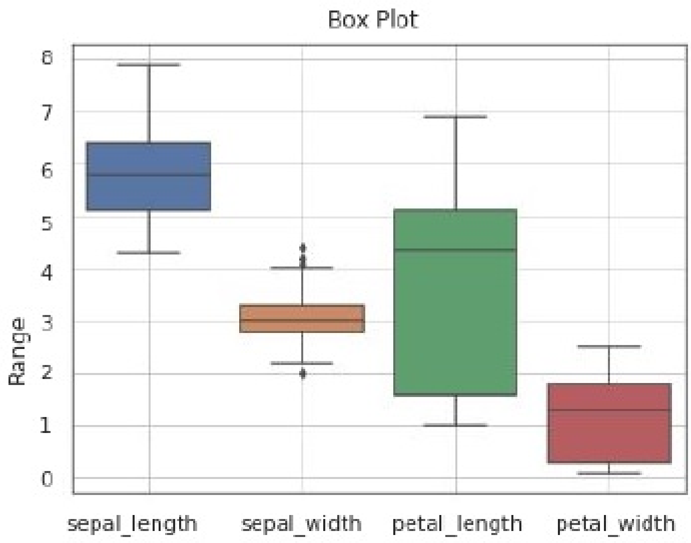
D. ✓ FeatureUnion and Pipeline can be used to create complex transformers.

Question Number : 199 Question Type : MSQ

Correct Marks : 3

Question Label : Multiple Select Question

Consider the following box plots of features of iris data and select the correct options.



Options :

A. ✔ Petal length of data is left-skewed

B. ✖ Petal length of data is right-skewed

C. ✔ Median of petal width lies between 1 and 2

D. ✖ Median of petal length lies between 1 and 2

Question Number : 200 Question Type : MSQ

Correct Marks : 3

Question Label : Multiple Select Question

Select the correct options regarding SGDRegressor:

Options :

A. ✖ With every SGD iteration, the loss steadily decreases.

B. ✔ More iterations require more computational time.

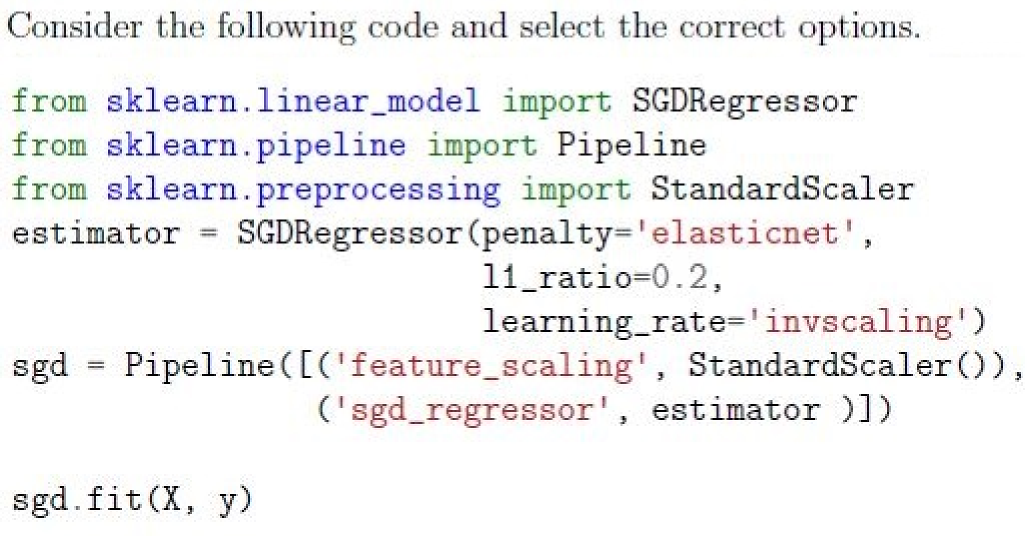
C. ✔ The tol(error tolerance) parameter restricts the number of iterations performed.

D. ✔ Training error might actually increase while performing SGD iterations.

Question Number : 201 Question Type : MSQ

Correct Marks : 3

Question Label : Multiple Select Question



Consider the following code and select the correct options.

```
from sklearn.linear_model import SGDRegressor
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import StandardScaler
estimator = SGDRegressor(penalty='elasticnet',
                         l1_ratio=0.2,
                         learning_rate='invscaling')
sgd = Pipeline([('feature_scaling', StandardScaler()),
                ('sgd_regressor', estimator )])

sgd.fit(X, y)
```

Options :

A. ✓ Regularization rate for L2 penalty will be 0.8

B. ✖ Regularization rate for L2 penalty will be 0.2

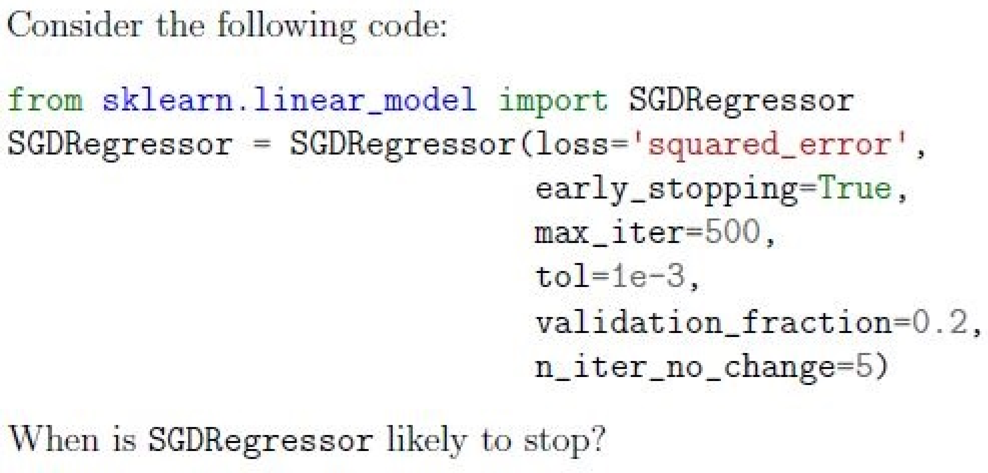
C. ✓ Learning rate decreases with every iteration

D. ✖ Learning rate increases with every iteration

Question Number : 202 Question Type : MSQ

Correct Marks : 3

Question Label : Multiple Select Question



Consider the following code:

```
from sklearn.linear_model import SGDRegressor
SGDRegressor = SGDRegressor(loss='squared_error',
                            early_stopping=True,
                            max_iter=500,
                            tol=1e-3,
                            validation_fraction=0.2,
                            n_iter_no_change=5)
```

When is SGDRegressor likely to stop?

Options :

A. ✓ validation score does not improve by at least 10^{-3} for 5 consecutive epochs.

B. ✖ validation score does not improve by at least 10^{-3} for the first epoch.

C. ✓ after at most 500 iterations.

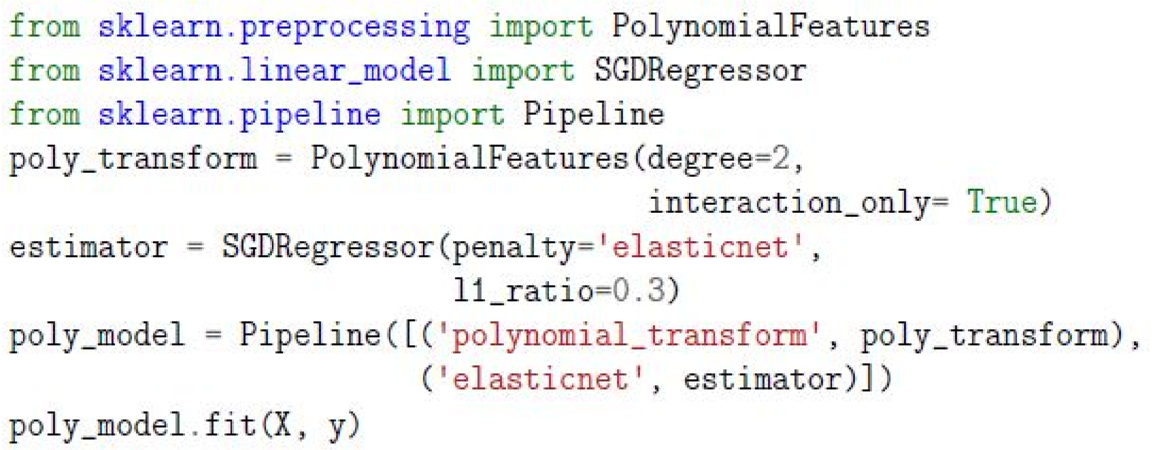
D. ✖ after exactly 5 iterations.

Question Number : 203 Question Type : MSQ

Correct Marks : 3

Question Label : Multiple Select Question

Consider the following code and select the correct options. Assume that all the necessary libraries are imported.



```
from sklearn.preprocessing import PolynomialFeatures
from sklearn.linear_model import SGDRegressor
from sklearn.pipeline import Pipeline
poly_transform = PolynomialFeatures(degree=2,
                                    interaction_only= True)
estimator = SGDRegressor(penalty='elasticnet',
                         l1_ratio=0.3)
poly_model = Pipeline([('polynomial_transform', poly_transform),
                       ('elasticnet', estimator)])
poly_model.fit(X, y)
```

Options :

A. ✓ It will first transform the features into degree 2 such that power of 2 and higher of the same input feature are excluded.

B. ✗ It will first transform the features into degree 2 such that only power of 2 the same input feature are included.

C. ✗ It will not perform any regularization.

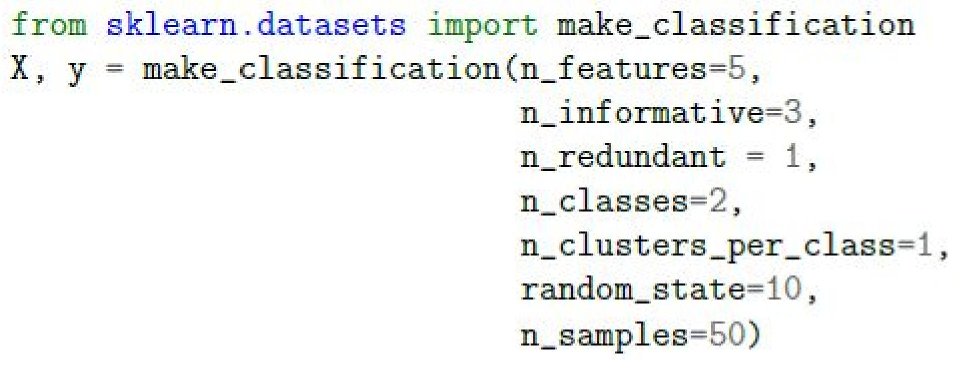
D. ✓ It will apply both lasso and ridge penalties to the model.

Question Number : 204 Question Type : MSQ

Correct Marks : 4

Question Label : Multiple Select Question

Consider the following code and select the correct options.



```
from sklearn.datasets import make_classification
X, y = make_classification(n_features=5,
                           n_informative=3,
                           n_redundant = 1,
                           n_classes=2,
                           n_clusters_per_class=1,
                           random_state=10,
                           n_samples=50)
```

Options :

A. ✗ The shapes of X and y are $(5, 2)$ and $(5,)$, respectively.

B. ✓ The shapes of X and y are $(50, 5)$ and $(50,)$, respectively.

C. ✓ X has one feature such that it can be written as linear combination of other features.

D. ✗ X has three features such that they can be written as linear combination of other features.