✔ $h(x_i) = sign(w^T x_i)$

Sub-Section Number : 14

Sub-Section Id : 640653103299

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 174 Question Id : 640653697754 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 6 Max. Selectable Options : 0

Question Label : Multiple Select Question

Choose all the correct statements about neural networks

Options :

6406532330358. ✖ It can not be used for both regression and classification problems

6406532330359. ✔ It can have more than two hidden layers

6406532330360. ✔ The activation functions have to be non-linear to separate not linearly separable data points

6406532330361. ✔ Each neuron in the neural network may or may not have bias associated with it

MLP

Section Id : 64065349267

Section Number : 8

Section type : Online

Mandatory or Optional : Mandatory

Number of Questions : 36

Number of Questions to be attempted : 36

Section Marks :	100
Display Number Panel :	Yes
Section Negative Marks :	0
Group All Questions :	No
Enable Mark as Answered Mark for Review and Clear Response :	Yes
Maximum Instruction Time :	0
Sub-Section Number :	1
Sub-Section Id :	640653103300
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Number : 175 Question Id : 640653697755 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 :

6406532330362. ✔ YES

6406532330363. ✖ NO

Sub-Section Number :	2
Sub-Section Id :	640653103301
Question Shuffling Allowed :	Yes

Is Section Default? : null

Question Number : 176 Question Id : 640653697756 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 Python code snippet using Pandas:

```
import pandas as pd

# Assume a DataFrame df is already loaded with appropriate data.

grouped_data = df.groupby('Category')['Quantity'].sum()
```

What does the groupby operation in the code achieve?

Options :

6406532330364. ✔ It calculates the total quantity for each category.

6406532330365. ✖ It calculates the average quantity for each category.

6406532330366. ✖ It groups the data by the 'Category' column.

6406532330367. ✖ It filters out rows where the 'Quantity' column is zero.

Question Number : 177 Question Id : 640653697758 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 below code:

```
data = [[-6, 5],
        [-6, 5],
        [ 3, 1],
        [ 3, 1]]

from sklearn.preprocessing import StandardScaler
ss = StandardScaler()
print(ss.fit_transform(data))
```

Which of the following option represents the print output :

Options :

6406532330372. ✖

```
[[0, -1],
 [0, -1],
 [1,  1],
 [1,  1]]
```

6406532330373. ✖

```
[[-0.5, -2],
 [-0.5, -2],
 [  1,   2],
 [  1,   2]]
```

6406532330374. ✔

```
[[-1,  1],
 [-1,  1],
 [ 1, -1],
 [ 1, -1]]
```

6406532330375. ✖

```
[[-1, -1],
 [-1, -1],
 [ 1,  1],
 [ 1,  1]]
```

Question Number : 178 Question Id : 640653697765 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 Ridge regression model in scikit-learn, represented by the Ridge class.

Which of the following statements about Ridge regression is correct?

Options :

6406532330403. ✖ Ridge regression is specifically designed for handling non-linear relationships in the data.

6406532330404. ✔ The regularization term in Ridge regression is added to the sum of squared residuals.

6406532330405. ✖ Increasing the value of the regularization parameter (**alpha**) in Ridge regression tends to overfit the model.

6406532330406. ✖ Ridge regression is equivalent to ElasticNet regression when the regularization parameter (**alpha**) is set to one.

Question Number : 179 Question Id : 640653697775 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 it relevant to add a preprocessing step to scale the data using a `StandardScaler` when working with a `KNeighborsClassifier`?

Options :

6406532330433. ✖ Speeds up the process of finding neighbors on unscaled data.

6406532330434. ✔ k-nearest neighbors relies on computing distances. Normalizing features ensures that each feature contributes approximately equally to the distance computation.

6406532330435. ✖ Scaling the data significantly improves the accuracy of k-nearest neighbor models.

6406532330436. ✖ It doesn't matter. K-nearest neighbors works equally well with or without normalizing the dataset.

Question Number : 180 Question Id : 640653697776 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

When using `CountVectorizer` in scikit-learn, what does the `max_features` parameter control?

Options :

6406532330437. ✖ The upper limit on the number of documents considered during vectorization.

6406532330438. ✔ The maximum number of features (words) to be extracted based on term frequency.

6406532330439. ✖ The maximum number of randomly selected features (words).

6406532330440. ✖ The upper limit on the number of characters allowed in each document.

Question Number : 181 Question Id : 640653697787 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

Rahul is working on an unsupervised machine learning algorithm. He is not able to choose the optimum number of clusters(k) for his model (model is based on K- means algorithm). He tried to plot the elbow chart which is shown below. By looking at this chart, which option would you recommend Rahul as the most suitable value of number of clusters (k).

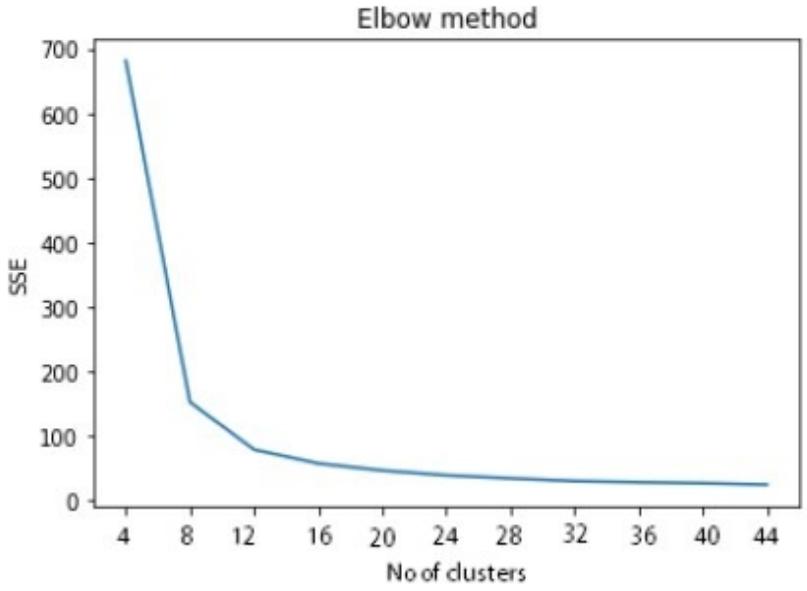

Figure 1: Elbow chart

Options :

6406532330472. ✖ 4

6406532330473. ✖ 8

6406532330474. ✔ 12

6406532330475. ✖ 40

6406532330476. ✖ 44

Question Number : 182 Question Id : 640653697791 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 training a multi-layer perceptron (MLP) classifier on a dataset for a multi-class classification task. The following code snippet demonstrates the process using the 'MLPClassifier' from scikit-learn:

```
from sklearn.neural_network import MLPClassifier
import numpy as np

# Simulated data (features: X, target: y)
X = np.array([[1, 2], [3, 4], [5, 6]])
y = np.array([0, 1, 2])

# Create an MLPClassifier with a specified maximum number of iterations

model = MLPClassifier(max_iter=100, alpha = 0.01, random_state=42)

# Fit the model on the training data
model.fit(X, y)

# Get the number of iterations used in training
iterations_used = model.n_iter_
```

What does the parameter 'alpha' in the MLPClassifier signify?

Options :

6406532330492. ✖ It determines the maximum fraction of neurons to be used to train the model

6406532330493. ✖ It is the learning rate

6406532330494. ✖ It is the L1 regularization rate.

6406532330495. ✔ It is the L2 regularization rate.

6406532330496. ✖ None of these

Sub-Section Number : 3

Sub-Section Id : 640653103302

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 183 Question Id : 640653697757 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

Suppose you have loaded a dataset representing various attributes of red wine into a Pandas DataFrame named `wine_data` using the following code:

```
import pandas as pd

data_url = '''https://archive.ics.uci.edu/ml/machine-learning-databases/
wine-quality/winequality-red.csv'''
wine_data = pd.read_csv(data_url, sep=";")
```

Now, suppose you want to retrieve the value of the `alcohol` attribute for the fifth sample in the dataset (4th by index). Which of the following expressions correctly achieves this? The `alcohol` attribute is at the 3rd index of columns.

Options :

6406532330368. ✔ `wine_data.alcohol[4]`

6406532330369. ✔ `wine_data['alcohol'][4]`

6406532330370. ✖ `wine_data[4][3]`

6406532330371. ✔ `wine_data.iloc[4, 3]`

Question Number : 184 Question Id : 640653697763 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

You're building a machine learning pipeline to preprocess data and train a model on a classification task. You decide to use a pipeline that includes data pre-processing and a support vector machine (SVM) classifier. The following code snippet demonstrates the pipeline creation and usage:

```
from sklearn.pipeline import Pipeline
from sklearn.svm import SVC
from sklearn.preprocessing import StandardScaler
import numpy as np

# Simulated data (features: X, target: y)
X = np.array([[2, 3], [5, 7], [8, 10]])
y = np.array([0, 1, 0])

# Create a pipeline with StandardScaler and SVM classifier
pipe = Pipeline([
    ('scaler', StandardScaler()),
    ('svm', SVC())])

# Fit the pipeline on training data
pipe.fit(X, y)

# Make predictions using the trained pipeline
predictions = pipe.predict(X)
```

Which of the following can be used to get number of support vectors of the classifier?

Options :

6406532330394. ✔ `pipe[1].n_support_`

6406532330395. ✖ `pipe[1]._n_support_vecs`

6406532330396. ✔ `pipe['svm'].n_support_`

6406532330397. ✖ `pipe['svm']._n_support_vecs`

6406532330398. ✖ None of these

Question Number : 185 Question Id : 640653697771 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

You are building a sentiment analysis model using scikit-learn's `SGDClassifier` to classify movie reviews as positive or negative. The dataset is quite large, and you're dealing with a high-dimensional feature space. You want to fine-tune the hyperparameters of the classifier to achieve better convergence and classification performance.

Here's how you're setting up the `SGDClassifier`:

```
from sklearn.linear_model import SGDClassifier
classifier = SGDClassifier(loss='hinge', alpha=0.0001,
                           max_iter=1000, tol=1e-3, power_t=0.5)
```

Which of the following are correct?

Options :

6406532330417. ✖ The classifier is based on perceptron model.

6406532330418. ✔ The classifier is based on logistic regression model.

6406532330419. ✔ The learning rate decays every iteration.

6406532330420. ✖ The learning rate is 0.0001.

6406532330421. ✖ It will run exactly for 1000 iterations, which is independent of the data.

6406532330422. ✖ None of these.

Question Number : 186 Question Id : 640653697778 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

When working with a `KNeighborsClassifier` in scikit-learn, consider the following scenarios. Choose all the correct statements:

Options :

6406532330442. ✖ A higher value of `n_neighbors` tends to overfit the model.

6406532330443. ✔ Reducing the value of `n_neighbors` typically increases the risk of overfitting.

6406532330444. ✔ Opting for a small `n_neighbors` can make the model sensitive to noise in the data.

6406532330445. ✖ Increasing `n_neighbors` always enhances the model's ability to generalize to new, unseen data.

Question Number : 187 Question Id : 640653697780 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

Fill in the missing parameter value in the following estimator that can be used to classify the data

```
from sklearn.svm import SVC
clf = SVC(kernel = ______)
clf.fit(X, y)
```

Options :

6406532330450. ✖ 'lasso'

6406532330451. ✔ 'poly'

6406532330452. ✖ 'scale'

6406532330453. ✔ 'rbf'

Question Number : 188 Question Id : 640653697789 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 are true about multilayer perceptron model in sklearn?

Options :

6406532330482. ✔ It is an iterative algorithm/model.

6406532330483. ✔ It can be used to capture non-linear relationships between features and labels.

6406532330484. ✔ It can be used for regression as well as classification.

6406532330485. ✖ It can be used for clustering.

6406532330486. ✖ None of these

Sub-Section Number :	4
Sub-Section Id :	640653103303
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 189 Question Id : 640653697759 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

consider the below code : keep following symbols in mind:

- >>>: Represents input code
- # : Represents comment in a code
- ... : Represents code continuation
- Without any symbols at the beginning of a line then it is output of just above input line of code.

```
>>> from sklearn.feature_selection import SelectKBest, f_regression
>>> from sklearn.datasets import load_diabetes
>>> X,y = load_diabetes(return_X_y=True,as_frame=True)
>>> print(X.shape)
(442,10)

>>> print(X.columns)
['age', 'sex', 'bmi', 'bp', 's1', 's2', 's3', 's4', 's5', 's6']

>>> skb = SelectKBest(f_regression, k=4)
>>> X_selected = skb.fit_transform(X, y)

>>> print(skb.scores_)
[16.10, 0.81, 230.65, 106.52, 20.71, 13.74,  81.23, 100.06, 207.27, 75.39]

>>> print(skb.pvalues_)
[7.0e-05, 3.6e-01, 3.4e-42, 1.6e-22, 6.9e-06,
 2.3e-04, 6.1e-18, 2.3e-21, 8.8e-39, 7.5e-17]

>>> print(skb.pvalues_.argsort())
[2, 8, 3, 7, 6, 9, 4, 0, 5, 1]
```

Which of the following feature(s) will be selected in X_selected from X ?

Options :

6406532330376. ✖ age

6406532330377. ✔ bmi

6406532330378. ✔ bp

6406532330379. ✖ s3

6406532330380. ✔ s4

6406532330381. ✔ s5

Question Number : 190 Question Id : 640653697761 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

You've built a linear regression model to predict the temperature of a particular place 5 days from today, based on today's humidity and wind speed. The model's coefficients for the features are as follows:

- Coefficient for humidity: +2.4
- Coefficient for wind speed: −2

What does the coefficient for "wind speed" (5000) represent in this context?

Options :

6406532330386. ✔ For each additional unit of increase in wind speed, the temperature is expected to decrease by 2 units.

6406532330387. ✖ For each additional unit of increase in wind speed, the temperature is expected to increase by 2 units.

6406532330388. ✔ Humidity has a stronger impact on salary than years of experience.

6406532330389. ✖ The coefficient doesn't have any meaningful interpretation in this scenario.

Question Number : 191 Question Id : 640653697766 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

Consider the following code snippet using scikit-learn:

```
sgd_regressor = SGDRegressor()


param_dist = { 'loss': ['squared_loss', 'huber', 'epsilon_insensitive'],
    'alpha': loguniform(1e-5, 1e1),
    'penalty': ['l1', 'l2', 'elasticnet'],
    'epsilon': loguniform(1e-5, 1e-2),}


random_search = RandomizedSearchCV(sgd_regressor,
↪  param_distributions=param_dist, n_iter=15, cv=3,
↪  scoring='neg_mean_squared_error')
random_search.fit(X, y)
```

Assume all the necessary imports and X, y to be the training dataset. Which of the following statements about the given code are correct?

Options :

6406532330407. ✔ The `n_iter` parameter in `RandomizedSearchCV` controls the number of hyper-parameter combinations to try.

6406532330408. ✖ The actual number of combinations tried in the fit operation is 36.

6406532330409. ✔ The hyperparameter search space for the `alpha` parameter follows a log-uniform distribution.

6406532330410. ✔ The scoring metric used for the search is the negative mean squared error.

Question Number : 192 Question Id : 640653697772 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

Which of the following estimators are exactly same?

Options :

6406532330423.

```
SGDClassifier(loss="perceptron",
              eta0=1,
              learning_rate="constant",
              penalty=None)
```

6406532330424. ✔ `Perceptron()`

6406532330425. ✖
```
SGDClassifier(loss="percept",
              eta0=1,
              learning_rate="constant",
              penalty=None)
```

6406532330426. ✖
```
SGDRegressor(loss="percept",
             eta0=1,
             learning_rate="constant",
             penalty=None)
```

6406532330427. ✖ All of these are different.

Question Number : 193 Question Id : 640653697779 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

When employing a Support Vector Classifier (`SVC`) in scikit-learn with different values of the regularization parameter (`C`), how does the complexity of the decision boundary change? Pick all the correct statements:

Options :

6406532330446. ✖ Smaller values of `C` result in a more complex decision boundary.

6406532330447. ✔ Larger values of `C` lead to a more complex decision boundary.

6406532330448. ✖ The decision boundary tends to become simpler with increasing values of C.

6406532330449. ✔ Excessively large values of C may result in overfitting.

Question Number : 194 Question Id : 640653697790 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

Choose the correct statements.

Options :

6406532330487. ✔ Hierarchical agglomerative clustering does not need the initial number of clusters to group the data.

6406532330488. ✖ K-means clustering does not need the number of clusters to group the data.

6406532330489. ✔ K-means clustering can be agglomerative or divisive.

6406532330490. ✖ Hierarchical clustering can be agglomerative or divisive.

6406532330491. ✖ None of these

Sub-Section Number :	5
Sub-Section Id :	640653103304
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 195 Question Id : 640653697760 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 below code ?

```
import numpy as np
from sklearn.impute import SimpleImputer
data =[[5, 8, 4],[8, 4, 1],[7, np.nan, 6],[4, 2, 3],[np.nan, 6, 6]]
si = SimpleImputer(missing_values= np.nan,strategy="mean" )
si.fit(data)
print(si.statistics_)
```

Options :

6406532330382. ✖ [6,5,3.75]

6406532330383. ✖ [5,5,3]

6406532330384. ✔ [6,5,4]

6406532330385. ✖ [5,5,5]

Question Number : 196 Question Id : 640653697762 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're working on a dataset with two features and a target variable. You decide to use polynomial regression with a degree of 3 to capture potential cubic relationships. The following code snippet demonstrates the process:

```
from sklearn.preprocessing import PolynomialFeatures
import numpy as np

# Simulated data (feature: X)
X = np.array([1, 2, 3, 4])

# Reshape the features
X = X.reshape(-1, 2)

# Transform features into polynomial features
poly = PolynomialFeatures(degree=3)
X_poly = poly.fit_transform(X)
```

What will be the shape of the X_poly matrix after transforming the feature 'X' into polynomial features of degree 3?

Options :

6406532330390. ✔ (2, 10)

6406532330391. ✖ (2, 11)

6406532330392. ✖ (2, 9)

6406532330393. ✖ (2, 8)

Question Number : 197 Question Id : 640653697764 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 snippet using scikit-learn:

```
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import MinMaxScaler
from sklearn.svm import SVC
from sklearn.model_selection import GridSearchCV

pipeline = Pipeline([
    ('scaler', MinMaxScaler()),
    ('classifier', SVC())])

param_grid = {'scaler__feature_range': [(0, 1), (1, 2)],
              'classifier__C': [0.01, 0.1, 1],
              'classifier__kernel': ['linear', 'poly'],
              'classifier__degree': [2, 3]}

grid_search = GridSearchCV(pipeline, param_grid, cv=5, scoring='accuracy')
grid_search.fit(X_train, y_train)
```

Assume that `X_train` and `y_train` are the training feature matrix and label vector, respectively. Which of the following statements about the given code is correct?

Options :

6406532330399. ✖ The `grid_search.score(X_train, y_train)` will give the accuracy on the test dataset by the 4 folds in which the model gives the best parameters.

6406532330400. ✔ The `scaler__feature_range` hyperparameter is being tuned for the MinMaxScaler.

6406532330401. ✖ The pipeline always uses a polynomial kernel (`'poly'`) as the kernel for the SVC classifier.

6406532330402. ✖ A total of 12 combinations of hyperparameters were tried during the GridSearchCV.

Question Number : 198 Question Id : 640653697773 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 is likely to be the correct output of the code given below?

```
from sklearn import linear_model
clf = linear_model.Lasso(alpha=0.1)
clf.fit([[1,2], [2, 1], [2, 3]], [1, 2, 3])

linear_model.Lasso(alpha=0.1,max_iter=1000, tol=0.0001, warm_start=False,
                   fit_intercept=True)
print(clf.coef_)
```

Options :

6406532330428. ✖ [0.85,0.1,0.05]

6406532330429. ✔ [1.05 0.35]

6406532330430. ✖ [3,2,1]

6406532330431. ✖ There are some mistakes in the 3rd /4th line of code, hence it will produce error.

Question Number : 199 Question Id : 640653697783 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 :

6406532330459. ✔ dtc1

6406532330460. ✖ dtc2

6406532330461. ✖ dtc3

6406532330462. ✖ dtc4

Question Number : 200 Question Id : 640653697788 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

Output of the following code snippet is mentioned below.

```
from sklearn.cluster import KMeans
import numpy as np
X = np.array([[5, 4], [5, 6],[10, 8], [10, 12]])
kmeans = KMeans(n_clusters=2, random_state=0).fit(X)
kmeans.labels_
```

Output: array([0, 0, 1, 1], dtype=int32)

Considering the above code, Which of the following do you think as correct output of Print(kmeans.predict([[6, 5]]))

Options :

6406532330477. ✔ 0

6406532330478. ✖ 1

6406532330479. ✖ 5

6406532330480. ✖ 10

6406532330481. ✖ None of these

Sub-Section Number :	6
Sub-Section Id :	640653103305
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 201 Question Id : 640653697770 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 and its output:

keep following symbols in mind:

- >>>: Represents input code
- # : Represents comment in a code
- ... : Represents code continuation
- Without any symbols at the beginning of a line then it is output of just above input line of code.

```
>>> from sklearn.datasets import load_iris
>>> from sklearn.linear_model import SGDClasssifier

>>> X, y = load_iris(return_X_y=True)
>>> clf = SGDClasssifier(random_state=0).fit(X, y)

>>> print(y[70:80])
[1 1 0 1 1 1 1 1 1 0]

>>> print(clf.predict(X[70:80, :]))
[0 1 1 1 1 1 1 0 1 0]

>>> print(clf.score(X[70:80, :], y[70:80]))
```

What will be the output of the above code? Enter your answer correct to one decimal place.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

0.7

Question Number : 202 Question Id : 640653697774 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

What will be the output of the following code?

```
from sklearn.neighbors import KNeighborsClassifier
X_train = [[1,100],[4,400],[5,500],[6,600],[8,800],[9,900],
           [11,1100],[12,1200],[15,1500], [18,1800],[19,1900]]
y_train = [2,2,2,2,2,2,2,2,1,1,1]

X_test = [[2,200]]

knn = KNeighborsClassifier(n_neighbors= 7,
                           metric="euclidean",
                           weights= 'uniform')
knn.fit(X_train,y_train)

print(knn.predict(X_test))
```

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

2

Question Number : 203 Question Id : 640653697782 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. How many DecisionTreeClassifier models will be trained internally?

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

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

250

Sub-Section Number :	7
Sub-Section Id :	640653103306
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 204 Question Id : 640653697777 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

Consider the following code snippet:

```
from sklearn.neighbors import KNeighborsClassifier

X_train = [[1, 2], [3, 4], [5, 6]]
y_train = [0, 1, 2]

knn = KNeighborsClassifier(n_neighbors=4)
knn.fit(X_train, y_train)
```

What will be the output of the following code:

```
print(len(knn.classes_))
```

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

3

Sub-Section Number :	8
Sub-Section Id :	640653103307
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 205 Question Id : 640653697781 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 4 Max. Selectable Options : 0

Question Label : Multiple Select Question

Consider the following block of code:

```
from sklearn.datasets import load_breast_cancer
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
X,y = load_breast_cancer(as_frame = True,
                         return_X_y = True)
X_train,X_test,y_train,y_test = train_test_split(X,y,
                                                 test_size = 0.2,
                                                 random_state = 1)
clf = DecisionTreeClassifier(min_samples_split = 5,
                             min_samples_leaf = 3,
                             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 made at node N?

Options :

6406532330454. ✖ Number of samples at node N = 10. If it is split, it will result in 5 nodes in the left child and 5 nodes in the right child.

6406532330455. ✔ Number of samples at node N = 5. If it is split, it will result in 2 nodes in the left child and 3 nodes in the right child.

6406532330456. ✖ Number of samples at node N = 12. If it is split, it will result in 5 nodes in the left child and 7 nodes in the right child.

6406532330457. ✔ Number of samples at node N = 4. If it is split, it will result in 3 nodes in the left child and 1 node in the right child.

Sub-Section Number :	9
Sub-Section Id :	640653103308
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 206 Question Id : 640653697784 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 block of code for the binary classification dataset. Shape of feature matrix is (10000,4) and labels (10000,) respectively.

keep following symbols in mind:

- >>>: Represents input code
- # : Represents comment in a code
- ... : Represents code continuation
- Without any symbols at the beginning of a line then it is output of just above input line of code.

```
>>> from sklearn.linear_model import LogisticRegression,SGDClassifier
>>> from sklearn.naive_bayes import GaussianNB
>>> from sklearn.ensemble import VotingClassifier

>>> clf1 = LogisticRegression(multi_class='multinomial', random_state=1)
>>> clf2 = SGDClassifier(random_state=1)
>>> clf3 = GaussianNB()

>>> eclf = VotingClassifier(estimators=[('lr', clf1),
...                                     ('sgd', clf2),
...                                     ('gnb', clf3)],
...                         voting='soft')

>>> eclf.fit(X,y)

>>> eclf.named_estimators_['lr'].predict_proba(X[0:1])
[0.3,0.7]
>>> eclf.named_estimators_['sgd'].predict_proba(X[0:1])
[0.35,0.65]
>>> eclf.named_estimators_['gnb'].predict_proba(X[0:1])
[0.9,0.1]
```

what will be the predicted class for $\mathbf{X[0:1]}$ sample using the code given above

Options :

6406532330463. ✔ 0

6406532330464. ✖ 1

6406532330465. ✖ 2

6406532330466. ✖ 3

Question Number : 207 Question Id : 640653697785 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

Given the following code using `BaggingClassifier` with `KNeighborsClassifier` as the base estimator:

```
from sklearn.ensemble import BaggingClassifier
from sklearn.neighbors import KNeighborsClassifier
base_knn = KNeighborsClassifier(n_neighbors=3,
                                weights='distance')

bag_clf = BaggingClassifier(base_knn,
                            n_estimators=30, max_samples=100,
                            bootstrap=False, random_state=42)
```

Which of the following statements is correct?

Options :

6406532330467. ✖ Above code uses bootstrapping to generate samples for each base classifier.

6406532330468. ✖ Each base KNN classifier will be trained on a random subset of unknown number of samples.

6406532330469. ✔ Due to `weights='distance'`, nearest neighbors in each base KNN classifier will have voting power inversely proporional to their distances from the sample (sample from test set).

6406532330470. ✖ The ensemble will consist of 3 base KNN classifiers.

Sub-Section Number : 10

Sub-Section Id : 640653103309

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 208 Question Id : 640653697786 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

Consider the following code. How many different parameter combinations will be tried in GridSearchCV?

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

X, y = load_iris(as_frame = True, return_X_y = True)

param_grid = [{'max_depth':range(1, 10, 2),
               'min_samples_split': range(2, 10, 3)},

              {'min_samples_leaf': range(1, 11, 3)}]
gs = GridSearchCV(DecisionTreeClassifier(),
                  param_grid, cv = 5)
gs.fit(X,y)
```

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

19

Sub-Section Number :	11
Sub-Section Id :	640653103310
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Id : 640653697767 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 : (209 to 210)

Question Label : Comprehension

Please consider the following data and code for a regression problem with following symbols in mind:

- >>>: Represents input code
- # : Represents comment in a code
- ... : Represents code continuation
- Without any symbols at the beginning of a line then it is output of just above input line of code.

	Age	Car_color	Accidents_per_1000_Driver
0	20	Black	72
1	20	Blue	75
2	20	Red	88
3	25	Black	70
4	25	Blue	73
5	25	Red	90
6	30	Black	84
7	30	Blue	76
8	30	Red	92

Target column: **Accidents_per_1000_Driver**

```
>>> import pandas as pd
>>> from sklearn.preprocessing import OneHotEncoder
>>> from sklearn.linear_model import LinearRegression

>>> data = pd.DataFrame([[20, 'Black', 72],
...                      [20, 'Blue', 75],
...                      [20, 'Red', 88],
...                      [25, 'Black', 70],
...                      [25, 'Blue', 73],
...                      [25, 'Red', 90],
...                      [30, 'Black', 84],
...                      [30, 'Blue', 76],
...                      [30, 'Red', 92]],
                        columns=["Age","Car_color","Accidents_per_1000_Driver"])

>>> X = data.drop("Accidents_per_1000_Driver", axis=1)
>>> y = data["Accidents_per_1000_Driver"]

>>> ohe = OneHotEncoder(sparse_output=False)

>>> X[['Black', 'Blue', 'Red']] = ohe.fit_transform(X[["Car_color"]])
>>> X.drop("Car_color", axis=1, inplace=True)

>>> lr = LinearRegression().fit(X, y)

>>> print(lr.coef_)
[ 0.57, -4.67, -5.33, 10.00]

>>> print(lr.intercept_)
65.83
```

Based on the above data, answer the given subquestions.

Sub questions

Question Number : 209 Question Id : 640653697768 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

Predict number of Accidents per 1000 Driver to happen for Age 28 and driving red car ?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

91 to 93

Question Number : 210 Question Id : 640653697769 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 improve the given Linear Regression model which of the following preprocessing step you suggest?

Options :

6406532330412. ✖ OrdinalEncoder()

6406532330413. ✖ DecisionTree()

6406532330414. ✖ LabelEncoder()

6406532330415. ✔ StandardScaler()

Sub-Section Number :	12
Sub-Section Id :	640653103311
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Id : 640653697792 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 : (211 to 212)

Question Label : Comprehension

Consider the following code snippet and its output:
Code:

```
from sklearn.neural_network import MLPClassifier
from sklearn.datasets import make_classification
from sklearn.model_selection import train_test_split
X, y = make_classification(n_samples=3000, random_state=1)
X_t, X_test, y_t, y_test = train_test_split(X, y, stratify=y, random_state=1)
print(y_test[30:35])
```

Output:

```
[0 0 0 1 0]
```

Code:

```
clf = MLPClassifier(random_state=1).fit(X_t, y_t)
print(clf.predict_proba(X_test[30:35]))
```

Output:

```
[[9.97710379e-01 2.28962074e-03]
 [3.59471505e-01 6.40528495e-01]
 [9.99405675e-01 5.94325017e-04]
 [1.72404926e-03 9.98275951e-01]
 [9.86367828e-01 1.36321719e-02]]
```

Based on the above data, answer the given subquestions.

Sub questions

Question Number : 211 Question Id : 640653697793 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 following code?

```
print(clf.predict(X_test[30:35]))
```

Options :

6406532330497. ✖ [1 1 0 1 0]

6406532330498. ✖ [0 0 0 1 0]

6406532330499. ✖ [0 1 0 0 0]

6406532330500. ✔ [0 1 0 1 0]

Question Number : 212 Question Id : 640653697794 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

What will be the output of following code?

```
print(clf.score(X_test[30:35], y_test[30:35]))
```

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

0.78 to 0.82

System Commands

Section Id :	64065349268
Section Number :	9
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	16
Number of Questions to be attempted :	16
Section Marks :	105
Display Number Panel :	Yes
Section Negative Marks :	0