Response Type : Alphanumeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Answers Case Sensitive : No

Text Areas : PlainText

Possible Answers :

987987987

Question Number : 246 Question Id : 640653611165 Question Type : SA Calculator : None

Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 7

Question Label : Short Answer Question

What is the output of the **second loop** after the execution of the given script?

Response Type : Alphanumeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Answers Case Sensitive : No

Text Areas : PlainText

Possible Answers :

998877

MLP

Section Id :	64065341313
Section Number :	10
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	35

Number of Questions to be attempted :	35
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 :	64065388078
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Number : 247 Question Id : 640653611168 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 :

6406532041136. ✔ YES

6406532041137. ✖ NO

Sub-Section Number :	2
Sub-Section Id :	64065388079

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 248 Question Id : 640653611169 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 MaxAbsScaler
a = [[-3],[ 0],[-2],[ 2],[-1],[-4]]
mas = MaxAbsScaler()
scaled_a = mas.fit_transform(a)
print(scaled_a.max())
```

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

0.5

Sub-Section Number : 3

Sub-Section Id : 64065388080

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 249 Question Id : 640653611170 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Given below a y_train list which consists of pizza's ordered by the customers in a shop.

```
y_train =  [['regular', 'veg'],
            ['medium', 'veg'],
            ['regular', 'non-veg'],
            ['medium', 'non-veg']]
```

MultiLabelBinarizer from sklearn library has been used to convert the y_train into numbers, so which of the following option matches with the output using the following code ?

```
from sklearn.preprocessing import MultiLabelBinarizer
mlb = MultiLabelBinarizer(classes=['regular','medium', 'veg', 'non-veg'])
print(mlb.fit_transform(y_train))
```

Options :

6406532041139. ✖ `[1 0 1 0 1 0 0 1 0 1 0 1 0 1 1 0]`

6406532041140. ✔ `[[1 0 1 0],[0 1 1 0],[1 0 0 1],[0 1 0 1]]`

6406532041141. ✖ `[[1 0 1 0 1 0 1 0],[0 0 1 0 1 1 1 0]]`

6406532041142. ✖ `[[1 0],[1 0],[0 1],[1 0],[1 0],[0 1],[0 1],[0 1]]`

Question Number : 250 Question Id : 640653611176 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 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 preprocessing 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
pipeline = Pipeline([
    ('scaler', StandardScaler()),
    ('svm', SVC())])
# Fit the pipeline on training data
pipeline.fit(X, y)

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

What is the purpose of using the pipeline in this code snippet?

Options :

6406532041163. ✖ The pipeline combines multiple models for better model performance.

6406532041164. ✖ The pipeline allows for simultaneous training of the scaler and classifier.

6406532041165. ✔ The pipeline simplifies the code by encapsulating preprocessing and modeling steps.

6406532041166. ✖ The pipeline ensures that only linear SVM can be used for this classification task.

Question Number : 251 Question Id : 640653611177 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've developed a binary classification model to predict whether an email is spam or not. You want to evaluate the model's performance using appropriate metrics. The following options represent different evaluation metrics. Choose the one that is most suitable for assessing the model's performance in this scenario:

Options :

6406532041167. ✖ R-squared (R^2)

6406532041168. ✖ Mean Absolute Error (MAE)

6406532041169. ✔ F1-score

6406532041170. ✖ Root Mean Squared Error (RMSE)

Question Number : 252 Question Id : 640653611184 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 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)
```

In the context of the given code and scenario, what does the `power_t` parameter value of 0.5 influence during the training process?

Options :

6406532041189. ✖ It determines the degree of L2 regularization applied to the model's weights.

6406532041190.

✔ It controls the decay rate of the learning rate during each iteration.

6406532041191. ✖ It sets the threshold for early stopping based on the loss function improvement.

6406532041192. ✖ It adjusts the aggressiveness of stochastic gradient updates for faster convergence.

Question Number : 253 Question Id : 640653611191 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 is the output of the following code?

```
corpus = ['An overfitted model is a mathematical model that contains more
parameters than can be justified by the data.']
from sklearn.feature_extraction.text import CountVectorizer
vectorizer = CountVectorizer()
vectors = vectorizer.fit_transform(corpus)
print(vectors.shape)
```

Options :

6406532041214. ✖ (16,1)

6406532041215. ✖ (8,2)

6406532041216. ✖ (4,4)

6406532041217. ✔ (1,16)

Question Number : 254 Question Id : 640653611200 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 machine learning algorithms, setting the `random state` parameter to a specific value provides which of the following benefits?

Options :

6406532041248. ✖ It ensures that the model will always produce the same predictions for any input data.

6406532041249. ✖ It guarantees that the model will converge to the global optimum during training.

6406532041250. ✖ It prevents overfitting by adding randomness to the model's predictions.

6406532041251. ✔ It allows for reproducibility, ensuring consistent results across different runs.

Sub-Section Number :	4
Sub-Section Id :	64065388081
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 255 Question Id : 640653611171 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 involving the use of LabelEncoder from sklearn.preprocessing:

```
from sklearn.preprocessing import LabelEncoder

data = ["apple", "orange", "banana", "apple", "grape", "orange", "grape"]

encoder = LabelEncoder()
encoded_data = encoder.fit_transform(data)

print(encoded_data)
```

Which of the following statements is true based on the code?

Options :

6406532041143. ✔ After calling `encoder.transform(["pineapple"])`, an error will be raised.

6406532041144. ✖ LabelEncoder assigns smaller integer values to labels appearing earlier in the dataset.

6406532041145. ✖ If the `encoded_data` contains the integer 3, it corresponds to the label "apple".

6406532041146. ✔ LabelEncoder encodes the data based on the alphabetical order of the labels.

Question Number : 256 Question Id : 640653611189 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 statements are true?

Options :

6406532041206. ✖ KNeighborsClassifier with high values of n_neighbors produces complex decision boundaries.

6406532041207. ✔ KNeighborsClassifier with high values of n_neighbors produces smooth decision boundaries.

6406532041208. ✔ In KNeighborsClassifier the scale of the features(columns) can impact the decision boundaries.

6406532041209. ✖ None of these

Question Number : 257 Question Id : 640653611190 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 option(s) are correct regarding regularization?

Options :

6406532041210. ✖ Regularization is the error given by any model while predicting the values for the test set.

6406532041211. ✖ It helps in decreasing the bias of the training model.

6406532041212. ✔ Compare to without regularized model regularization decreases the variance of the training model

6406532041213. ✔ Predictions made with the Ridge Regression are less sensitive to weights(coefficients) than the Linear Regression.

Sub-Section Number :	5
Sub-Section Id :	64065388082
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 258 Question Id : 640653611172 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 sklearn code snippet that employs a Pipeline for data preprocessing:

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

data = [[1, 5], [5, 15], [0, 10], [1, 15]]

pipeline = Pipeline([('scaler', MinMaxScaler()),
                     ('binarize', Binarizer(threshold=0.1))])

transformed_data = pipeline.fit_transform(data)
```

What will be the output of the following code?

```
print(transformed_data[0])
```

Options :

6406532041147. ✖ `[0.  0.]`

6406532041148. ✔ [1. 0.]

6406532041149. ✖ [1. 1.]

6406532041150. ✖ [0. 1.]

Question Number : 259 Question Id : 640653611173 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 that uses PCA from sklearn.decomposition for dimensionality reduction on a dataset with 10 features:

```
import numpy as np
from sklearn.decomposition import PCA

np.random.seed(42)
X = np.random.rand(100, 10)

pca = PCA(n_components=4)
X_pca = pca.fit_transform(X)

explained_variance = np.sum(pca.explained_variance_ratio_)
```

After executing this code, which of the following statements is true regarding the `explained_variance`?

Options :

6406532041151. ✖ The explained_variance represents the combined explained variance of all 10 principal components.

6406532041152. ✔ To get a dataset with more than 90% of the original variance, you can set `n_components` to 0.9 when initializing PCA.

6406532041153. ✖ The `X_pca` dataset contains the original 10 features but with reduced dimensionality.

6406532041154. ✖ The principal components in `X_pca` are based only on the most significant original features.

Question Number : 260 Question Id : 640653611174 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've built a linear regression model to predict the salary of employees based on their years of experience and education level. The model's coefficients for the features are as follows:

- Coefficient for Years of Experience: 5000
- Coefficient for Education Level: 3000

What does the coefficient for "Years of Experience" (5000) represent in this context?

Options :

6406532041155. ✔ For each additional year of experience, an employee's salary is expected to increase by 5000.

6406532041156. ✖ For each additional year of experience, an employee's salary is expected to decrease by 5000.

6406532041157. ✖ Education level has a stronger impact on salary than years of experience.

6406532041158. ✖ The coefficient doesn't have any meaningful interpretation in this scenario.

Question Number : 261 Question Id : 640653611175 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 that includes data points for a single feature 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, 5])

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

# 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 :

6406532041159. ✖ (5, 1)

6406532041160. ✖ (5, 3)

6406532041161. ✔ (5, 4)

6406532041162. ✖ (5, 6)

Question Number : 262 Question Id : 640653611178 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

Choose the correct output of the following code?

```
data = [[1, 3],
        [2, 4]]
from sklearn.preprocessing import PolynomialFeatures
pf = PolynomialFeatures(degree=3,interaction_only=False)
print(pf.fit_transform(data))
```

Options :

6406532041171. ✖
```
[[1, 2, 3, 2, 4],
[1, 2, 4, 6, 8]]
```

6406532041172. ✖
```
[[1, 1, 3, 3],
[1, 2, 4, 8]]
```

6406532041173. ✔
```
[[1, 1, 3, 1, 3,  9, 1,  3,  9, 27],
 [1, 2, 4, 4, 8, 16, 8, 16, 32, 64]]
```

6406532041174. ✖
```
[[1, 1, 2, 2, 3, 3,  4,  4],
 [1, 1, 4, 4, 9, 9, 16, 16]]
```

Question Number : 263 Question Id : 640653611182 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

Upon successfully training a `LogisticRegression` model from `sklearn.linear model` on a binary classification problem, you decide to inspect some of its attributes. Which of the following attributes would provide you with the estimated probabilities that each sample belongs to a particular class?

Options :

6406532041181. ✖ `model.classes_`

6406532041182. ✖ `model.decision_function()`

6406532041183. ✔ `model.predict_proba()`

6406532041184. ✖ `model.n_iter_`

Question Number : 264 Question Id : 640653611183 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

Imagine you're training a Perceptron using sklearn with the following code:

```
from sklearn.linear_model import Perceptron

X = [[0, 0.5], [1, 1.5], [1, 2], [2, 3]]
y = [-1, -1, 1, 1]


clf = Perceptron( tol=None, shuffle=True, random_state=42)
clf.fit(X, y)

iterations = clf.n_iter_
```

Given the linearly separable nature of the data, how many iterations would it most likely take for the perceptron to converge? What will be the value of `iterations`?

Options :

6406532041185. ✖ `iterations = 1`

6406532041186. ✖ `iterations = 10`

6406532041187. ✔ `iterations` value can vary since the data is being shuffled in each epoch.

6406532041188. ✖ `iterations = 5`

Question Number : 265 Question Id : 640653611185 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 multi-class classification task with high-dimensional data. You decide to use the `RidgeClassifier` algorithm for its regularization properties. How-ever, you're concerned about selecting an appropriate value for the regularization parameter alpha. Which of the following statements about choosing the alpha value for 'RidgeClassifier' is correct?

Options :

6406532041193. ✖ A higher alpha value results in stronger regularization, leading to larger feature coefficients.

6406532041194. ✖ The alpha value has no impact on the regularization strength in 'RidgeClassifier'.

6406532041195. ✔ Cross-validation can be used to find the optimal alpha value.

6406532041196. ✖ 'RidgeClassifier' automatically determines the alpha value based on the data distribution.

Question Number : 266 Question Id : 640653611198 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 applying the K-means clustering algorithm to a dataset with a large number of data points. You decide to use the K-means++ initialization method for better convergence and clustering quality. What is the primary advantage of using K-means ++ initialization over random initialization?

Options :

6406532041240. ✖ K-means++ initialization guarantees finding the global optimum of the clustering solution.

6406532041241. ✖ K-means++ initialization speeds up the convergence of the algorithm.

6406532041242. ✖ K-means++ initialization reduces the number of clusters needed for accurate

results.

6406532041243. ✔ K-means++ initialization helps avoid local optima during clustering.

Question Number : 267 Question Id : 640653611199 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 you are working on a classification task involving handwritten digit recognition using scikit-learn's MLPClassifier. You have a large dataset of digit images and want to train a neural network for this task. However, you're concerned about overfitting and want to make sure the training process stops at the right time to avoid this issue. In the context of training an MLPClassifier for handwritten digit recognition,how does the early stopping parameter help prevent overfitting?

Options :

6406532041244. ✖ It terminates the training process as soon as the specified number of hidden layers are trained.

6406532041245. ✖ It automatically decreases the learning rate during training to slow down the optimization process.

6406532041246. ✔ It pauses the training when the model's performance on a validation set stops improving.

6406532041247. ✖ It enforces a maximum limit on the number of epochs during the training process.

Question Number : 268 Question Id : 640653611201 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 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, 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 max_iter=100 in the MLPClassifier signify?

Options :

6406532041252. ✖ It specifies the maximum number of features to be used during training.

6406532041253. ✖ It determines the maximum number of neurons in the hidden layers.

6406532041254. ✔ It sets the maximum number of iterations for the training process.

6406532041255. ✖ It controls the maximum number of epochs for the training process.

Question Number : 269 Question Id : 640653611202 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 multi-class classification task using the 'MLPClassifier' from scikit-learn. The dataset contains features with varying scales, and you're considering whether to scale the features

before training the model. How might feature scaling impact the prediction accuracy of the 'MLPClassifier'?

Options :

6406532041256. ✖ Feature scaling has no effect on the accuracy of 'MLPClassifier' predictions.

6406532041257. ✖ Scaling features is only relevant if the dataset contains categorical features.

6406532041258. ✖ Feature scaling can lead to overfitting and decreased prediction accuracy.

6406532041259. ✔ Scaling features can help the model converge faster and improve prediction accuracy.

Sub-Section Number :	6
Sub-Section Id :	64065388083
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 270 Question Id : 640653611179 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 $Y = W_0X_0 + W_1X_1 + W_2X_2 + \epsilon$ and given that $W_2 = \frac{6}{5} * W_1$ 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], [2,2.4], [4,4.8], [6,7.2]]
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 :

0.2039 to 0.2052

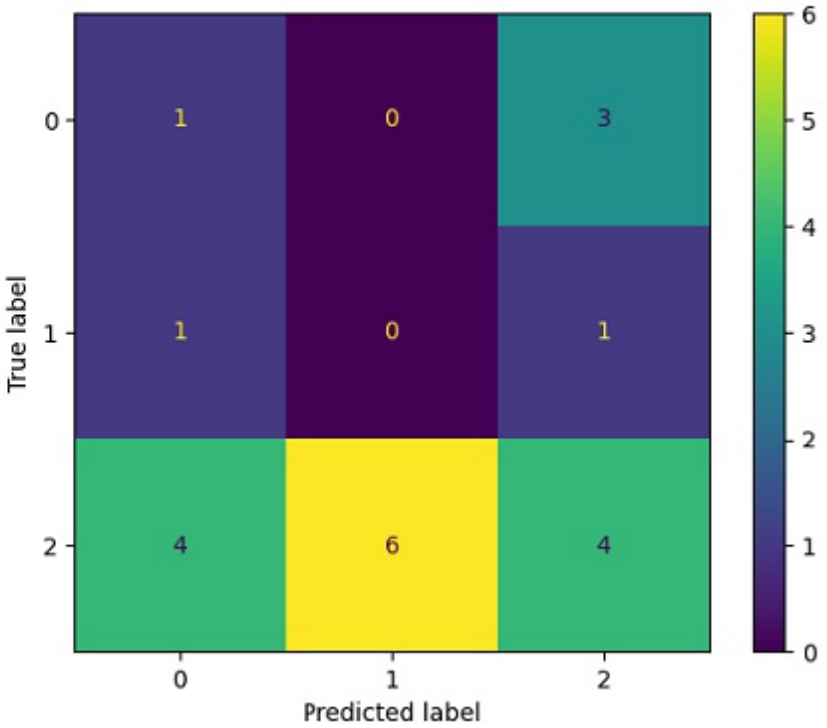
Question Number : 271 Question Id : 640653611180 Question Type : SA Calculator : None

Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3

Question Label : Short Answer Question

Using the confusion matrix given below. What is the recall score for the label (class) 2? (upto 3 digits after the decimal)



Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

0.283 to 0.291

Question Number : 272 Question Id : 640653611186 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 is the output of the following code?

```
from sklearn.neighbors import KNeighborsClassifier
X = [[2,3], [5,6], [10, 11], [15,16], [20,21]]
y = [2, 2, 0, 1, 0]
knn = KNeighborsClassifier (n_neighbors=3, metric='euclidean',weights='uniform')
knn.fit (X, y)
print (knn.predict ([ [8,9]]))
```

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

2

Sub-Section Number :	7
Sub-Section Id :	64065388084
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 273 Question Id : 640653611181 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 code snippet that employs LogisticRegression from sklearn on a feature matrix X and corresponding label vector y:

```
from sklearn.linear_model import LogisticRegression

model = LogisticRegression(C=0.8, multi_class='multinomial', max_iter=1000)
model.fit(X, y)
```

Given the code above, which of the following statements is true?

Options :

6406532041177. ✖ The logistic regression model is set up for binary classification.

6406532041178. ✖ The model does not use any regularization because the parameter C is set.

6406532041179. ✔ The model has been specifically set up to handle a multi-class classification problem using a softmax regression approach.

6406532041180. ✔ The model will iterate over the data a maximum of 1000 times, irrespective of convergence.

Sub-Section Number : 8

Sub-Section Id : 64065388085

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 274 Question Id : 640653611187 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 estimators can implement the `partial_fit` method ?

Options :

6406532041198. ✔ MultinomialNB

6406532041199. ✖ RandomForestRegressor

6406532041200. ✔ MiniBatchKMeans

6406532041201. ✖ LogisticRegressor

Question Number : 275 Question Id : 640653611188 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 :

6406532041202. ✔ 'poly',

6406532041203. ✖ 'lasso'

6406532041204. ✖ 'scale'

6406532041205. ✔ 'sigmoid'

Sub-Section Number :	9
Sub-Section Id :	64065388086
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 276 Question Id : 640653611192 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 snippet where two decision trees are trained on the same dataset:

```
from sklearn.tree import DecisionTreeClassifier
from sklearn.datasets import load_digits

data = load_digits()
X, y = data.data, data.target

tree_1 = DecisionTreeClassifier(splitter='best', max_leaf_nodes=5)
tree_1.fit(X, y)

tree_2 = DecisionTreeClassifier(splitter='random', max_leaf_nodes=None)
tree_2.fit(X, y)
```

Given the configurations of tree_1 and tree_2, which decision tree is more likely to overfit the training data?

Options :

6406532041218. ✖ tree_1

6406532041219. ✔ tree_2

Question Number : 277 Question Id : 640653611193 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 :

6406532041220. ✖ The ensemble uses bootstrapping to generate samples for each base classifier.

6406532041221. ✔ Each base KNN classifier will be trained on a random subset of exactly 100 samples.

6406532041222. ✖ Due to `weights='distance'`, each base KNN classifier will treat all neighbors equally in terms of voting power.

6406532041223. ✖ The ensemble will consist of 3 base KNN classifiers.

Question Number : 278 Question Id : 640653611194 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 you are training a `RandomForestClassifier` from `sklearn` for a binary classification problem. You're particularly interested in reducing overfitting and enhancing the model's generalization to unseen data. Which combination of parameter settings would be the MOST effective in achieving this goal?

Options :

6406532041224. ✖ Setting `n_estimators` to a high value (e.g., 1000) and `max_depth` to None.

6406532041225. ✖ Using `criterion='gini'`, setting `min_samples_split` to 2, and `min_samples_leaf` to 1.

6406532041226. ✖ Using `criterion='entropy'` with `oob_score=True` and `max_samples=0.5`.

6406532041227. ✔ Setting `max_features` to 'sqrt', `bootstrap` to True, and increasing the value of `min_samples_leaf`.

Question Number : 279 Question Id : 640653611195 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 training a classification model using the `DecisionTreeClassifier` from `sklearn.tree`. Which of the following statements about its parameters and attributes is MOST accurate?

Options :

6406532041228. ✔ The `ccp_alpha` parameter determines the complexity of the tree, with higher values resulting in simpler trees. The `tree_.node_count` attribute gives the total number of nodes in the tree.

6406532041229. ✖ If `splitter='best'`, the classifier guarantees that the resulting tree will have the highest possible accuracy on the training set.

6406532041230. ✖ The `criterion='entropy'` parameter means that the decision tree will split nodes to maximize information gain, while the `tree_.impurity` attribute retrieves the impurity of the root node.

6406532041231. ✖ If the `class_weight` parameter is set to 'balanced', the decision tree will always have balanced classes in its leaf nodes.

Question Number : 280 Question Id : 640653611196 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

You're aiming to optimize an AdaBoostClassifier that uses a DecisionTreeClassifier as its base estimator. You decide to use GridSearchCV from sklearn.model selection to search for the best hyperparameters. In the given parameter grids for GridSearchCV, parameters n_estimator and learning_rate are meant for the 'AdaBoostClassifier', while the others are for the 'DecisionTreeClassifier'. Which of the following sets of parameters is the MOST comprehensive in testing the capabilities of both the 'AdaBoostClassifier' and its base estimator?

Options :

6406532041232. ✖

```
{'AdaBoostClassifier_n_estimators': [50, 100, 150],
    ''AdaBoostClassifier_learning_rate': [0.01, 0.1, 1]}
```

6406532041233. ✖

```
{'max_depth': [1, 2, 3],
    'n_estimators': [50, 100],
    'learning_rate': [0.01, 0.1, 1]}
```

6406532041234. ✖

```
{'DecisionTreeClassifier_criterion': ['gini', 'entropy'],
    'DecisionTreeClassifier_splitter': ['best', 'random'],
    'n_estimators': [50, 100],
    'learning_rate': [0.1, 1]}
```

6406532041235. ✔

```
{'base_estimator__max_depth': [1, 2, 3],
    'base_estimator__criterion': ['gini', 'entropy'],
    'n_estimators': [30, 50],
    'learning_rate': [0.05, 0.1, 0.5]}
```

Question Number : 281 Question Id : 640653611197 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

You're evaluating the results of a clustering algorithm and you calculate the silhouette score for the clustering solution. The obtained silhouette score is -0.15. What can you infer from this silhouette score?

Options :

6406532041236. ✖ The clustering solution has well-defined and distinct clusters.

6406532041237. ✖ The silhouette score indicates a random distribution of data points across clusters.

6406532041238. ✖ The data points are equally spaced across clusters.

6406532041239. ✔ The clustering solution is flawed and the data points might be assigned to incorrect clusters.

PDSA

Section Id :	64065341314
Section Number :	11
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	27
Number of Questions to be attempted :	27
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 :	64065388087
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Number : 282 Question Id : 640653611203 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction