6406532731254. ✖ Points except {F, A, C, I} do not play any role in determining optimal weight vector.

MLP

Section Id :	64065356662
Section Number :	12
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	37
Number of Questions to be attempted :	37
Section Marks :	100
Display Number Panel :	Yes
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 :	640653118710
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Number : 250 Question Id : 640653815196 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 PRACTICES (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 :

6406532731268. ✔ YES

6406532731269. ✖ NO

Question Number : 251 Question Id : 640653815197 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

Instruction:

Note: *For numerical answer type questions enter your answer correct upto 2 decimal places without rounding off, unless stated otherwise.*

Options :

6406532731270. ✔ Instruction has been mentioned above.

6406532731271. ✖ This Instruction is just for a reference & not for an evaluation.

Sub-Section Number :	2
Sub-Section Id :	640653118711
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 252 Question Id : 640653815204 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

Select multilabel multiclass classification problems:

Options :

6406532731285. ✔ There is a collection of photographs. Each photograph can have multiple animals, e.g., cats, dogs and birds. Your model should indicate all the animals which are present.

6406532731286. ✖ From appropriate weather data, your model must predict, average temperature and average humidity for next seven days.

6406532731287. ✖ Predicting expected price of a second hand car with appropriate features.

6406532731288. ✖ None of these.

Question Number : 253 Question Id : 640653815211 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Which of the following options is the correct method to shuffle training data after each epoch in SGDRegressor?

Options :

6406532731307. ✔
```
from sklearn.linear_model import SGDRegressor
linear_regressor = SGDRegressor(shuffle=True)
```

6406532731308. ✖
```
from sklearn.preprocessing import SGDRegressor
linear_regressor = SGDRegressor(shuffle_per_epoch=True)
```

6406532731309. ✖
```
from sklearn.SGDRegressor import linear_model
linear_regressor = SGDRegressor(learning_rate='constant',
                                eta0=1e-2)
```

6406532731310. ✖ None of these

Question Number : 254 Question Id : 640653815212 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Which of the following can be used (with appropriate supporting code) to compute training error after each iteration?

Options :

6406532731311. ✔

```
SGDRegressor(max_iter=1,
             warm_start=True,
             fit_intercept=True,
             random_state=0,
             learning_rate='optimal')
```

6406532731312. ✖

```
SGDRegressor(max_iter=2,
             warm_start=True,
             fit_intercept=True,
             random_state=42,
             learning_rate='optimal')
```

6406532731313. ✖

```
SGDRegressor(max_iter=1,
             warm_start=False,
             fit_intercept=True,
             random_state=0,
             learning_rate='optimal')
```

6406532731314. ✖

```
SGDRegressor(max_iter=1,
             warm_start=False,
             fit_intercept=True,
             random_state=0,
             learning_rate='invscaling')
```

6406532731315. ✖ None of these

Question Number : 255 Question Id : 640653815213 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

For the given code below in which you use how many models will get trained or what will be the length of scores variable ?

```
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import LeaveOneOut
from sklearn.linear_model import LogisticRegression
from sklearn.datasets import make_classification
X, y = make_classification(n_samples=1024, n_features=82, n_classes=2,
↪  random_state=42)
estimator = LogisticRegression()
loocv = LeaveOneOut()
scores = cross_val_score(estimator, X, y, cv=loocv)
```

Options :

6406532731316. ✖ 1106

6406532731317. ✖ 82

6406532731318. ✔ 1024

6406532731319. ✖ 5

6406532731320. ✖ None of these

Question Number : 256 Question Id : 640653815214 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 purpose of k-fold cross-validation ?

Options :

6406532731321. ✖ To split data into training and testing sets.

6406532731322. ✖ To tune hyperparameters.

6406532731323. ✔ To evaluate model performance on multiple subsets.

6406532731324. ✖ To preprocess data.

Question Number : 257 Question Id : 640653815215 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Consider below statements and choose the correct option
Statement 1 : fit method in scikit-learn helps in learning the parameters of any Classifier or Regressor.
Statement 2 : The transform method in scikit-learn is utilized to apply a transforming function and convert the feature matrix after fitting the Classifier or Regressor.

Options :

6406532731325. ✔ Statement 1 is True and Statement 2 is False

6406532731326. ✖ Statement 2 is True and Statement 1 is False

6406532731327. ✖ Both the statements are True

6406532731328. ✖ Both the statements are False

Question Number : 258 Question Id : 640653815219 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Which assumption does Naive Bayes make about the features?

Options :

6406532731332. ✔ They are independent of each other.

6406532731333. ✖ They are always a numerical representation of the text data.

6406532731334. ✖ They are linearly related.

6406532731335. ✖ They are categorical.

6406532731336. ✖ None of these.

Sub-Section Number :	3
Sub-Section Id :	640653118712
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 259 Question Id : 640653815207 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

Given the following code snippet that preprocesses a dataset with both continuous and categorical features using `sklearn.preprocessing` tools, what will be the first row of the `X_transformed` array after preprocessing?

```
import numpy as np
from sklearn.preprocessing import OneHotEncoder
from sklearn.preprocessing import MinMaxScaler, StandardScaler
from sklearn.compose import ColumnTransformer


X = np.array([[4.0, 'avocado'],
              [3.0, 'dragon fruit'],
              [2.0, 'sapodilla'],
              [7.0, 'papaya']])

preprocessor = ColumnTransformer(
                transformers=[('num1', MinMaxScaler(), [0]),
                              ('cat', OneHotEncoder(), [1]),
                              ('num2', StandardScaler(), [0]),])

X_transformed = preprocessor.fit_transform(X)
print(X_transformed[0])
```

Options :

6406532731295. ✔ $[0.4, 1, 0, 0, 0, 0]$

6406532731296. ✖ $[0.5, 0, 1, 0, 0, 1.66]$

6406532731297. ✖ $[0.25, 0, 1, 0, -1.66]$

6406532731298. ✖ $[0.4, 0, 0, 0, 1, 0.66]$

6406532731299. ✖ None of these

Question Number : 260 Question Id : 640653815231 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 containing customer purchase data, and you want to segment the customers into distinct groups based on their purchasing behavior. Each data point represents a customer and includes features like "Total Amount Spent" and "Number of Items Purchased." Which algorithm is suitable for this scenario?

Options :

6406532731376. ✖ Linear Regression

6406532731377. ✖ Decision Tree

6406532731378. ✔ K-means Clustering

6406532731379. ✖ Support Vector Machine

6406532731380. ✖ None of these

Question Number : 261 Question Id : 640653815232 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 building an `MLPClassifier` for a dataset with a large number of features. The goal is to predict whether an online user will purchase a product based on their browsing behavior. You're trying to decide the appropriate number of neurons in the hidden layers of the neural network. Which statement about adjusting the `hidden_layer_sizes` parameter is correct?

Options :

6406532731381. ✖ Increasing the number of neurons in hidden layers will always lead to better model performance.

6406532731382. ✖ Decreasing the number of neurons in hidden layers reduces the model's capacity to capture complex patterns.

6406532731383. ✖ The number of neurons in hidden layers does not significantly affect the

model's performance.

6406532731384. ✔ Finding the optimal number of neurons is a trial-and-error process and may require experimentation.

Question Number : 262 Question Id : 640653815234 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

How does agglomerative clustering handle outliers?

Options :

6406532731389. ✖ It ignores outliers during the clustering process.

6406532731390. ✔ It assigns outliers to the nearest cluster.

6406532731391. ✖ It creates separate clusters for outliers.

6406532731392. ✖ It removes outliers from the dataset before clustering.

Question Number : 263 Question Id : 640653815235 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 is agglomerative clustering?

Options :

6406532731393. ✔ A hierarchical clustering technique that starts with each data point as its cluster and merges the closest clusters iteratively.

6406532731394. ✖ A method for partitioning data into a predefined number of clusters.

6406532731395. ✖ A clustering algorithm that uses centroids to iteratively assign data points to clusters.

6406532731396. ✖ A dimensionality reduction technique that projects data onto a lower-dimensional space.

Question Number : 264 Question Id : 640653815236 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 initialization method is best in KMeans to select initial cluster centroids?

Options :

6406532731397. ✖ Random initialization

6406532731398. ✔ K-means++ initialization

6406532731399. ✖ Hierarchical agglomerative initialization

6406532731400. ✖ Weighted initialization

Question Number : 265 Question Id : 640653815237 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

In K-Means clustering, what does the `inertia_` attribute represent?

Options :

6406532731401. ✖ The distance between cluster centroids

6406532731402. ✖ The number of clusters formed

6406532731403. ✔ Sum of squared distances of samples to their closest cluster center.

6406532731404. ✖ The silhouette coefficient

Sub-Section Number : 4

Sub-Section Id : 640653118713

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 266 Question Id : 640653815226 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 using a `DecisionTreeClassifier` from `sklearn.tree` to build a classification model. Which of the following statements is MOST accurate regarding the parameters and attributes of this classifier?

Options :

6406532731355. ✔ The depth of the tree always must be less than equal to the `max_depth` parameter value, while the `tree_.max_depth` attribute retrieves the depth of the actual tree that was built.

6406532731356. ✖ Setting `min_samples_split` to a value greater than 2 can prevent the tree from splitting on features that have very minimal influence, but this guarantees that all leaf nodes will contain fewer samples than this value.

6406532731357. ✖ 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.

6406532731358. ✖ If the `class_weight` parameter is set to 'balanced', the decision tree will always have balanced classes in its leaf nodes.

Question Number : 267 Question Id : 640653815228 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 :

6406532731363. ✖

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

6406532731364. ✖

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

6406532731365. ✖

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

6406532731366. ✔

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

Question Number : 268 Question Id : 640653815229 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=5)

bag_clf = BaggingClassifier(base_knn, n_estimators=50, max_samples=0.5,
↪  bootstrap=True, n_jobs=-1)
```

Which of the following statements is correct?

Options :

6406532731367. ✖ `Bag_clf` will throw an error as it only accepts decision tree classifiers as base classifiers.

6406532731368. ✖ Each base KNN classifier will be trained on the entire dataset.

6406532731369. ✔ The `max_samples=0.5` parameter means each base estimator in the ensemble is trained on 50% of the training samples,

6406532731370. ✖ The ensemble will use sequential computation due to `n_jobs=-1`.

6406532731371. ✖ None of these

Question Number : 269 Question Id : 640653815230 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 using a `RandomForestClassifier` from `sklearn` for a multi-class classification problem. You want to ensure diversity among the trees to avoid overfitting and increase robustness. Which combination of parameter settings would contribute MOST to achieving this objective?

Options :

6406532731372. ✖ Setting `n_estimators` to 10, `max_depth` to 3, and using `criterion='entropy'`.

6406532731373. ✖ Increasing the value of `n_estimators`, setting `max_features` to a value less than the total number of features, and setting `bootstrap` to False.

6406532731374. ✔ Setting `n_estimators` to a high value, using `criterion='gini'`, and setting `max_samples` to a value less than the total number of samples.

6406532731375. ✖ Setting `max_depth` to None, `min_samples_split` to 2, and `min_samples_leaf` to 1.

Sub-Section Number :	5
Sub-Section Id :	640653118714
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 270 Question Id : 640653815205 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

What is the solution for overfitting?

Options :

6406532731289. ✖ To have less constraints/regularization

6406532731290. ✔ To have more constraints/regularization

6406532731291. ✖ Delete a significant portion of data

6406532731292. ✔ Increase the dataset size.

6406532731293.

✖ None of these.

Question Number : 271 Question Id : 640653815210 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

Select all the correct options:

Options :

6406532731302. ✔ The more the SGD iterations, the lesser the fluctuations in training error.

6406532731303. ✔ More iterations require more computation time.

6406532731304. ✔ The tol (error tolerance) parameter restricts the number of iterations performed.

6406532731305. ✔ Training error might not consistently decrease while performing SGD iterations.

6406532731306. ✖ None of these

Question Number : 272 Question Id : 640653815222 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 :

6406532731339. ✖ 'lasso'

6406532731340. ✔ 'linear',

6406532731341. ✔ 'rbf',

6406532731342. ✖ 'scale'

Question Number : 273 Question Id : 640653815223 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 statements are true?

Options :

6406532731343. ✔ `KNeighborsClassifier` with low values of `n_neighbors` produces complex decision boundaries.

6406532731344. ✖ `KNeighborsClassifier` with low values of `n_neighbors` produces smooth decision boundaries.

6406532731345. ✔ In `KNeighborsClassifier` the scale of the features (columns) can impact the decision boundaries.

6406532731346. ✖ None of these.

Question Number : 274 Question Id : 640653815224 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 options are correct regarding regularization?

Options :

6406532731347. ✖ It is a technique used to minimize the adjusted loss function and avoid underfitting.

6406532731348. ✔ It helps in increasing the bias of the training model.

6406532731349. ✖ It determines the rows to be selected as a training dataset.

6406532731350. ✔ Elastic net regularization is a combination of L1 and L2 regularization both.

Sub-Section Number :	6
Sub-Section Id :	640653118715
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 275 Question Id : 640653815227 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 techniques are used in decision trees to make decisions or to measure the quality of a split while training the model?

Options :

6406532731359. ✔ Entropy

6406532731360. ✖ RoC Curve

6406532731361. ✖ Cross Entropy

6406532731362. ✔ Gini Impurity

Question Number : 276 Question Id : 640653815233 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 approaches is(are) helpful to find a good value for *k* in *k*−means clustering algorithm?

Options :

6406532731385. ✔ Plotting an Elbow curve.

6406532731386. ✖ Using classifiers before making clusters.

6406532731387. ✔ Plotting Silhouette coefficient for various values of *k*.

6406532731388. ✖ Using k-fold cross validation

Sub-Section Number :	7
Sub-Section Id :	640653118716
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 277 Question Id : 640653815225 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 :

6406532731351. ✔ 10 number of samples at node N. If it is split, it can split such that 2 samples in the left child and 8 samples in the right child.

6406532731352. ✖ 6 number of samples at node N. If it is split, it can split such that 3 samples in the left child and 3 samples in the right child.

6406532731353. ✖ 12 number of samples at node N. If it is split, it can split such that 5 samples in the left child and 7 samples in the right child.

6406532731354. ✔ 4 number of samples at node N. If it is split, it can split such that 3 samples in the left child and 1 samples in the right child.

Sub-Section Number :	8
Sub-Section Id :	640653118717
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 278 Question Id : 640653815206 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 = [[-3.5],[ 0],[-2.8],[ 2.0],[-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

Question Number : 279 Question Id : 640653815216 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 following data points:

```
import numpy as np
X = np.array([[1,1],[10,11],[5,5],[25,18],[-1,-1]])
y = np.array([0,1,0,1,1]).reshape(-1,1)
```

What will be the highest accuracy a perceptron model can achieve on this dataset without any feature engineering?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

0.8

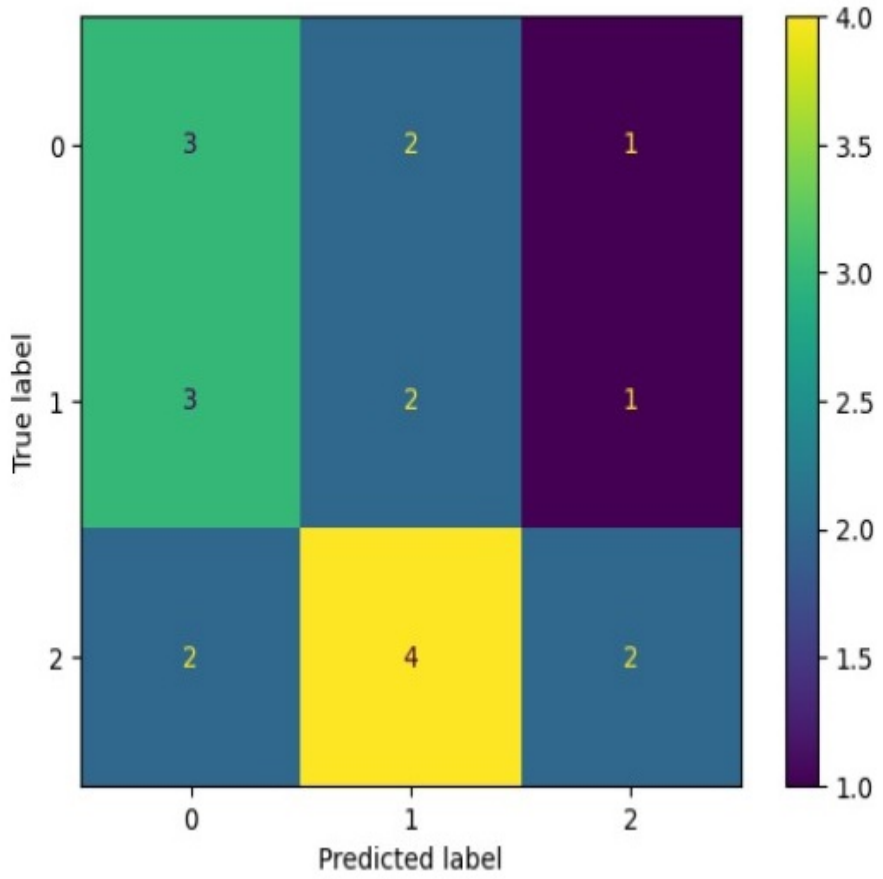
Question Number : 280 Question Id : 640653815217 Question Type : SA Calculator : None

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

Correct Marks : 2

Question Label : Short Answer Question

Using the confusion matrix given below. What is the precision score for the label (class) 0 ?



Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

0.37 to 0.38

Question Number : 281 Question Id : 640653815220 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 Python code snippet and the given dataset that has been stored in the `data` variable that demonstrates the use of `GaussianNB` from scikit-learn:

	Outlook	Temperature	Humidity	Windy	Play Golf
0	Rainy	Hot	High	False	No
1	Overcast	Hot	High	False	Yes
2	Sunny	Cool	Normal	False	Yes
3	Sunny	Cool	Normal	True	No
4	Overcast	Cool	Normal	True	Yes
5	Rainy	Cool	Normal	False	Yes
6	Rainy	Hot	High	True	No
7	Overcast	Hot	High	False	Yes
8	Sunny	Cool	Normal	True	No
9	Overcast	Cool	Normal	True	Yes
10	Rainy	Cool	Normal	False	Yes
11	Overcast	Hot	Normal	False	Yes

```
from sklearn.preprocessing import OneHotEncoder, LabelEncoder
from sklearn.naive_bayes import GaussianNB

X = data.drop("Play Golf", axis=1)
y = data["Play Golf"]

X = OneHotEncoder(sparse_output=False).fit_transform(X)
y = LabelEncoder().fit_transform(y)

estimator = GaussianNB()

estimator.fit(X,y)

print(estimator.class_prior_) # gives the priori of labels(y)
```

What is the prior probability of the label being "No"? i.e. $p(y = \text{“}No\text{”})$

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

0.32 to 0.35

Question Number : 282 Question Id : 640653815221 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 is the output of the following code?

```
from sklearn.neighbors import KNeighborsClassifier
X = [[2,3], [5,6], [10, 11], [15,16], [20,21]]
y = [0, 1, 1, 1, 2]
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 :

1

Sub-Section Number :	9
Sub-Section Id :	640653118718
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 283 Question Id : 640653815208 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 snippet:

```
from sklearn.datasets import fetch_california_housing, load_iris
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler, PolynomialFeatures
from sklearn.pipeline import Pipeline, FeatureUnion

X,y = load_iris(return_X_y= True)

polynomial_transform = PolynomialFeatures(degree=3,
                                          interaction_only=False,
                                          include_bias=False)



combined_features = FeatureUnion([('poly', polynomial_transform),
                                  ('pca', PCA(n_components=2))])

pipeline = Pipeline([('features', combined_features),
                     ('scaler', MinMaxScaler())])

X_transformed = pipeline.fit_transform(X)

print(X_transformed.shape)
```

If the shape of X is $(150, 4)$, what will be the number of features in X_transformed?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

36

Question Number : 284 Question Id : 640653815209 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 snippet? Assume necessary imports.

```
X = np.array([[1,6],
              [-2,0],
              [-0.25, 3.5]])

pca_transform = PCA(n_components=2)

X_transformed = pca_transform.fit_transform(X)

print(pca_transform.explained_variance_ratio_[0])
```

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

1.00

Question Number : 285 Question Id : 640653815218 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 snippet?

```
from sklearn.linear_model import Perceptron
# Sample data
X = [[0, 0], [0, 1], [1, 0], [1, 1]]
y = [0, 0, 0, 1]

clf = Perceptron(tol=None, shuffle=False)
clf.fit(X, y)

print(clf.predict([[2, 2]]))
```

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

1

Sub-Section Number :	10
Sub-Section Id :	640653118719
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Id : 640653815198 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 : (286 to 290)

Question Label : Comprehension

Consider following code snippet, assume necessary imports:

```
df=pd.DataFrame(data={"Name":['Akash',
                              'Brajesh',
                              'Charu',
                              'Deepak'],
                      "Maths":[34,43,56,77],
                      "English":[23,45,67,82],
                      "Hindi":[53,35,np.nan,"hi"],},
                index = range(11,15))
```

Based on the above data, answer the given subquestions.

Sub questions

Question Number : 286 Question Id : 640653815199 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

Consider the following code snippet:

```
selection = df.loc[13:15, 'Maths':'English']
```

Which of the following will be equivalent to the given code snippet?

Options :

6406532731272. ✔ `selection = df.iloc[[2,3], [1,2]]`

6406532731273. ✖ `selection = df.iloc[[1,3], [1,2]]`

6406532731274. ✔ `selection = df.iloc[2:4, 1:3]`

6406532731275. ✖ `selection = df.iloc[1:3, 1:2]`

6406532731276. ✖ None of these

Question Number : 287 Question Id : 640653815200 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 is the output of the following code snippet:

```
df.loc[13, 'English'] + df.iloc[0,3]
```

Enter -1, if you think the given statement will generate an error.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

120

Question Number : 288 Question Id : 640653815201 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 is the output of the following code snippet:

```
print(df.Hindi.apply(type).nunique())
```

Enter -1, if you think the given statement will generate an error.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

3

Question Number : 289 Question Id : 640653815202 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 is the output of the following code snippet:

```
def getPassing(aRow):
  if aRow['Maths']>=35 and aRow['English']>=35:
    return True
  return False

df.apply(getPassing).sum()
```

Enter -1, if you think the given statement will generate an error.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

-1

Question Number : 290 Question Id : 640653815203 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 snippet:

```
df=df.dropna()
```

Choose correct options from following:

Options :

6406532731280. ✔ The number of rows will decrease in the dataset.

6406532731281. ✖ The number of columns will decrease in the dataset.

6406532731282. ✖ The number of columns and number of rows, both, will decrease in the

dataset.

6406532731283. ✖ There will be no change.

6406532731284. ✖ Insufficient information.

System Commands

Section Id :	64065356663
Section Number :	13
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	16
Number of Questions to be attempted :	16
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 :	640653118720
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Number : 291 Question Id : 640653815238 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