

Calculate the value of \hat{p}_2^0 .

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

0.5

Question Number : 173 **Question Id :** 640653587084 **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

Calculate the value of \hat{p}_2^1 .

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

0

MLP

Section Id :	64065339718
Section Number :	12
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	24

Number of Questions to be attempted :	24
Section Marks :	50
Display Number Panel :	Yes
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 :	64065384409
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Number : 174 Question Id : 640653587085 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 :

6406531958936. ✓ YES

6406531958937. ✗ NO

Sub-Section Number :	2
Sub-Section Id :	64065384410
Question Shuffling Allowed :	Yes

Is Section Default? :

null

Question Number : 175 Question Id : 640653587086 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 and its output:

Code:

```
from sklearn.datasets import load_iris
from sklearn.linear_model import LogisticRegression
```

```
X, y = load_iris(return_X_y=True)
clf = LogisticRegression(random_state=0).fit(X, y)
```

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

Output:

```
[1 1 1 1 1 1 1 1 1 1]
[2 1 1 1 1 1 1 2 1 1]
```

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

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

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

0.8

Question Number : 176 Question Id : 640653587095 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 might be the possible output of the following code:

```
from sklearn.metrics import precision_score
y_true = [1,1,0,1,0,0,1,0,1]
y_pred = [1,1,0,0,0,0,0,0,1]
print(precision_score(y_true,y_pred))
```

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

1.00

Question Number : 177 **Question Id :** 640653587097 **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.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 = [0,0,1,1,1,2,2,2,2,2,2]

X_test = [[2,200]]

knn = KNeighborsClassifier(n_neighbors= len(y_train),
                           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 : 178 Question Id : 640653587100 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?

```
import numpy as np
from sklearn.impute import KNNImputer
X = np.array([[5,6,3],[np.nan,1,5],[0,2,8],[4,4,2]])
knn = KNNImputer(n_neighbors=2,weights="uniform")
X_trf= knn.fit_transform(X)
print(X_trf[1][0])
```

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

2

Sub-Section Number : 3

Sub-Section Id : 64065384411

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 179 Question Id : 640653587087 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

The parameter C in a logistic regression is:

Options :

6406531958939. ✖ similar to the parameter α in a ridge regressor.

6406531958940. ✔ similar to $1 / \alpha$ where α is the parameter of a ridge regressor.

6406531958941. ✖ not controlling the regularization.

6406531958942. ✖ Weights associated with classes while fitting the model.

Question Number : 180 Question Id : 640653587089 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

Your task to design a model that can predict label of an article, in order to help an online news website. The labels could be “political”, “sports” and “international”.

Following is the label matrix for random 3 articles:

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

What type of classification problem is this?

Options :

6406531958947. ✖ Binary class, single label classification.

6406531958948. ✖ Binary class, multi label classification.

6406531958949. ✔ Multi class, multi label classification.

6406531958950. ✖ Multi class, single label classification.

Question Number : 181 Question Id : 640653587091 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

How does strong correlation between features given the labels impact the classification performance in Naive Bayes?

Options :

6406531958955. ✖ It has no impact because Naive Bayes assumes feature independence.

6406531958956. ✖ It improves the classification performance.

6406531958957. ✔ It degrades the classification performance.

6406531958958. ✖ It depends on the type of Naive Bayes variant used.

Question Number : 182 Question Id : 640653587092 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 might the Precision-Recall curve be more informative than the ROC curve?

Options :

6406531958959. ✔ When the dataset is imbalanced.

6406531958960. ✖ When the dataset has equal numbers of positive and negative instances.

6406531958961. ✖ When the classifier has high accuracy.

6406531958962. ✖ When the classifier produces balanced precision and recall values.

Question Number : 183 Question Id : 640653587093 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 ordinal data of the sizes of cups used in a coffee shop. Which of the following code will correctly transform the dataset as given in output array ?

```
dataset = [['Small'], ['Large'], ['Large'], ['Large'], ['Normal'], ['Small'], ['Large'], ['Normal']]
```

```
output = [[0], [2], [2], [2], [1], [0], [2], [1]]
```

Options :

```
from sklearn.preprocessing import OrdinalEncoder
oe = OrdinalEncoder()
print(oe.fit_transform(dataset))
```

6406531958963. ✖

```
from sklearn.preprocessing import OrdinalEncoder
oe = OrdinalEncoder(categories = [['Small', 'Normal', 'Large']])
print(oe.fit_transform(dataset))
```

6406531958964. ✔

```
from sklearn.preprocessing import OrdinalEncoder
oe = OrdinalEncoder()
print(oe.transform(dataset))
```

6406531958965. ✖

```
from sklearn.preprocessing import OrdinalEncoder
oe = OrdinalEncoder({"small"=0, "Normal"=1, "Large"=2})
print(oe.fit_transform(dataset))
```

6406531958966. ✖

Question Number : 184 Question Id : 640653587094 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 parameters learned by the KNN(KNeighborRegressor) model while training?

Options :

6406531958967. ✖ `coef_`

6406531958968. ✖ `n_neighbors`

6406531958969. ✖ `weight`

6406531958970. ✔ `None of these`

Question Number : 185 Question Id : 640653587096 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 a support vector machine model, let X_i be an input instance with label y_i . If X_i is a support vector what will be the output of this formula :

$$y_i * (X_i^T W + W_0)$$

W_0 and W are the estimated parameters from the model

Options :

6406531958972. ✖ > 1

6406531958973. ✖ < 1

6406531958974. ✔ $= 1$

6406531958975. ✖ Cannot be determined

Question Number : 186 Question Id : 640653587098 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 code which of the following option is true for that

```
from sklearn.neighbors import NearestNeighbors
neigh = NearestNeighbors(n_neighbors=4)
neigh.fit(X_train)
print(neigh.kneighbors(X_test[0:1]))
```

Assume X_train and X_test are of type numpy.ndarray.

Options :

6406531958977. ✖ It will print nearest neighbours from the test point.

6406531958978. ✖ It will print the distance of test point from all the training points.

6406531958979. ✔ It will print the distance and the index of the n_neighbors (in training set) for the test point.

6406531958980. ✖ It will throw an error.

Question Number : 187 Question Id : 640653587104 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 these may NOT help in handling overfitting in decision trees?

Options :

6406531958995. ✖ Increasing the value of min_samples_split

6406531958996. ✖ Increasing the value of the pruning parameter

6406531958997. ✖ Increasing the value of `min_samples_leaf`

6406531958998. ✔ Increasing the depth of the tree

Question Number : 188 Question Id : 640653587107 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

In a `BaggingClassifier` or `BaggingRegressor`, the parameter `base_estimator` can be:

Options :

6406531959007. ✔ Any predictor

6406531959008. ✖ only a decision tree predictor

6406531959009. ✖ only a linear model predictor

6406531959010. ✖ only a support vector predictor

Question Number : 189 Question Id : 640653587108 Question Type : MCQ Is Question

Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

You are working on a classification problem using

`sklearn.ensemble.RandomForestClassifier`. After training the model, you want to evaluate its performance on a test dataset. Which of the following method(s) can be used to obtain the predicted class probabilities for the test samples?

Options :

6406531959011. ✖ predict

6406531959012. ✔ predict_proba

6406531959013. ✖ decision_function

6406531959014. ✖ score

Sub-Section Number :	4
Sub-Section Id :	64065384412
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 190 Question Id : 640653587088 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

For which of the following cases, f1-score is the most suitable evaluation metric?

Options :

6406531958943. ✔ There are 10,000 images, each contains either a cat or a dog. Exactly 500 contain cats and others contain dogs. Your task is to train a binary classifier.

6406531958944. ✔ Train a binary classifier to detect if an MRI image contains carcinogenic cells or not. Number of true positives are 2%.

6406531958945. ✖ Predicting if a chest x-ray belongs to a male patient or a female patient. There are nearly equal number of samples of each category.

6406531958946. ✔ Based on a student's senior secondary marks and other features, predicting if he will fail a particular exam. The exam clearing rate is 98.23%.

Question Number : 191 Question Id : 640653587090 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 is correct?

Options :

6406531958951. ✖ `SGDClassifier(loss="percept")` is stochastic version of a perceptron model

6406531958952. ✔ `SGDClassifier(loss="log_loss")` is stochastic version of a logistic classifier model

6406531958953. ✖ `SGDClassifier(loss="log_loss")` is stochastic version of a SVM model

6406531958954. ✔ `SGDClassifier(loss="hinge")` is stochastic version of a SVM model

Question Number : 192 Question Id : 640653587099 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 option is true?

Options :

6406531958981. ✖ Distance between the datapoints varies as we change the `n_neighbors` parameter in `KNeighborsClassifier`.

6406531958982. ✖ `KNeighborsClassifier` model couldn't able to predict labels for the samples outside of the training dataset because it does not learn from dataset.

6406531958983. ✓ MinMaxScaler can impact the KNeighborsClassifier's accuracy score

6406531958984. ✖ KNeighborsClassifier can help in outlier detection

Question Number : 193 Question Id : 640653587101 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 value of C can overfit the SVM classifier model for the linearly inseparable data?

Options :

6406531958986. ✖ 0.0001

6406531958987. ✖ 1

6406531958988. ✓ 1000

6406531958989. ✖ Cannot be determined

Sub-Section Number :	5
Sub-Section Id :	64065384413
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 194 Question Id : 640653587102 Question Type : SA Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 1

Question Label : Short Answer Question

What will be the output of the following code:

```
from sklearn.feature_extraction.text import HashingVectorizer
corpus = ['You can have it all. Just not all at once.',
          'Train your mind to see the good in every situation.',
          'What we think, we become.',
          'If I got rid of my demons, I'd lose my angels.'],]
vectorizer = HashingVectorizer(n_features= 12,lowercase=True)
X = vectorizer.fit_transform(corpus)
print(X.shape[1])
```

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

12

Sub-Section Number : 6

Sub-Section Id : 64065384414

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 195 **Question Id :** 640653587103 **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. 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)
```

Options :

6406531958991. ✖ 12

6406531958992. ✖ 80

6406531958993. ✖ 60

6406531958994. ✔ 19

Sub-Section Number : 7

Sub-Section Id : 64065384415

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 196 Question Id : 640653587105 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 the most expected output for the code given below:

```
from sklearn.datasets import load_wine
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
X,y = load_wine(as_frame = True, return_X_y = True)

X_train,X_test,y_train,y_test = train_test_split(X,
                                                    y,
                                                    test_size = 0.10,
                                                    random_state = 12)

clf1 = DecisionTreeClassifier(ccp_alpha = 0.1,
                              random_state = 81)

clf2 = DecisionTreeClassifier(ccp_alpha = 0.25,
                              random_state = 81)

clf1.fit(X_train, y_train)
clf2.fit(X_train, y_train)

print(clf1.score(X_train, y_train))
print(clf2.score(X_train, y_train))
print(clf1.get_depth())
print(clf2.get_depth())
```

Options :

6406531958999. ✖ 0.9875

0.9125

2

3

6406531959000. ✖ 0.9125

0.9875

2

3

6406531959001. ✔ 0.9875

0.9125

3

2

6406531959002. ✖ 0.9125

0.9875

3

2

Sub-Section Number :	8
Sub-Section Id :	64065384416
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 197 Question Id : 640653587106 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_wine
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 = 8,
                            min_samples_leaf = 5,
                            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 be done at a node N?

Options :

- 6406531959003. ✖ Number of samples at node N = 5. If it is split, it will result in 3 samples in the left child and 2 samples in the right child.
- 6406531959004. ✔ Number of samples at node N = 10. If it is split, it will result in 5 samples in the left child and 5 samples in the right child.
- 6406531959005. ✔ Number of samples at node N = 15. If it is split, it will result in 9 samples in the left child and 6 samples in the right child.

6406531959006. ✖ Number of samples at node N = 8. If it is split, it will result in 5 samples in the left child and 3 samples in the right child.

BDM

Section Id :	64065339719
Section Number :	13
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	10
Number of Questions to be attempted :	10
Section Marks :	17
Display Number Panel :	Yes
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 :	64065384417
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Number : 198 Question Id : 640653587109 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 : BUSINESS DATA MANAGEMENT (COMPUTER BASED EXAM)"