

Indian Institute of Technology, Madras - Centre for Continuing Education

Notations :

- 1.Options shown in **green** color and with ✓ icon are correct.
- 2.Options shown in **red** color and with ✗ icon are incorrect.

Question Paper Name :

IIT M FOUNDATION DIPLOMA ENDTERM
QPB1 07 Aug 2022

Subject Name :

2022 Aug: IIT M FOUNDATION DIPLOMA
ENDTERM QPB1

Creation Date :

2022-08-03 15:35:26

Duration :

90

Total Marks :

991

Display Marks:

Yes

Share Answer Key With Delivery Engine :

Yes

Actual Answer Key :

Yes

Calculator :

Scientific

Magnifying Glass Required? :

No

Ruler Required? :

No

Eraser Required? :

No

Scratch Pad Required? :

No

Rough Sketch/Notepad Required? :

No

Protractor Required? :

No

Show Watermark on Console? :

Yes

Highlighter :

No

Auto Save on Console?

Yes

Change Font Color :

No

Find the joint distribution of U and V

Options :

$$g(u, v) = \begin{cases} \frac{1}{2}ve^{-(u-v)} & \text{if } 0 < v < 2, u > v \\ 0 & \text{otherwise} \end{cases}$$

6406531190679. ✓

$$g(u, v) = \begin{cases} \frac{1}{2}ve^{-(u+v)} & \text{if } 0 < v < 2, u > v \\ 0 & \text{otherwise} \end{cases}$$

6406531190680. ✗

$$g(u, v) = \begin{cases} \frac{1}{2}ve^{-(u-v)} & \text{if } 0 < v < 2, v > u \\ 0 & \text{otherwise} \end{cases}$$

6406531190681. ✗

$$g(u, v) = \begin{cases} \frac{1}{2}ve^{-(u-v)} & \text{if } 0 < x < 2, u > v \\ 0 & \text{otherwise} \end{cases}$$

6406531190682. ✗

MLP

Section Id :	64065322414
Section Number :	10
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	38
Number of Questions to be attempted :	38
Section Marks :	100
Display Number Panel :	Yes
Group All Questions :	No
Enable Mark as Answered Mark for Review and	Yes

Clear Response :

Maximum Instruction Time : 0
Sub-Section Number : 1
Sub-Section Id : 64065352349
Question Shuffling Allowed : No

Question Number : 236 Question Id : 640653359385 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"

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 :

6406531190683. ✓ Yes

6406531190684. ✗ No

Sub-Section Number : 2

Sub-Section Id : 64065352350

Question Shuffling Allowed : Yes

Question Number : 237 Question Id : 640653359386 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 similarity between loaders and fetchers of sklearn is -

Options :

6406531190685. ✗ Both loaders and fetchers generate controlled synthetic datasets

6406531190686. ✓ Both loaders and fetchers return a Bunch object, which is a dictionary with two keys

6406531190687. ✗ Both techniques load small standard datasets

6406531190688. ✗ None of these options are correct

Question Number : 238 Question Id : 640653359387 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 represent the main purpose of using the FeatureUnion?

Options :

6406531190689. ✗ Enables different transformations on various columns of data based on their types

6406531190690. ✓ Combines output from several transformer objects by creating a new transformer from them

6406531190691. ✗ To chain multiple estimators to execute a fixed sequence of steps in data preprocessing and modeling

6406531190692. ✗ None of these

Question Number : 239 Question Id : 640653359388 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 will be the output of the following block of code?

```
from sklearn.datasets import make_blobs
X, y = make_blobs(n_samples=10,
                    centers=3,
                    n_features=2,
                    random_state=0)
print(X.shape)
```

Options :

6406531190693. ✓ (10, 2)

6406531190694. ✗ (2, 10)

6406531190695. ✗ (10, 3)

6406531190696. ✗ (3, 10)

Question Number : 240 Question Id : 640653359390 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 load the wine dataset and print its shape?

Options :

```
import load_wine  
data = load_wine()  
6406531190701. ✗ print(data.data.shape())
```

```
from sklearn.datasets import load_wine  
data = load_wine()  
6406531190702. ✗ print(data.shape())
```

```
from sklearn.datasets import load_wine  
data = load_wine()  
6406531190703. ✗ print(data.data.shape())
```

```
from sklearn.datasets import load_wine  
data = load_wine()  
6406531190704. ✓ print(data.data.shape)
```

Question Number : 241 Question Id : 640653359391 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 block of code?

```
import numpy as np
from sklearn.preprocessing import MaxAbsScaler
X = [[ 1., -1.,  2.],
      [ 2.,  0.,  0.],
      [ 0.,  1., -1.]]
transformer = MaxAbsScaler().fit(X)
transformer.transform(X)
```

Options :

6406531190705. ✓
array([[0.5, -1. , 1.],
 [1. , 0. , 0.],
 [0. , 1. , -0.5]])

6406531190706. ✗
array([[[-0.5, -1. , -1.],
 [1. , 0. , 0.],
 [0. , 1. , -0.5]])

6406531190707. ✗
array([[[-0.5, -1. , -1.],
 [-1. , 0. , 0.],
 [0. , -1. , -0.5]])

6406531190708. ✗
array([[[-0.5, -0.5, -1.],
 [-1. , -1. , 0.],
 [0. , -0.5, -0.5]])

Question Number : 242 Question Id : 640653359392 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?

```
from sklearn.preprocessing import StandardScaler
data = [[0, 0], [0, 0], [1, 1], [1, 2]]
scaler = StandardScaler()
scaler.fit(data)
print(scaler.mean_)
```

Options :

6406531190709. ✘ [0.75 0.5]

6406531190710. ✘ [0.25 0.5]

6406531190711. ✓ [0.5 0.75]

6406531190712. ✘ [0.5 0.5]

Question Number : 243 Question Id : 640653359393 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 options will be the correct output for the following code snippet?

```
import numpy as np
x = np.array(
    [[5, 4],
     [3, 8],
     [2, 0],
     [9, 6]])
from sklearn.preprocessing import add_dummy_feature
x_new = add_dummy_feature(x)
print(x_new)
```

Options :

[[5, 4]

[3, 8]

[2, 0]

6406531190713. ✘ [9, 6]]

6406531190714. ✘

[[5. 1. 4.]
[3. 1. 8.]
[2. 1. 0.]
[9. 1. 6.]]

[[5. 4. 1.]
[3. 8. 1.]
[2. 0. 1.]
[9. 6. 1.]]

6406531190715. *

[[1. 5. 4.]
[1. 3. 8.]
[1. 2. 0.]
[1. 9. 6.]]

6406531190716. ✓

Question Number : 244 Question Id : 640653359394 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 options will be the correct output for the following code snippet?

```
from sklearn.preprocessing import MultiLabelBinarizer  
mlb = MultiLabelBinarizer()  
mlb.fit_transform([(1, 2), (3,)])
```

Options :

array([[1, 1, 0, 0],
 [0, 0, 1, 1]])

6406531190717. *

array([[1, 1, 0, 0],
 [0, 1, 1, 1]])

6406531190718. *

array([[1, 1, 0],
6406531190719. ✘ [1, 0, 1]])

array([[1, 1, 0],
6406531190720. ✓ [0, 0, 1]])

Question Number : 245 Question Id : 640653359395 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 correct explained variance score for the following code snippet applying PCA technique.

```
import numpy as np
from sklearn.decomposition import PCA
X = np.array([[1, 1], [2, 2]])
pca = PCA(n_components=2)
pca.fit(X)
print(pca.explained_variance_ratio_)
```

Options :

6406531190721. ✘ [1. 1.]

6406531190722. ✓ [1. 0.]

6406531190723. ✘ [0. 1.]

6406531190724. ✘ [0. 0.]

Question Number : 246 Question Id : 640653359398 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 output of the following block of code?

```
from sklearn.model_selection import LeaveOneOut
X = [1, 2, 3, 4]
loo = LeaveOneOut()
split = 1
for train, test in loo.split(X):
    print(f"split# {split}, {train} {test}")
    split += 1
```

Options :

6406531190730. ✓
split# 1, [1 2 3] [0]
split# 2, [0 2 3] [1]
split# 3, [0 1 3] [2]
split# 4, [0 1 2] [3]

6406531190731. ✗
split# 1, [1 2 3] [4]
split# 2, [4 2 3] [1]
split# 3, [4 1 3] [2]
split# 4, [4 1 2] [3]

6406531190732. ✗
split# 1, [1 2 3] [4]
split# 2, [4 2 3] [1]
split# 3, [4 1 3] [2]

6406531190733. ✗
split# 1, [1 2 3] [0]
split# 2, [0 2 3] [1]
split# 3, [0 1 3] [2]

Question Number : 247 Question Id : 640653359399 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 :

from sklearn.linear_model import SGDRegressor
6406531190734. ✓ linear_regressor = SGDRegressor(shuffle=True)

from sklearn.preprocessing import SGDRegressor
6406531190735. ✗ linear_regressor = SGDRegressor(shuffle=True)

from sklearn.SGDRegressor import linear_model
6406531190736. ✗ linear_regressor = SGDRegressor(learning_rate='constant', eta0=1e-2)

6406531190737. ✗ None of these

Question Number : 248 Question Id : 640653359401 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 represent the correct output of the following block of code?

```
import numpy as np
from sklearn.linear_model import LinearRegression
X = np.array([[0,0], [1, 1], [2, 2], [3, 3]])
y = np.dot(X, np.array([1, 1]))-1
reg = LinearRegression().fit(X, y)
reg.predict(np.array([[5, 6]]))
```

Options :

6406531190742. ✗ array([9.])

6406531190743. ✓ array([10.])

6406531190744. ✗ array([11.])

6406531190745. ✗ SyntaxError: unexpected EOF while parsing

Question Number : 249 Question Id : 640653359402 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 represent the set of transformed features of $[x_1, x_2]$ after execution of the following code?

```
from sklearn.preprocessing import PolynomialFeatures  
poly_transform = PolynomialFeatures(degree=2, interaction_only=True)
```

Options :

6406531190746. ✖ $[1, x_1, x_2, x_1^2, x_2^2]$

6406531190747. ✖ $[1, x_1^2, x_2^2]$

6406531190748. ✓ $[1, x_1, x_2, x_1x_2]$

6406531190749. ✖ $[1, x_1, x_2, x_1^2, x_2^2, x_1x_2]$

Question Number : 250 Question Id : 640653359403 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 possible output for the following block of code?

```
from sklearn.datasets import load_digits  
from sklearn.linear_model import Perceptron  
X, y = load_digits(return_X_y=True)  
clf = Perceptron(tol=1e-3, random_state=0)  
clf.fit(X, y)  
clf.score(X, y)
```

Options :

6406531190750. ✓ 0.94

6406531190751. ✘ 1.25

6406531190752. ✘ 1.94

6406531190753. ✘ -2.64

Question Number : 251 Question Id : 640653359405 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 operation does the following code implement?

```
from sklearn.linear_model import LogisticRegression  
logreg = LogisticRegression()  
logreg.fit(X,y)
```

where X and y are the training data.

Options :

6406531190758. ✘ It will perform linear regression on the given data.

6406531190759. ✘ It will generate synthetic regression data.

6406531190760. ✘ It will generate synthetic classification data.

6406531190761. ✓ It will perform classification on the given data.

Question Number : 252 Question Id : 640653359408 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 represent the recall value for the **class 2** in the confusion matrix shown in Figure 1 if $y_{true} = [0, 1, 2, 2, 2]$, $y_{pred} = [0, 0, 2, 2, 1]$?

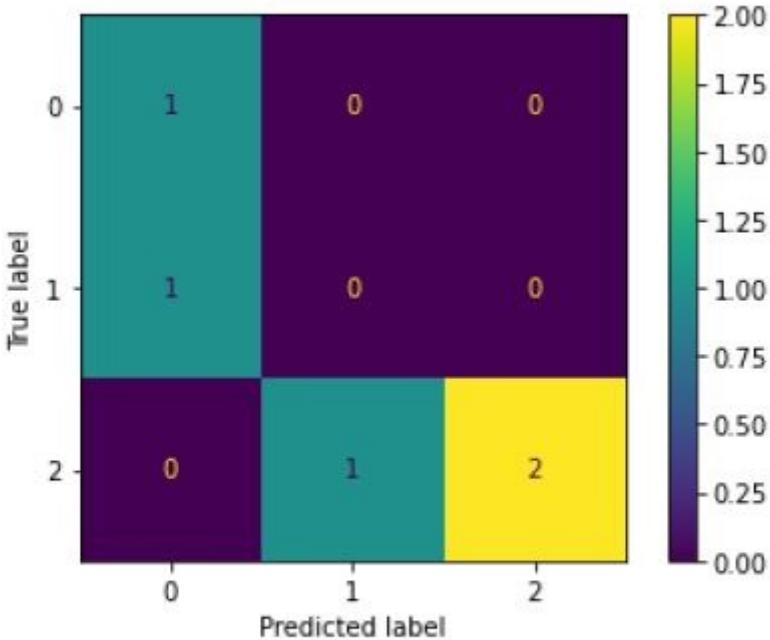


Figure 1

Options :

6406531190770. ✓ 0.67

6406531190771. ✗ 1.00

6406531190772. ✗ 0.50

6406531190773. ✗ 0.25

Question Number : 253 Question Id : 640653359420 Question Type : MCQ Is Question

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

Correct Marks : 2

Question Label : Multiple Choice Question

Consider the following code block and mark the correct output.

```
from sklearn.cluster import KMeans
import numpy as np
X = np.array([[2, 4], [4, 2], [8, 10], [10, 12], [20, 21], [17, 19]])
kmeans = KMeans(n_clusters=3, random_state=5).fit(X)
print(kmeans.labels_)
```

Options :

6406531190808. ✗ [0 0 0 1 1 1]

6406531190809. ✓ [2 2 0 0 1 1]

6406531190810. ✘ [0 1 2 0 1 2]

6406531190811. ✘ [0 1 0 1 0 1]

Question Number : 254 Question Id : 640653359422 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 X and y as the training dataset. What will be the output of the following code?

```
from sklearn.datasets import load_iris
X,y = load_iris(as_frame = True, return_X_y = True)
from sklearn.neural_network import MLPClassifier
rs = MLPClassifier(activation= 'logistic', random_state=12)
rs.fit(X, y)
print(rs.out_activation_)
```

Options :

6406531190813. ✘ logistic

6406531190814. ✘ relu

6406531190815. ✘ identity

6406531190816. ✓ softmax

Sub-Section Number : 3

Sub-Section Id : 64065352351

Question Shuffling Allowed : Yes

Question Number : 255 Question Id : 640653359389 Question Type : MSQ Is Question

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

Correct Marks : 2

Question Label : Multiple Select Question

Which of the following options represent the design philosophy of the sklearn API?

Options :

6406531190697. ✓ Nonproliferation of classes

6406531190698. ✗ Execution of all codes within 99.9 ms.

6406531190699. ✓ Sensible defaults

6406531190700. ✓ Direct accessibility of hyperparameters of all estimators

Question Number : 256 Question Id : 640653359400 Question Type : MSQ Is Question

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

Time : 0

Correct Marks : 2

Question Label : Multiple Select Question

Which of the following options represent(s) the correct method(s) to search for the best regularization parameter for ridge regression?

Options :

6406531190738. ✓ Search for the best regularization rate with built-in cross validation in RidgeCV estimator

6406531190739. ✗ Step 1: Instantiate object of Ridge estimator.

Step 2: Set parameter alpha to the maximum regularization rate.

6406531190740. ✓ Use cross validation with Ridge to search for best regularization.

6406531190741. ✓ Apply cross validation with SGDRegressor for searching the best regularization parameter.

Question Number : 257 Question Id : 640653359428 Question Type : MSQ Is Question

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

Time : 0

Correct Marks : 2

Question Label : Multiple Select Question

Consider the following block of code and mark all possible correct options.

```

from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier, VotingClassifier
clf1 = LogisticRegression()
clf2 = RandomForestClassifier()
vote=VotingClassifier(estimators=[('lr', clf1), ('rf', clf2)],
                      weights=None,
                      n_jobs= 1,
                      flatten_transform=True,
                      verbose=False,
                      voting='hard')
vote = vote.fit(X, y)
print(vote.predict(X))

```

Options :

6406531190830. ✓ Given model uses majority voting rule to predict class labels.
6406531190831. ✗ Given model predicts the class labels based on the argmax of the sums of the predicted probabilities.
6406531190832. ✓ Setting flatten_transform =True with voting='soft' will flatten the output shape of transform.
6406531190833. ✗ None of these are correct

Sub-Section Number : 4

Sub-Section Id : 64065352352

Question Shuffling Allowed : Yes

Question Number : 258 Question Id : 640653359396 Question Type : MCQ Is Question

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

Time : 0

Correct Marks : 3

Question Label : Multiple Choice Question

What will be the output of the following code block:

```

from sklearn.preprocessing import PolynomialFeatures
import numpy as np
X = np.arange(6).reshape(3, 2)
poly = PolynomialFeatures(degree=2)
print (poly.fit_transform(X))

```

Options :

```
[[ 1.  0.  1.  0.  0.  1.]  
[ 1.  2.  3.  4.  6.  9.]  
[ 1.  4.  5.  16. 20. 25.]]
```

6406531190725. ✓

```
[[ 1.  0.  1.  0.  0.  1.]  
[ 1.  2.  3.  4.  6.  9.]  
[ 1.  4.  5.  16. 20. 25.]  
[ 1.  6.  7.  36. 42. 49.]]
```

6406531190726. ✗

```
array([[1, 1, 0, 6, 8],  
       [1, 0, 1, 6, 10]])
```

6406531190727. ✗

```
[[ 1.  0.  2.  1.  1.  1.]  
[ 1.  1.  3.  6.  6.  9.]  
[ 1.  4.  5.  16. 26. 25.]]
```

6406531190728. ✗

Question Number : 259 Question Id : 640653359404 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 options represent the output of the following block of code?

```
from sklearn.metrics import confusion_matrix  
y_true = ["cat", "ant", "cat", "cat", "ant"]  
y_pred = ["ant", "ant", "cat", "cat", "ant"]  
confusion_matrix(y_true, y_pred, labels=["ant", "cat"])
```

Options :

```
array([[2, 0],  
       [1, 2]])
```

6406531190754. ✓

```
array([[2, 1],  
       [0, 2]])
```

6406531190755. ✗

array([[2, 2],
6406531190756. ✘ [1, 0]])

array([[2, 2],
6406531190757. ✘ [0, 2]])

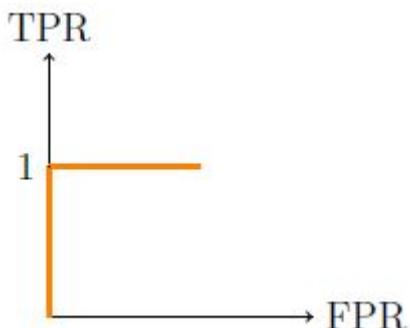
Question Number : 260 Question Id : 640653359406 Question Type : MCQ Is Question

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 statements will be true based on the following diagram?



Options :

6406531190762. ✘ The classifier will predict all positive classes as negatives and all negative classes as positives.

6406531190763. ✓ The classifier will perfectly distinguish between the positive and negative classes.

6406531190764. ✘ There is a good probability that the classifier can properly identify the classes.

6406531190765. ✘ We can not comment on the classifier performance based on the given figure

Question Number : 261 Question Id : 640653359407 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 options is the correct output of the following block of code?

```
import numpy as np
from sklearn.dummy import DummyRegressor
X = np.array([1.0, 1.5, 2.0])
y = np.array([1.0, 2.0, 3.0])
dummy_regr = DummyRegressor(strategy="mean")
dummy_regr.fit(X, y)
dummy_regr.predict(X)
```

Options :

6406531190766. ✘ array([2.25, 2.25, 2.25])

6406531190767. ✘ array([1., 1., 1.])

6406531190768. ✓ array([2., 2., 2.])

6406531190769. ✘ DummyRegressor()

Question Number : 262 Question Id : 640653359412 Question Type : MCQ Is Question

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

Time : 0

Correct Marks : 3

Question Label : Multiple Choice Question

Which of the following is likely to be the correct output of the code given below?

```
X = [[0], [5], [7], [10]]
y = [0, 0, 1, 1]
from sklearn.neighbors import KNeighborsClassifier
neigh = KNeighborsClassifier(n_neighbors=3)
neigh.fit(X, y)
A = neigh.kneighbors_graph(X)
A.toarray()
```

Options :

6406531190782. ✘ array([[1., 1., 1., 0.], [1., 1., 1., 0.], [0., 1., 0., 1.], [0., 1., 1., 1.]])

6406531190783. ✘ array([[1., 1., 1., 0.], [1., 1., 1., 0.], [0., 1., 0., 1.], [0., 1., 0., 1.]])

6406531190784. ✘ array([[1., 1., 0., 1.], [0., 1., 1., 0.], [0., 1., 1., 0.], [1., 0., 1., 1.]])

6406531190785. ✓ array([[1., 1., 1., 0.], [1., 1., 1., 0.], [0., 1., 1., 1.], [0., 1., 1., 1.]])

Question Number : 263 Question Id : 640653359416 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

For the following block of code, we get the output as 0.9875. How would the output change if we decrease max depth value ?:

```
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)

clf = DecisionTreeClassifier(max_depth = 6,
                             min_samples_split = 2,
                             min_samples_leaf=3,
                             random_state = 81)

clf.fit(X_train, y_train)
print(clf.score(X_train, y_train))
```

Options :

6406531190794. ✘ Output score will definitely increase.

6406531190795. ✘ Output score will definitely decrease.

6406531190796. ✓ Output score may decrease or remain the same.

6406531190797. ✘ Code will throw an error because max_depth can't be less than 6.

Question Number : 264 Question Id : 640653359417 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 block of code and mark the correct option.

```
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,
                             random_state = 5)
clf.fit(X_train, y_train)
print(clf.score(X_test, y_test))
```

Options :

6406531190798. ❌ The minimum number of samples required to split a leaf node is 5.

6406531190799. ❌ The minimum number of samples required to split an internal node is 6.

6406531190800. ❌ The minimum number of samples required to be at a leaf node is 4.

6406531190801. ✓ The minimum number of samples required to split an internal node is 5.

Question Number : 265 Question Id : 640653359419 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. What value should we fill at the blank place, so that given code block creates graph as shown below:

```

from sklearn.datasets import make_blobs
from sklearn.cluster import KMeans
X, y = make_blobs(n_samples=1000,
                   n_features=9,
                   centers=.....,
                   random_state=42)

wcss = [ ]
for i in range(1, 11):
    km = KMeans(n_clusters = i, random_state = 10)
    km.fit(X)
    wcss.append(km.inertia_)

plt.plot(range(1, 11), wcss)
plt.title('The Elbow Method', fontsize = 12)
plt.xlabel('No. of Clusters')
plt.ylabel('wcss')
plt.show()

```

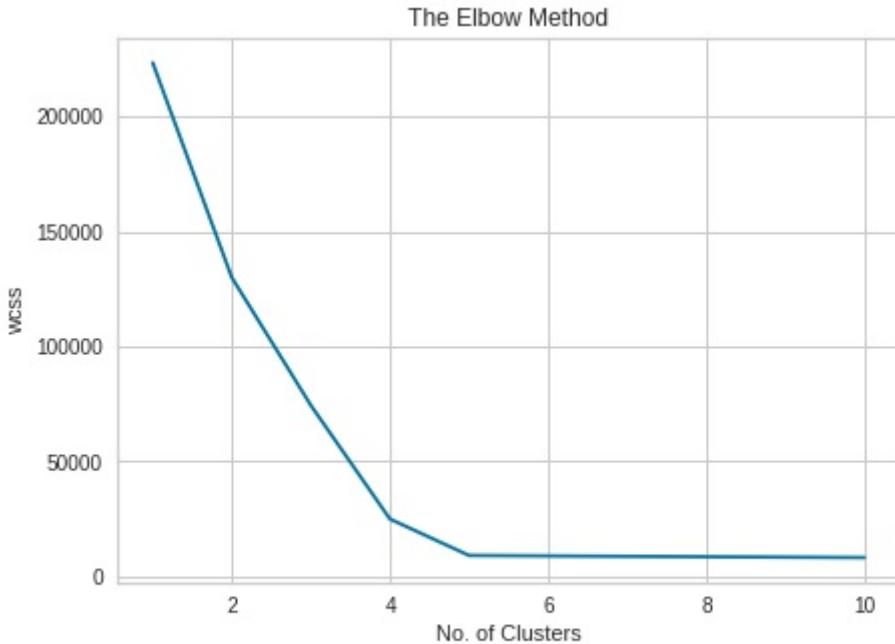


Figure 2: Elbow chart

Options :

6406531190803. ✘ 2

6406531190804. ✘ 4

6406531190805. ✓ 5

6406531190806. ✘ 6

6406531190807. ✘ 10

Sub-Section Number : 5

Sub-Section Id : 64065352353

Question Shuffling Allowed :

Yes

Question Number : 266 Question Id : 640653359397 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

Calculate the coefficient of determination (R^2) [up to 1 decimal point] based on the following block of code.

```
from sklearn.metrics import r2_score
y_true = [2, 5, 2, 7]
y_pred = [3, 9, 2, 8]
print(r2_score(y_true, y_pred))
```

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

-0.05 to 0.05

Question Number : 267 Question Id : 640653359418 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

In the given block of code, the Iris dataset having shape (150,4) has been loaded for model training. What do you think will be the output of the code given below?

```
from sklearn.datasets import load_iris
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
X,y = load_iris(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(max_features=2,
                             max_depth = 6,
                             min_samples_split = 2,
                             min_samples_leaf = 3,
                             random_state = 81 )

clf.fit(X_train, y_train)
print(clf.n_features_in_)
```

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

4

Question Number : 268 **Question Id :** 640653359421 **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

Code snippet written below plots a graph for some dataset X as shown below.

```
import scipy.cluster.hierarchy as shc  
plt.figure(figsize=(5, 7))  
plt.title("Dendograms")  
dend = shc.dendrogram(shc.linkage(X, method='ward'))
```

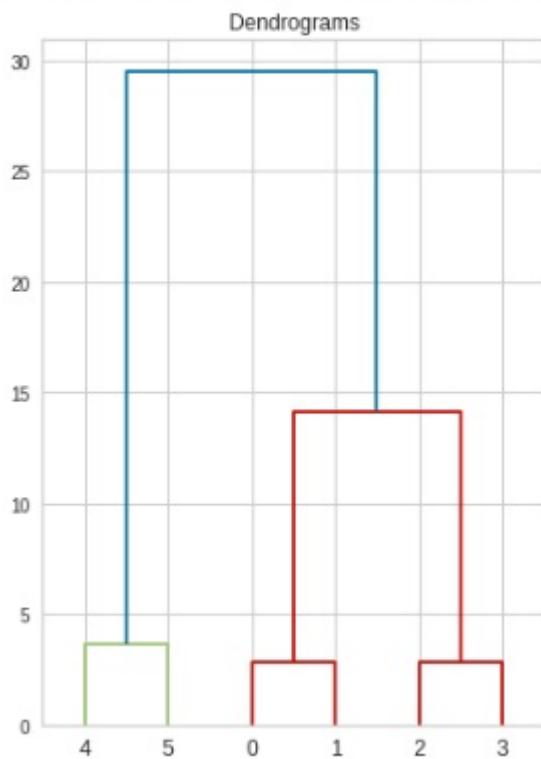


Figure 3: Elbow chart

In the given dendrogram which sample is most similar to the sample having label as 1 ?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

0

Sub-Section Number : 6

Sub-Section Id : 64065352354

Question Shuffling Allowed : No

Question Id : 640653359409 **Question Type :** COMPREHENSION **Sub Question Shuffling**

Allowed : No **Group Comprehension Questions :** No **Calculator :** None **Response Time :** N.A

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

Question Numbers : (269 to 270)

Question Label : Comprehension

Go through the code snippet given below and answer the subquestions.

```
X = [[0,1],[3,4], [4,5],[1,2],[2,3], [5,6]]  
y = [0, 1, 1,0,0,1]  
from sklearn.neighbors import RadiusNeighborsClassifier  
neigh = RadiusNeighborsClassifier(radius=1.0, P=1)  
neigh.fit(X, y)  
print(neigh.predict([[2.8,3.8]]))
```

Sub questions

Question Number : 269 Question Id : 640653359410 Question Type : MCQ Is Question

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

Correct Marks : 3

Question Label : Multiple Choice Question

Which of the following is likely to be the correct output of the given code ?

Options :

6406531190774. ✘ [2]

6406531190775. ✘ [0]

6406531190776. ✓ [1]

6406531190777. ✘ [3]

Question Number : 270 Question Id : 640653359411 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 distance metrics is used in the given code.

Options :

6406531190778. ✓ Euclidean_distance

6406531190779. ✘ Manhattan_distance

6406531190780. ✘ Hamming Distance

6406531190781. ✘ None of these

Question Id : 640653359423 Question Type : COMPREHENSION Sub Question Shuffling Allowed : No Group Comprehension Questions : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Question Numbers : (271 to 272)

Question Label : Comprehension

Consider the following code snippet:

```
from sklearn.datasets import fetch_california_housing
from sklearn.neural_network import MLPRegressor

X,y = fetch_california_housing(as_frame = True,
                                return_X_y = True)
rs = MLPRegressor(activation='logistic',
                  hidden_layer_sizes=(12,15,13,11,12,8),
                  random_state=12)
rs.fit(X, y)
```

Based on the above data, answer the given subquestions.

Sub questions

Question Number : 271 Question Id : 640653359424 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?

```
print(rs.n_layers_)
```

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

8

Question Number : 272 Question Id : 640653359425 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3

Question Label : Multiple Choice Question

What will be the output of the following code?

```
print(rs.coefs_[5].shape)
```

Options :

6406531190818. ✘ (5, 12)

6406531190819. ✘ (6, 11)

6406531190820. ✘ (8, 15)

6406531190821. ✓ (12, 8)

Sub-Section Number : 7

Sub-Section Id : 64065352355

Question Shuffling Allowed : No

Question Id : 640653359413 Question Type : COMPREHENSION Sub Question Shuffling Allowed : No Group Comprehension Questions : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Question Numbers : (273 to 274)

Question Label : Comprehension

Go through the code snippet given below and answer the subquestions.

```
import numpy as np
from sklearn.pipeline import make_pipeline
from sklearn.preprocessing import StandardScaler
X = np.array([[-8, -3], [3, 3], [5, 3], [-4, -3],])
y = np.array([-3, 3, 3, -3])
from sklearn.svm import SVC
clf = SVC(gamma='auto', kernel="linear")
clf.fit(X, y)
print(clf.predict([-4, -3]))
```

Sub questions

Question Number : 273 Question Id : 640653359414 Question Type : MCQ Is Question

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

Correct Marks : 3

Question Label : Multiple Choice Question

Which of the following is likely to be the correct output of the given code ?

Options :

6406531190786. ✘ [3]

6406531190787. ✘ [-1]

6406531190788. ✘ [1]

6406531190789. ✓ [-3]

Question Number : 274 Question Id : 640653359415 Question Type : MCQ Is Question

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

Correct Marks : 3

Question Label : Multiple Choice Question

Which of the following is likely to be the correct output:

```
print(clf.support_vectors_)
```

Options :

6406531190790. ✘ array([[-8., -3.], [3., 3.]])

6406531190791. ✓ array([[-4., -3.], [3., 3.]])

6406531190792. ✘ array([[-4., -3.], [5., 3.]])

6406531190793. ✘ array([[-8., -3.], [5., 3.]])

Sub-Section Number : 8

Sub-Section Id : 64065352356

Question Shuffling Allowed : Yes

Question Number : 275 Question Id : 640653359426 Question Type : MSQ Is Question

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

Correct Marks : 4

Question Label : Multiple Select Question

Which of the following is/are true about the following block of the code?

```
from sklearn.neural_network import MLPClassifier
X = [[0., 0.], [1., 1.], [2., 3.], [5., 4.]]
y = [0, 0, 1, 1]
rs = MLPClassifier(solver='adam',
                    alpha=0.0001,
                    batch_size="auto",
                    hidden_layer_sizes=(8,12),
                    random_state=12)
rs.fit(X,y)
rs.score(X,y)
```

Options :

6406531190822. ✓ Given dataset belongs to binary classification.

6406531190823. ✓ Number of neurons in the 1st hidden layer is 8.

6406531190824. ✘ Strength of the L1 regularization term is 0.0001

6406531190825. ✓ The activation function for the output layer in the given code is the '*logistic*' function.

Sub-Section Number : 9

Sub-Section Id : 64065352357

Question Shuffling Allowed : Yes

Question Number : 276 Question Id : 640653359427 Question Type : MSQ Is Question

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

Time : 0

Correct Marks : 5

Question Label : Multiple Select Question

Which of the following code blocks will throw an error ?

Options :

```
from sklearn.svm import SVC
from sklearn.ensemble import BaggingClassifier
from sklearn.model_selection import GridSearchCV
from sklearn.datasets import make_classification
X, y = make_classification(n_samples=100,
                           n_features=22,
                           n_informative=2,
                           n_redundant=0,
                           random_state=0,
                           shuffle=False)
param_grid = [{}'n_estimators':range(10,50,12)}]
clf = BaggingClassifier(base_estimator=SVC(),
                        n_estimators=10,
                        random_state=0)
est= GridSearchCV(clf,
                  param_grid,
                  cv = 5,
                  return_train_score=True)
est.fit(X, y)
6406531190826. ✘ est.best_score_
```

6406531190827. ✓

```
from sklearn.svm import SVC
from sklearn.ensemble import BaggingClassifier
from sklearn.model_selection import GridSearchCV
from sklearn.datasets import make_classification
X, y = make_classification(n_samples=100,
                           n_features=22,
                           n_informative=2,
                           n_redundant=0,
                           random_state=0,
                           shuffle=False)
param_grid = [{n_estimators:range(10,50,12)}]
clf = BaggingClassifier(base_estimator=SVC(),
                        n_estimators=10, random_state=0)
est= GridSearchCV(clf, param_grid, cv = 1,return_train_score=True)
est.fit(X, y)
est.cv_results_
```

```
from sklearn.svm import SVC
from sklearn.ensemble import BaggingClassifier
from sklearn.model_selection import GridSearchCV
from sklearn.datasets import make_classification
X, y = make_classification(n_samples=100,
                           n_features=22,
                           n_informative=2,
                           n_redundant=0,
                           random_state=0,
                           shuffle=False)
param_grid = [{n_estimators:range(10,50,12)}]
clf = BaggingClassifier(base_estimator=SVC(),
                        n_estimators=10,
                        random_state=0)
est= GridSearchCV(clf,
                  param_grid,
                  return_train_score=True)
est.fit(X, y)
est.score()
6406531190828. ✓
```

6406531190829. ✘ None of these

Section Id :	64065322415
Section Number :	11
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	33
Number of Questions to be attempted :	33
Section Marks :	100
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 :	64065352358
Question Shuffling Allowed :	No

Question Number : 277 Question Id : 640653359429 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 TECHNIQUES"

**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 :

6406531190834. ✓ Yes

6406531190835. ✗ No

Question Number : 278 Question Id : 640653359430 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

Note:

Do not write your answer as percentage. Always enter fractions as they are. e.g. if your answer is 0.245, enter the same, do not enter 24.5 %.

Options :

6406531190836. ✓ Useful Data has been mentioned above.

6406531190837. ✗ This data attachment is just for a reference & not for an evaluation.

Sub-Section Number : 2

Sub-Section Id : 64065352359

Question Shuffling Allowed : Yes

Question Number : 279 Question Id : 640653359441 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

During classification of linearly separable data-set using perceptron algorithm, as the value of learning rate α is increased,

Options :

6406531190868. ✗ The number of steps required for convergence increases.

6406531190869. ✗ The number of steps required for convergence decreases.

6406531190870. ✓ The number of steps required for convergence is independent of the learning rate.

6406531190871. ✗ The perceptron algorithm may not converge.

Question Number : 280 Question Id : 640653359445 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

A knn algorithm with $k = 10$ gives low training error and high validation error. What value of k should we choose to get the better performance of the algorithm?

Options :

6406531190879. ❌ Less than 10

6406531190880. ✓ Greater than 10

Question Number : 281 Question Id : 640653359451 Question Type : MCQ Is Question

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

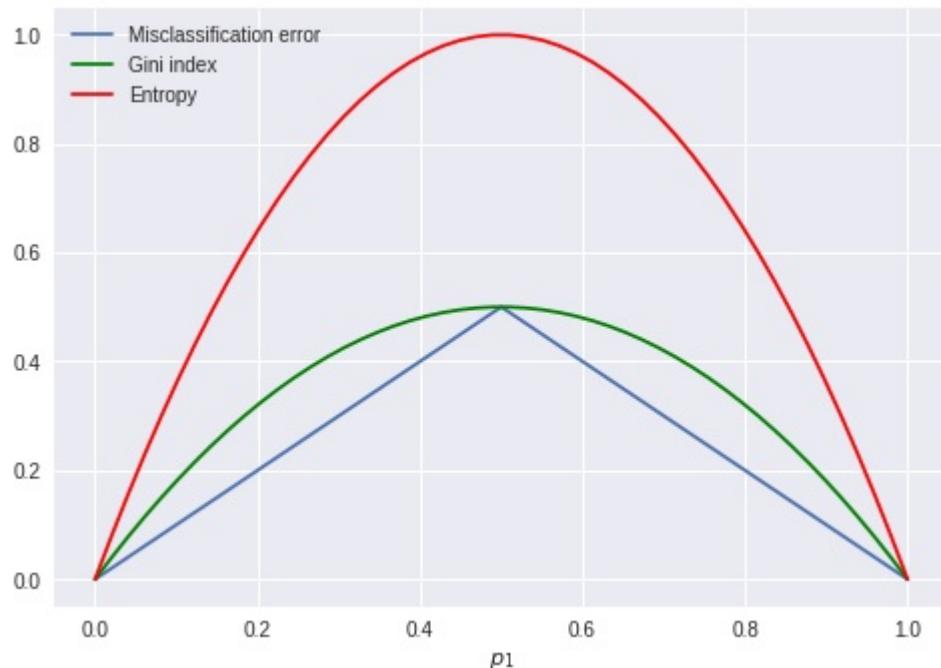
Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

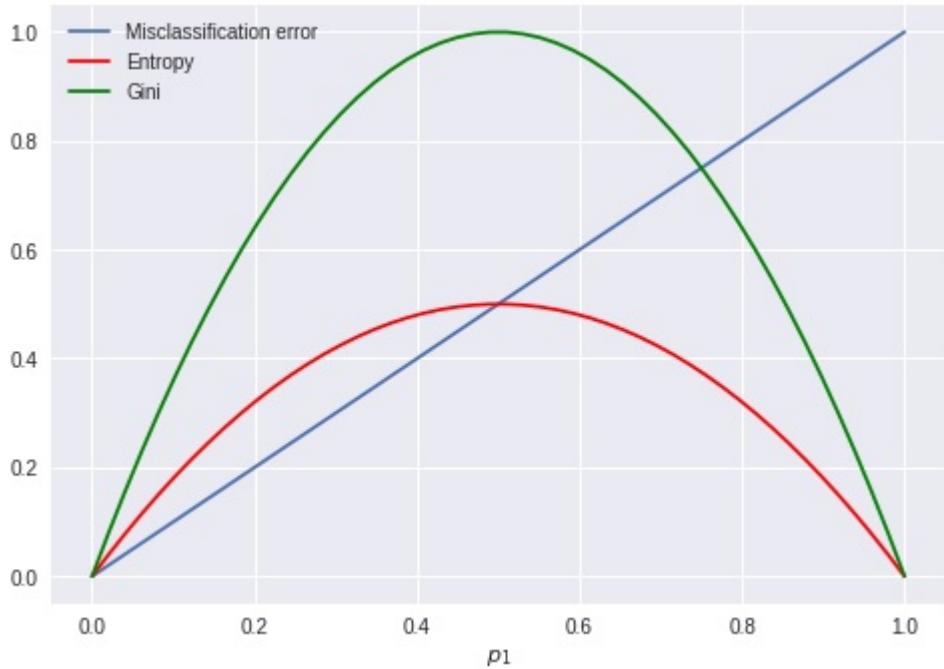
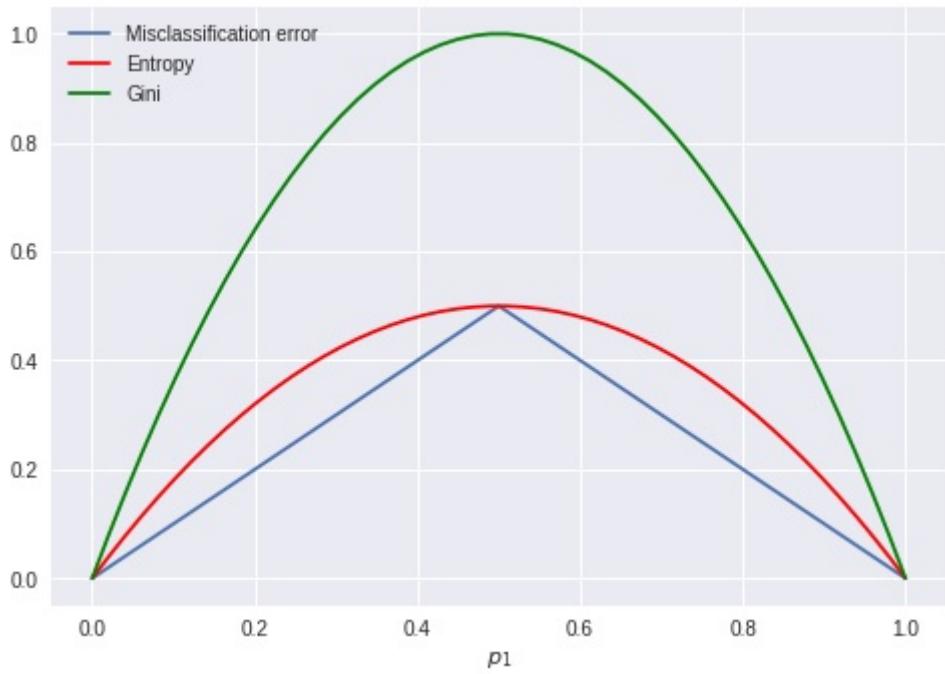
Consider a binary classification problem. Let p_1 denote the proportion of class 0 examples in a particular node. Which of the following graphs shows correct curves for the Gini-index, Entropy and misclassification error of that node?

Options :



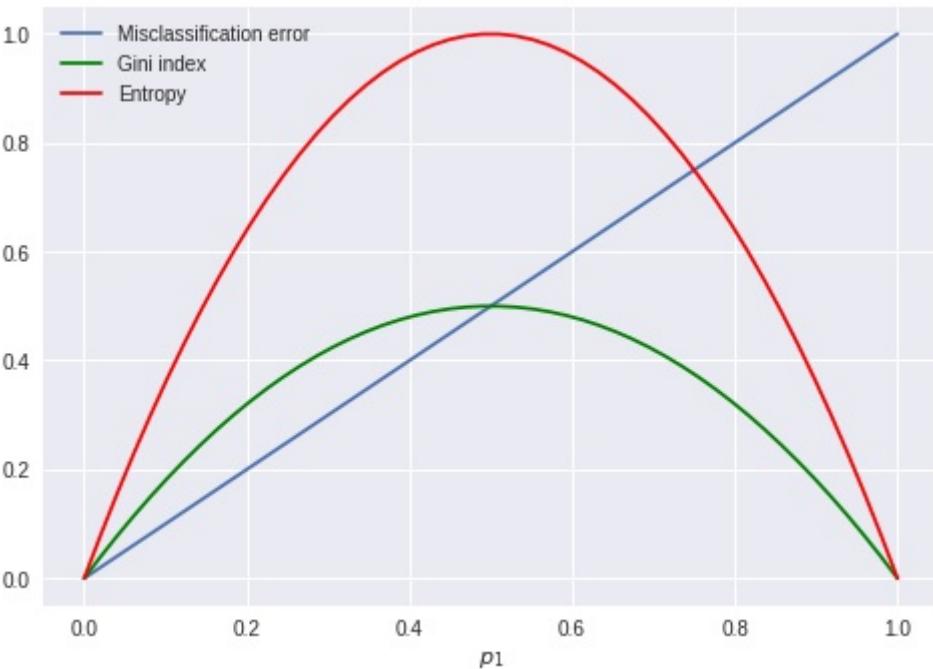
6406531190885. ✓

6406531190886. ❌



6406531190887. ✎

6406531190888. ✎



Question Number : 282 Question Id : 640653359462 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 following is the activation vector output by some hidden layer in a neural network when some input vector is given to it:

$$\begin{bmatrix} -0.2 \\ 0.8 \\ -0.9 \\ 0.1 \\ 0 \\ -0.3 \end{bmatrix}$$

Which of the following could be the activation function used in this layer?

Options :

6406531190911. ✘ Softmax

6406531190912. ✘ Sigmoid

6406531190913. ✘ ReLU

6406531190914. ✓ Tanh

Sub-Section Number :

3

Sub-Section Id :

64065352360

Question Shuffling Allowed :

Yes

Question Number : 283 Question Id : 640653359431 Question Type : MSQ Is Question

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

Correct Marks : 2

Question Label : Multiple Select Question

Which of the following are multi label problems? For each option, assume suitable features are available.

Options :

- 6406531190838. ❌ Predicted number of runs the Indian Cricket team will score in their next ODI.
- 6406531190839. ✓ Amount of money that Amitabh Bachchan's next three movies each will make.
- 6406531190840. ✓ Predicting number of goals Ronaldo will score in next 10 matches each.
- 6406531190841. ❌ Predicting blood group of a person.
- 6406531190842. ✓ Predicting blood sugar level of a person for next 15 days each.

Question Number : 284 Question Id : 640653359455 Question Type : MSQ Is Question

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

Correct Marks : 2

Question Label : Multiple Select Question

In a random forest model let $p < m$ be the number of randomly selected features that are used to identify the best split at any node of a tree. Which of the following are true? (m is the original number of features)

Options :

- 6406531190891. ❌ Increasing p reduces the correlation between any two trees in the forest.
- 6406531190892. ✓ Decreasing p reduces the correlation between any two trees in the forest.
- 6406531190893. ✓ Increasing p increases the performance of individual trees in the forest.
- 6406531190894. ❌ Decreasing p increases the performance of individual trees in the forest.

Sub-Section Number :

4

Sub-Section Id :

64065352361

Question Shuffling Allowed :

Yes

Question Number : 285 Question Id : 640653359434 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 feature matrix $X = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 2 & 3 \end{bmatrix}$, label vector $y = \begin{bmatrix} 4 \\ 5 \\ 2 \end{bmatrix}$ and the weight

vector is $w = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$. Add a dummy feature to X .

What will be the value of the loss function if a regression model without regularization is fitted?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

20.5 to 21.5

Question Number : 286 Question Id : 640653359465 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 a test-dataset of 100 points for a binary classification problem, where 60 belong to the positive class (true label) and the rest belong to the negative class (true label). The following is a table for some classifier that has been prepared by an ML engineer:

	Predicted label (+)	Predicted label (-)
True label (+)	40	20
True label (-)	10	30

If this is a valid confusion matrix (just by looking at the numbers), enter the classifier's precision as the answer. If this is not a valid confusion matrix, enter 0 as the answer. Your answer should be in the interval [0, 1], endpoints inclusive. Enter your answer correct to two decimal places.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

0.79 to 0.81

Sub-Section Number : 5

Sub-Section Id : 64065352362

Question Shuffling Allowed : Yes

Question Number : 287 **Question Id :** 640653359438 **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 a modified loss function for linear regression that is of the following form for a training dataset that has n points:

$$L(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^n r_i (\mathbf{w}^T \mathbf{x}_i - y_i)^2$$

Here, r_i is some constant in $[0, 1]$ associated with each data-point in the training dataset. The dummy feature and the corresponding weight are already included in the vectors \mathbf{x} and \mathbf{w} respectively. What is the expression of the gradient of $L(\mathbf{w})$ with respect to \mathbf{w} ?

Options :

6406531190862. ✓ $\sum_{i=1}^n r_i (\mathbf{w}^T \mathbf{x}_i - y_i) \mathbf{x}_i$

6406531190863. ❌ $\sum_{i=1}^n r_i (\mathbf{w}^T \mathbf{x}_i - y_i)$

6406531190864. ❌ $\sum_{i=1}^n (\mathbf{w}^T \mathbf{x}_i - y_i) \mathbf{x}_i$

6406531190865. ❌ $\sum_{i=1}^n r_i (\mathbf{w}^T \mathbf{x}_i - y_i)^2 \mathbf{x}_i$

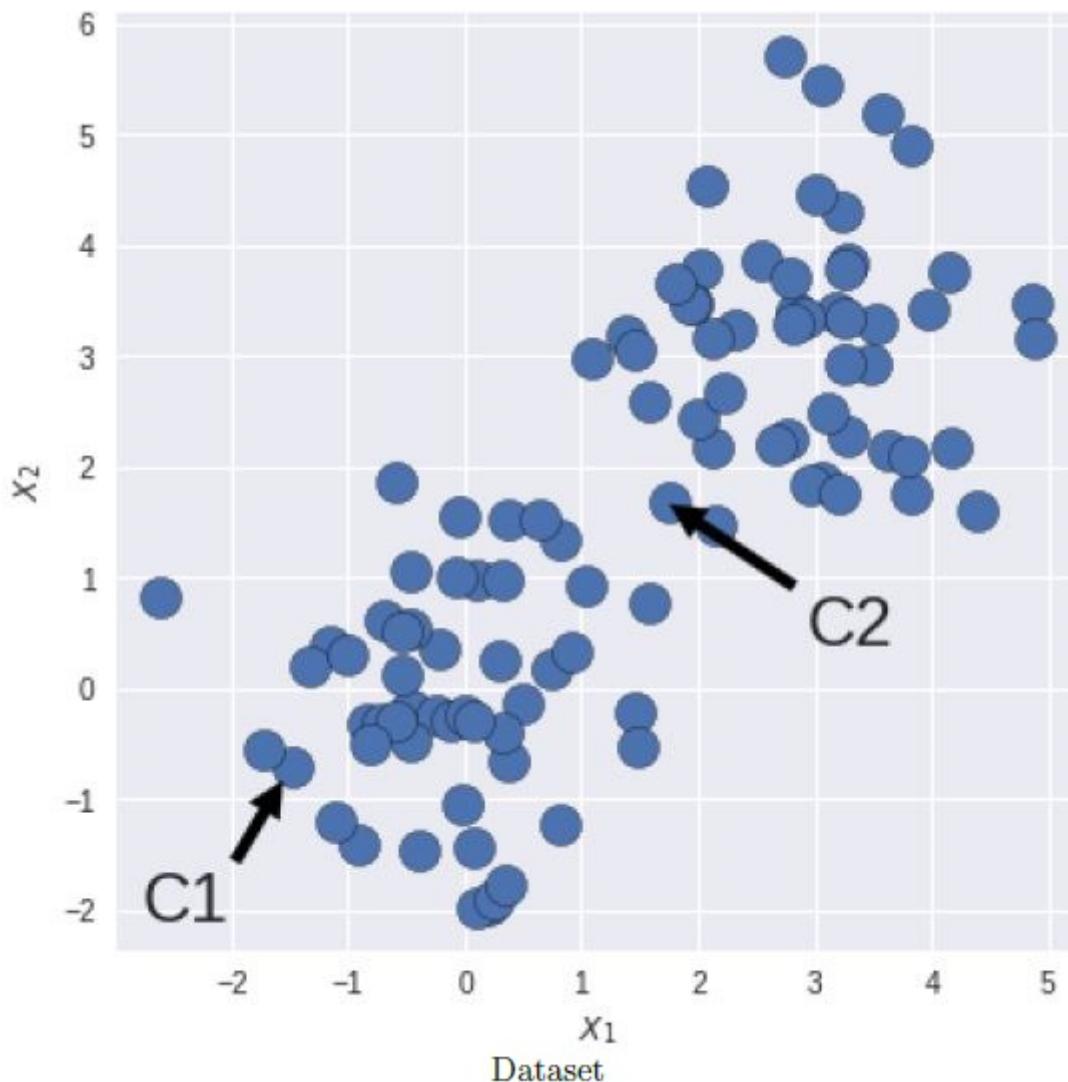
Question Number : 288 Question Id : 640653359456 Question Type : MCQ Is Question

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 unlabeled data with two features X_1 and X_2 as shown in the figure.



C_1 and C_2 are the coordinates of centroids obtained after certain iterations of the K -means algorithm with $K = 2$. Which of the following options is correct about the silhouette score S ? Note: Euclidean distance is used to calculate the distances.

Options :

- 6406531190895. ❌ S will be positive and close to 0.
- 6406531190896. ❌ S will be positive and close to 1.
- 6406531190897. ❌ S will be exact -1.
- 6406531190898. ✓ S will be negative but need not be exact -1.

Question Number : 289 Question Id : 640653359459 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 a neural network for a multi-class, image classification problem. When the network is trained on the images as they are, it does a good job on the test data. Call the dataset (train + test) for this setup D_1 and the network N_1 . Assume that we now turn all images upside down, in both the training and test dataset. Now, the network with the same architecture is trained from scratch on this modified dataset. Call the dataset (train + test) for this setup D_2 and network N_2 . Select the most appropriate option.

Options :

6406531190903. ❌ The network N_2 will not be able to learn anything from D_2 . Its test accuracy on D_2 will be very low.

6406531190904. ✓ The network N_2 will be able to learn useful patterns from D_2 . In fact, the performance of network N_2 on D_2 will be similar to N_1 on D_1 .

6406531190905. ❌ The network N_2 will be able to learn somewhat useful patterns from D_2 . But the performance of N_1 on D_1 will be much better than N_2 on D_2 .

Question Number : 290 Question Id : 640653359464 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

Two classifiers are trained on a dataset for a binary classification problem. They are then tested on the same dataset. The F_1 scores of both classifiers are the same. (p_1, r_1) and (p_2, r_2) are the precision-recall scores for the two classifiers.

Consider the following statements:

Statement-1

If $p_1 > p_2$, then $r_1 < r_2$

Statement-2

If $p_1 = p_2$, then $r_1 = r_2$

Select the most appropriate option.

Options :

6406531190916. ❌ Statement-1 is correct, statement-2 is incorrect

6406531190917. ❌ Statement-1 is incorrect, statement-2 is correct

6406531190918. ✓ Both statements 1 & 2 are correct

6406531190919. ❌ Both statements 1 & 2 are incorrect

Sub-Section Number : 6

Sub-Section Id : 64065352363

Question Shuffling Allowed : Yes

Question Number : 291 Question Id : 640653359432 Question Type : MSQ Is Question

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

Correct Marks : 3

Question Label : Multiple Select Question

Which of the following code blocks will produce the same value for the variable 'Ans'? Assume the 'numpy' library is imported as 'np'.

Options :

A = np.arange(6)
B = 5 - A

6406531190843. ✓ Ans = B + A - 3

A = np.ones((3,2)).astype('int')
B = A.reshape(-1)

6406531190844. ✓ Ans = B * 2

A = np.random.rand(2,3)
B = A.astype('int').reshape(-1)

6406531190845. ✓ Ans = B + 2

A = np.linspace(10, 60, 6)
B = (A.astype('int') - 10)//10

6406531190846. ❌ Ans = B + 2

6406531190847. ❌ All the code blocks produce unique results.

6406531190848. ✘ All the code blocks produce same results.

Question Number : 292 Question Id : 640653359433 Question Type : MSQ Is Question

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

Correct Marks : 3

Question Label : Multiple Select Question

Rajesh trained a model on house price prediction problem. After training he came up with following model:

$$y = 4000 - 20 \cdot x_1 + 1500 \cdot x_2 - 15 \cdot x_3 + 3 \cdot x_4$$

where,

y is predicted house price,

x_1 is age of the house in years,

x_2 is the area of house/apartment in square feet,

x_3 is the number of floors the house/building is,

x_4 represents if the house is facing the sea or not.

According to Rajesh's model, which of the following statements are correct?

Options :

6406531190849. ✘ The most important factor in determining the price of the house is if it is sea facing or not.

6406531190850. ✘ No. of floors in the building is the most important feature.

6406531190851. ✓ Carpet area is the most important feature.

6406531190852. ✘ Age is the most important feature.

6406531190853. ✓ Age is the second most important feature.

6406531190854. ✘ None of these.

Question Number : 293 Question Id : 640653359436 Question Type : MSQ Is Question

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

Time : 0

Correct Marks : 3

Question Label : Multiple Select Question

Consider that you have 100 non-linear data points randomly generated from a $\sin(2\pi x)$ function $0 \leq x \leq 2$. Choose the correct statement(s) from the following.

Options :

6406531190857. ✓ The polynomial regression model of degree 5 adds a smooth fitting to this data.

6406531190858. ✓ The polynomial regression model of degree 99 overfits this data.

6406531190859. ✗ The polynomial regression model of degree 1 adds a smooth fitting to this data.

6406531190860. ✓ The polynomial regression model of degree 2 underfits this data.

Sub-Section Number : 7

Sub-Section Id : 64065352364

Question Shuffling Allowed : Yes

Question Number : 294 Question Id : 640653359435 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?

```
import numpy as np

def aFunction(A, d):
    temp = np.ones(1)

    for i in range(1, d + 1):
        temp = np.concatenate((temp, A ** i))

    return temp
```

```
A = np.arange(3)
print(np.sum(aFunction(A, 3)))
```

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

18.0

Question Number : 295 Question Id : 640653359437 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 feature matrix $\mathbf{X} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 4 \end{bmatrix}$ and corresponding label vector

$\mathbf{y} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$. Let regularization rate, $\lambda = 0.1$. Compute ridge regression loss using weight

$$\mathbf{w} = \begin{bmatrix} 1 \\ 0.01 \\ -0.2 \\ 0.3 \end{bmatrix}.$$

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

2.48 to 2.58

Question Number : 296 Question Id : 640653359439 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

You are given a linearly separable dataset with feature matrix \mathbf{X} , for which a perceptron has been

trained until it converges (perfectly separates the data). The weight vector corresponding to it is \mathbf{w} . This dataset has 70 points from the positive class and 30 from the negative class. What is the output of the following snippet of code?

Notes

- The labels for a perceptron lie in $\{-1, 1\}$.
- The NumPy arrays \mathbf{X} and \mathbf{w} are compatible for matrix multiplication, the dummy feature and the corresponding weight are already built into the arrays.
- None of the 100 points lie on the decision boundary.

```
import numpy as np  
y_hat = np.where(X @ w > 0, 1, -1)  
print(np.sum(y_hat))
```

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

40

Question Number : 297 **Question Id :** 640653359444 **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

A Gaussian Naive Bayes model is trained for a multi-class classification problem that has 10 features and 3 classes. Find the total number of parameters that have to be estimated for this model. Consider each parameter to be a scalar value. In other words, if we decide to store all the parameters in a Python list, with each element of the list being a float value corresponding to a single parameter, what is the size of this list? Ignore the priors in the calculation. Only focus on the parameters of the class conditional densities.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

60

Question Number : 298 **Question Id :** 640653359458 **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 network architecture:

Layer	Number of Neurons
Input	10
Hidden layer-1	20
Hidden layer-2	30
Output layer	3

How many parameters (weights + biases) does this network have?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

943

Question Number : 299 **Question Id :** 640653359463 **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

The MNIST digit classification problem has 10 classes. The training dataset has n data points, with an equal number of points from each of the 10 classes. Consider a dummy classifier that does prediction as follows: for each input data-point, it picks one of the 10 classes at random (uniformly) and outputs that as its prediction. What is the accuracy of the model on the training dataset as n

becomes very large? Your answer should be between 0 and 1. Enter your answer correct to two decimal places.

Hint: Think about it in the probabilistic sense.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

0.09 to 0.11

Sub-Section Number : 8

Sub-Section Id : 64065352365

Question Shuffling Allowed : Yes

Question Number : 300 **Question Id :** 640653359443 **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 a logistic regression model for a binary classification problem with

two features x_1 and x_2 . The feature vector is $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and labels lie in $\{0, 1\}$. The threshold for inference is 0.5. The dummy feature and the weight corresponding to it can be ignored for this problem. Let x_1 be the horizontal axis and x_2 be the vertical axis. You are given two feature vectors:

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ \sqrt{3} \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} -1 \\ \sqrt{3} \end{bmatrix}$$

The weight vector makes an angle of θ with the positive x_1 axis (horizontal). Each θ corresponds to a different classifier. For what range of values of θ are both \mathbf{x}_1 and \mathbf{x}_2 predicted to belong to class-1?

Hints:

- To draw the weight vector $\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$, plot the point (w_1, w_2) and draw an arrow starting at the origin to this point.
- $\tan(60^\circ) = \sqrt{3}$

Options :

6406531190873. ✓ $30^\circ < \theta < 150^\circ$

6406531190874. ✗ $0^\circ < \theta < 60^\circ$

6406531190875. ✗ $60^\circ < \theta < 180^\circ$

6406531190876. ✗ $0^\circ < \theta < 180^\circ$

6406531190877. ✗ $0^\circ < \theta < 360^\circ$

Question Number : 301 Question Id : 640653359457 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 data points

(1,1)
(1,2)
(5,4)
(2.4,3)
(2.6, 3)

We perform k-means clustering on the above data, with $k = 2$, using Manhattan distance as the distance measure. At t th iteration, we have (1,1) as the centroid for cluster-1 and (5,4) as the centroid for cluster-2. After performing $(t + 1)$ th iteration the point (2.4,3) will belong to

Options :

6406531190899. ✘ cluster-1

6406531190900. ✓ cluster-2

6406531190901. ✘ can not be determined

Sub-Section Number : 9

Sub-Section Id : 64065352366

Question Shuffling Allowed : Yes

Question Number : 302 Question Id : 640653359460 Question Type : MSQ Is Question

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

Correct Marks : 4

Question Label : Multiple Select Question

If W_{ij} is the weight of the edge from neuron i in layer $l - 1$ to neuron j in layer l , which of the following statements about the matrix \mathbf{W} are true? Neurons in a layer are processed (indexed) from top to bottom. So, the first neuron in a layer is the top-most neuron in that layer.

Options :

6406531190906. ✓ The first row of the matrix corresponds to all outgoing connections from the first neuron in layer $l - 1$.

6406531190907. ✘ The first row of the matrix corresponds to all incoming connections to the first neuron in layer l .

6406531190908. ✓ The last column of the matrix corresponds to all incoming connections to the last neuron in layer l .

6406531190909. ✘ The last column of the matrix corresponds to all the outgoing connections

from the last neuron in layer $l - 1$.

Sub-Section Number : 10

Sub-Section Id : 64065352367

Question Shuffling Allowed : Yes

Question Number : 303 **Question Id :** 640653359440 **Question Type :** SA **Calculator :** None

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

Correct Marks : 4

Question Label : Short Answer Question

A perceptron model is trained on the following binary classification data-set.

X_1	X_2	Label (y)
0	0	-1
0	1	-1
1	0	-1
1	1	1

At one of the iterations, weights are: $w_0 = -0.5$, $w_1 = 0.9$ and $w_2 = 0.9$. Assume that w'_0 , w'_1 and w'_2 are the updated weights after one epoch (one epoch is completed on going through all the data points). Find the value of $w'_0 + w'_1 + w'_2$. Consider the learning rate to be one. Don't change the order of samples while updating the weight vector. Take the original order of samples.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

3.3

Question Number : 304 **Question Id :** 640653359442 **Question Type :** SA **Calculator :** None

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

Correct Marks : 4

Question Label : Short Answer Question

Inference using logistic regression happens as follows. T is called the threshold and is some real number in the interval $(0, 1)$. \hat{y} stands for the predicted label.

$$\hat{y} = \begin{cases} 1, & P(y = 1 | \mathbf{x}) \geq T \\ 0, & \text{otherwise} \end{cases}$$

Given this setup, the equation of the decision boundary is given below:

$$\mathbf{w}^T \mathbf{x} - u = 0$$

\mathbf{w} has the same dimensions as \mathbf{x} . The dummy feature is included in \mathbf{x} and the corresponding weight is included in \mathbf{w} . If $T = \frac{e}{1+e}$, what is the value of the unknown quantity u ? Here, e is Euler's number that is associated with the natural logarithm and is the same one found on your calculator. Enter the closest integer as your answer.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

1

Question Number : 305 **Question Id :** 640653359446 **Question Type :** SA **Calculator :** None

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

Correct Marks : 4

Question Label : Short Answer Question

An SVM has been trained for a 2D problem. The feature vector is $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

It has the following weight vector and bias:

$$\mathbf{w} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, b = -1$$

Recall that the labels in the SVM setup are +1 and -1. Consider a unit square whose vertices are at:

$$(0,0), (1,0), (0,1), (1,1)$$

The horizontal axis corresponds to x_1 and the vertical axis corresponds to x_2 . A point is picked at random (uniformly) from the region bounded by the square. What is the probability that this point is predicted as belonging to class +1 by the SVM? Enter your answer correct to two decimal places.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

0.74 to 0.76

Question Number : 306 **Question Id :** 640653359447 **Question Type :** SA **Calculator :** None

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

Correct Marks : 4

Question Label : Short Answer Question

Find the hinge loss for this soft-margin SVM classifier on the dataset that is given in the table. The weight vector and bias are as follows:

$$\mathbf{w} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad b = 1$$

The coefficient C can be assumed to be 1.

x_1	x_2	y
1	4	-1
-1	2	-1
0	0	-1
1	2	-1
1	3	1
1	0	1
2	1	1
2	3	1

Note that you just need to report the hinge loss. Do not compute margin loss which involves only the term w . Also, note that the hinge loss does not have a factor of 0.5 before it. Enter the closest integer as your answer.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

6

Question Number : 307 **Question Id :** 640653359461 **Question Type :** SA **Calculator :** None

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

Correct Marks : 4

Question Label : Short Answer Question

Consider two networks N_1 and N_2 for a binary classification task. N_1 has two neurons at the output layer and uses the softmax activation function. N_2 has one neuron at the output layer and uses sigmoid activation function. We don't need the information about the hidden layers for this problem.

For some test data-point x , the pre-activations at the output layer for N_1 is given below. The first neuron corresponds to class-0 and the second corresponds to class-1:

$$\mathbf{z} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

It turns out that for this data-point, both networks predict the same probability of this point belonging to class-1. That is $P(y = 1 | x)$ is the same for both networks. If this is the case, what should be the pre-activation value at the output layer of N_2 , call it z , corresponding to this data-point? Enter the closest integer as your answer.

Notes

- The activation of N_2 at the output-layer is interpreted as $P(y = 1 | x)$.
- The \mathbf{z} for N_1 is a vector and the z for N_2 is a scalar.
- Be careful about the distinction between pre-activation and activation.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

3

Sub-Section Number : 11

Sub-Section Id : 64065352368

Question Shuffling Allowed : No

Question Id : 640653359448 **Question Type :** COMPREHENSION **Sub Question Shuffling Allowed :** No **Group Comprehension Questions :** No **Calculator :** None **Response Time :** N.A

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

Question Numbers : (308 to 309)

Question Label : Comprehension

Consider the following training data-set with three features:

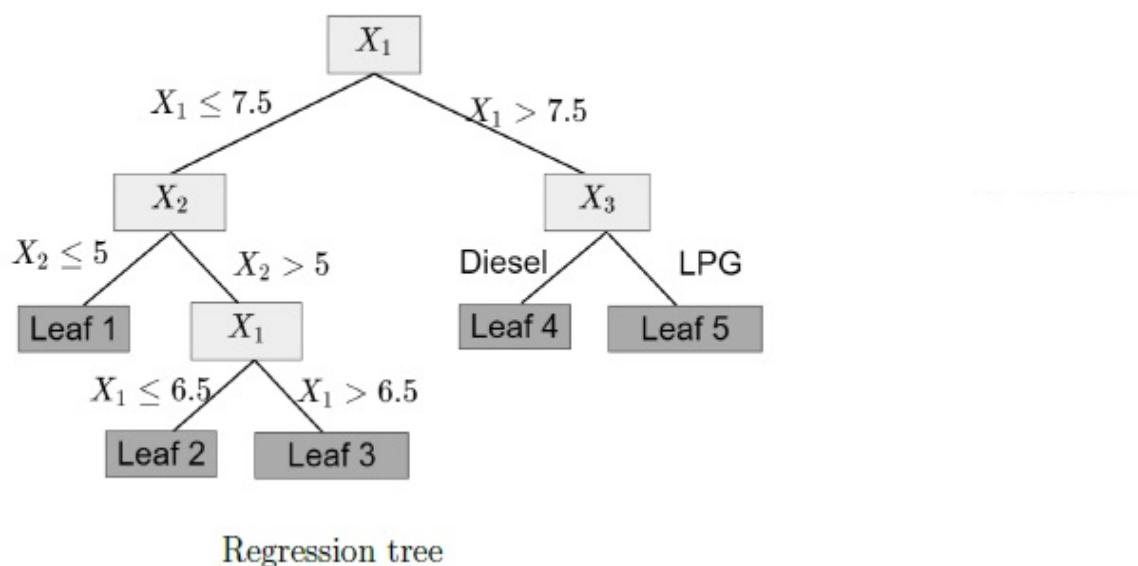
X_1 = Present price of the car (values are given in lac)

X_2 = Age (How many years the car has been driven)

X_3 = Fuel-type (diesel or LPG)

X_1	X_2	X_3	Selling price (y)
7.7	5	Diesel	4.5
6.5	6	Diesel	2.0
11.2	5	LPG	7.0
8.4	3	Diesel	6.0
5.0	4	LPG	2.5
7.5	3	LPG	4.5
10.0	5	LPG	6.5
18.0	5	Diesel	12.0
12.1	8	Diesel	6.0
7.5	6	Diesel	3.0

Its regression tree is given below. In the tree, each node is denoted by the feature along which it is split.



Based on the above data, answer the given subquestions.

Sub questions

Question Number : 308 Question Id : 640653359449 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

How many samples of the training data-set belong to Leaf 1?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

2

Question Number : 309 **Question Id :** 640653359450 **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

If the present price of a car which is driven for 7 years is 6.0 lac. If the car has a diesel engine, what will be the prediction for the selling price of the car according to the given regression tree? Enter your answer in lac. If your answer is 5.6 lac, enter the answer 5.6.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

2.0

Sub-Section Number : 12

Sub-Section Id : 64065352369

Question Shuffling Allowed : No

Question Id : 640653359452 **Question Type :** COMPREHENSION **Sub Question Shuffling Allowed :** No **Group Comprehension Questions :** No **Calculator :** None **Response Time :** N.A **Think Time :** N.A **Minimum Instruction Time :** 0

Question Numbers : (310 to 311)

Question Label : Comprehension

Consider that the AdaBoost model is trained on the following binary classification dataset.

X_1	X_2	Label (y)
2	5	true
2.5	6	false
3	5	true
4	3	false
4	4	false

The data-set is split according to the feature X_1 to create the first stump. Equal sample weights are assigned to each example to create the first stump and the Gini-index measure is used to split the data.

Based on the above data, answer the given subquestions.

Sub questions

Question Number : 310 Question Id : 640653359453 Question Type : SA Calculator : None

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

Correct Marks : 4

Question Label : Short Answer Question

What will be the performance of the first stump? Enter your answer correct to two decimal places.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

0.67 to 0.72

Question Number : 311 Question Id : 640653359454 Question Type : SA Calculator : None

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

Correct Marks : 4

Question Label : Short Answer Question

What sample weight will be assigned to the first example to create the next stump? Don't

normalize the sample weights. Write your answer correct to two decimal places.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

0.08 to 0.12

PDSA

Section Id :	64065322416
Section Number :	12
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	31
Number of Questions to be attempted :	31
Section Marks :	100
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 :	64065352370
Question Shuffling Allowed :	No

Question Number : 312 **Question Id :** 640653359466 **Question Type :** MCQ Is Question

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

Indian Institute of Technology, Madras - Centre for Continuing Education

Notations :

- 1.Options shown in **green** color and with ✓ icon are correct.
- 2.Options shown in **red** color and with ✗ icon are incorrect.

Question Paper Name :	IIT M DIPLOMA AN1 EXAM ETD1 11 Dec 2022
Subject Name :	2022 Dec: IIT M DIPLOMA AN1 EXAM ETD1
Creation Date :	2022-12-08 17:12:16
Duration :	90
Total Marks :	1030
Display Marks:	Yes
Share Answer Key With Delivery Engine :	Yes
Actual Answer Key :	Yes
Calculator :	Scientific
Magnifying Glass Required? :	No
Ruler Required? :	No
Eraser Required? :	No
Scratch Pad Required? :	No
Rough Sketch/Notepad Required? :	No
Protractor Required? :	No
Show Watermark on Console? :	Yes
Highlighter :	No
Auto Save on Console?	Yes
Change Font Color :	No
Change Background Color :	No

MLP

Section Id :	64065329423
Section Number :	7
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	38
Number of Questions to be attempted :	38
Section Marks :	100
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 :	64065365736
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Number : 179 Question Id : 640653454871 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"

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 :

6406531512751. ✓ YES

6406531512752. ✗ NO

Sub-Section Number : 2

Sub-Section Id : 64065365737

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 180 Question Id : 640653454873 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 will be the shape of X and y?

```
from sklearn.datasets import make_classification
X, y = make_classification(n_features=90, n_informative=10, n_redundant = 1,
n_classes=2, n_clusters_per_class=1, random_state=20, n_samples=120)
```

Options :

6406531512757. ✓ (120,90) and (120,)

6406531512758. ✗ (120,10) and (120,)

6406531512759. ✗ (90,120) and (90,)

6406531512760. ✗ (90,10) and (90,)

Question Number : 181 Question Id : 640653454877 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 will be the shape of the feature matrix (X.shape) and the label vector (y.shape) for the

following block of code?

```
import numpy as np
num_samples = 500
X = 2 * np.random.rand(num_samples, 1)
y = 4 + 3 * X + np.random.randn(num_samples, 1)
```

Options :

6406531512775. ❌ (699, 1) and (599, 1)

6406531512776. ✓ (500, 1) and (500, 1)

6406531512777. ❌ (1, 500) and (1,500)

6406531512778. ❌ (599, 1) and (699, 1)

Question Number : 182 Question Id : 640653454883 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 will be the output of the following code?

```
from sklearn.datasets import load_wine
one = load_wine()
two = load_wine(return_X_y = False)
three = load_wine(as_frame = False)
four = load_wine(as_frame = False, return_X_y = False)

print(type(one))
print(type(two))
print(type(three))
print(type(four))
```

Options :

6406531512793. ❌ Bunch, Tuple, DataFrame, DataFrame

6406531512794. ❌ Bunch, Tuple, DataFrame, Tuple

6406531512795. ✓ Bunch, Bunch, Bunch, Bunch

6406531512796. ❌ DataFrame, DataFrame, DataFrame, DataFrame

6406531512797. ❌ Bunch, Ndarray, Bunch, Ndarray

6406531512798. ❌ It will result into an error.

Question Number : 183 Question Id : 640653454889 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 following code snippet:

```
from sklearn.linear_model import SGDClassifier  
clf = SGDClassifier(loss='perceptron')
```

Above classifier uses:

Options :

6406531512814. ❌ hinge but quadratically penalized.

6406531512815. ❌ logistic regression.

6406531512816. ✓ linear SVM.

6406531512817. ❌ None of these.

6406531512818. ❌ perceptron.

Question Number : 184 Question Id : 640653454891 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 Naive Bayes algorithms can be applied if the data has discrete features that are categorically distributed?

Options :

6406531512823. ❌ GaussianNB

6406531512824. ✓ CategoricalNB

6406531512825. ❌ ComplementNB

6406531512826. ✘ None of these

Question Number : 185 Question Id : 640653454894 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 represents the correct output of the following block of code?

```
X = [[0], [1], [2], [3]]  
y = [0, 0, 2, 2]  
from sklearn.neighbors import KNeighborsClassifier  
neigh = KNeighborsClassifier(n_neighbors=2)  
neigh.fit(X, y)  
print(neigh.predict([[2.1]]))
```

Options :

6406531512835. ✘ [0]

6406531512836. ✓ [2]

6406531512837. ✘ [1.5]

6406531512838. ✘ [1]

Question Number : 186 Question Id : 640653454895 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 estimators implement partial_fit method in Regression?

Options :

6406531512839. ✘ Ridge Regression

6406531512840. ✓ SGDRegressor

6406531512841. ✘ Linear Regression

6406531512842. ✘ Polynomial Regression

Question Number : 187 Question Id : 640653454896 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

Mention which of the following statement(s) is / are correct?

Statement 1: Incremental learning is a dynamic technique which is applied when training data becomes available gradually over time or its size is out of system memory limits.

Statement 2: Sklearn can not process large data in batches.

Options :

6406531512843. ✓ Statement 1 is correct but statement 2 is incorrect.

6406531512844. ✗ Statement 1 is incorrect but statement 2 is correct.

6406531512845. ✗ Statement 1 and statement 2 both are incorrect.

6406531512846. ✗ Statement 1 and statement 2 both are correct.

Question Number : 188 Question Id : 640653454897 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 represents the correct value of gamma for **SVC_classifier = SVC(gamma = 'auto')**?

Options :

6406531512847. ✗ $\frac{1}{(n_features*X.var())}$

6406531512848. ✗ $\frac{1}{(n_features/X.var())}$

6406531512849. ✓ $\frac{1}{n_features}$

6406531512850. ✘ None of these

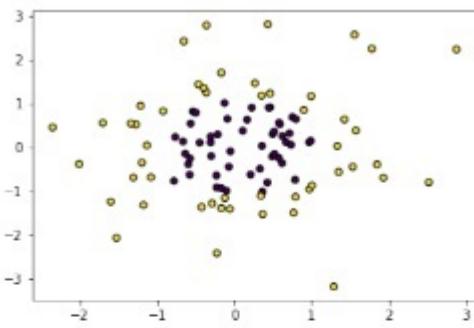
Question Number : 189 Question Id : 640653454901 Question Type : MCQ Is Question

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

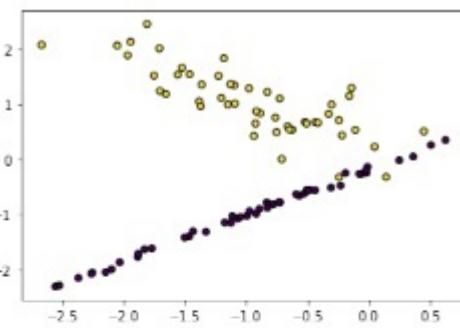
Correct Marks : 2

Question Label : Multiple Choice Question

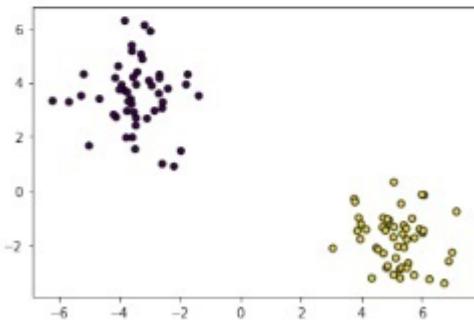
Consider the following scatter plots of four different input datasets:



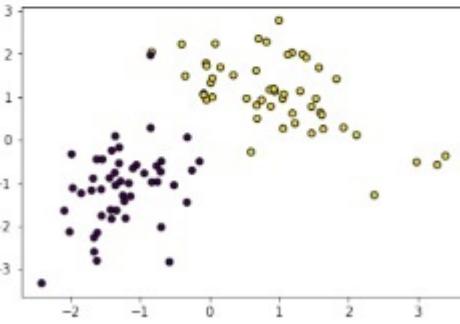
Data 1



Data 2



Data 3



Data 4

Which data will be classified better using gaussian rbf kernel?

Options :

6406531512860. ✓ Data 1

6406531512861. ✘ Data 3

6406531512862. ✘ Data 4

6406531512863. ✘ Data 2

Question Number : 190 Question Id : 640653454903 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 a node is split further in a decision tree:

Options :

6406531512870. ✓ It creates only two new children nodes always.

6406531512871. ✗ It can create two or more new children nodes always.

6406531512872. ✗ It depends on the cardinality of the categorical feature.

6406531512873. ✗ It depends on the scale of the numerical feature.

Question Number : 191 Question Id : 640653454907 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 is/are Correct for K-means algorithm.

Options :

6406531512888. ✗ The algorithm is sensitive to outliers.

6406531512889. ✗ It terminates at a local optimum if SSE is used.

6406531512890. ✗ k-means is considered to be a linear algorithm.

6406531512891. ✓ All of these

Question Number : 192 Question Id : 640653454908 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

Hierarchical Clustering can also be called:

Options :

- 6406531512892. ✘ Centroid Based Clustering
- 6406531512893. ✘ Distribution-based Clustering
- 6406531512894. ✘ Density-based Clustering
- 6406531512895. ✓ Connectivity-Based Clustering

Sub-Section Number : 3

Sub-Section Id : 64065365738

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 193 Question Id : 640653454893 Question Type : MSQ Is Question

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

Correct Marks : 2 Selectable Option : 0

Question Label : Multiple Select Question

Which of the following options represents the major difference between HashingVectorizer and CountVectorizer?

Options :

- 6406531512831. ✘ CountVectorizer does not store vocabulary, its object takes less space compared to HashingVectorizer
- 6406531512832. ✘ HashingVectorizer does not store vocabulary, its object takes less space compared to CountVectorizer
- 6406531512833. ✓ CountVectorizer alleviates any dependence with function calls performed on the previous chunk of data in case of incremental learning
- 6406531512834. ✓ HashingVectorizer alleviates any dependence with function calls performed on the previous chunk of data in case of incremental learning

Question Number : 194 Question Id : 640653454898 Question Type : MSQ Is Question

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

Correct Marks : 2 Selectable Option : 0

Question Label : Multiple Select Question

Which of the following statements correctly set L1 penalty to 0.2 and L2 Penalty to 0.8 in LogisticRegression?

Options :

6406531512851. ✓ `from sklearn.linear_model import LogisticRegression
clf_lr = LogisticRegression(l1_ratio=0.2)`

6406531512852. ✗ `from sklearn.linear_model import LogisticRegression
clf_lr = LogisticRegression(l2_ratio=0.8)`

6406531512853. ✗ `from sklearn.linear_model import LogisticRegression
clf_lr = LogisticRegression(l2_ratio=20)`

6406531512854. ✗ `from sklearn.linear_model import LogisticRegression
clf_lr = LogisticRegression(l1_ratio=80)`

Question Number : 195 Question Id : 640653454904 Question Type : MSQ Is Question

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

Correct Marks : 2 Selectable Option : 0

Question Label : Multiple Select Question

Which of the following are hyper parameters in a decision tree?

Options :

6406531512874. ✗ Number of features

6406531512875. ✓ Number of minimum samples to split

6406531512876. ✓ Height of the tree

6406531512877. ✗ Diameter of the tree

Sub-Section Number : 4

Sub-Section Id : 64065365739

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 196 Question Id : 640653454876 Question Type : MSQ Is Question

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

Correct Marks : 3 Selectable Option : 0

Question Label : Multiple Select Question

Which of the following statements are incorrect about the AUC?

Options :

6406531512770. ✓ The ROC curve plots sensitivity (recall) on the x-axis against specificity on the y-axis

6406531512771. ✗ A completely effective classifier is a diagonal line, and it will have an AUC of 0.5.

6406531512772. ✓ The larger the value of AUC, the less effective the classifier.

6406531512773. ✗ An AUC of 1 indicates a perfect classifier, which means it gets all the 1s correctly classified, and doesn't misclassify any 0s as 1s.

6406531512774. ✗ None of these

Sub-Section Number : 5

Sub-Section Id : 64065365740

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 197 Question Id : 640653454878 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

The output of the following block of code will be:

```
import numpy as np
from sklearn.model_selection import cross_validate
from sklearn.model_selection import ShuffleSplit
rs = ShuffleSplit(n_splits=4, random_state=8)
rs.get_n_splits(X)
```

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

4

Question Number : 198 Question Id : 640653454884 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

Refer the dataframe(df) given below and enter the correct output (up to 2 decimal points) for the following block of code:

	Name	Assignment-1	Assignment-2	Assignment-3
0	juli	22	84	78
1	nick	51	76	61
2	tom	41	76	81
3	Abhi	56	67	82
4	Krish	51	72	84

Figure 2: Sample dataframe

```
import pandas as pd
import numpy as np
df['total'] = df['Assignment-1']+df['Assignment-2']+df['Assignment-3']
print(df['total'].max()-df["total"].min())
```

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

22.95 to 23.05

Question Number : 199 **Question Id :** 640653454887 **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

Enter the output (up to 2 decimal points) for the following block of code.

```
import numpy as np
from sklearn.dummy import DummyClassifier
X = np.array([-2, 2, 3, 3])
y = np.array([0, 2, 2, 2])
dummy_clf = DummyClassifier(strategy="most_frequent")
dummy_clf.fit(X, y)
dummy_clf.score(X, y)
```

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

0.75

Question Number : 200 **Question Id :** 640653454900 **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 correct output of the following block of code?

[Note: LinearSVC implements “one-vs-the-rest” multi-class strategy]

```
from sklearn import svm
X = [[10], [11], [12], [13]]
Y = [0, 1, 2, 3]
lin_clf = svm.LinearSVC()
lin_clf.fit(X, Y)
dec = lin_clf.decision_function([[1]])
dec.shape[1]
```

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

4

Sub-Section Number : 6

Sub-Section Id : 64065365741

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 201 **Question Id :** 640653454872 **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 options represents all the correct precision, Recall, and accuracy values for the confusion matrix shown in Figure 1? [Note: All 3 values should be correct in the same option]

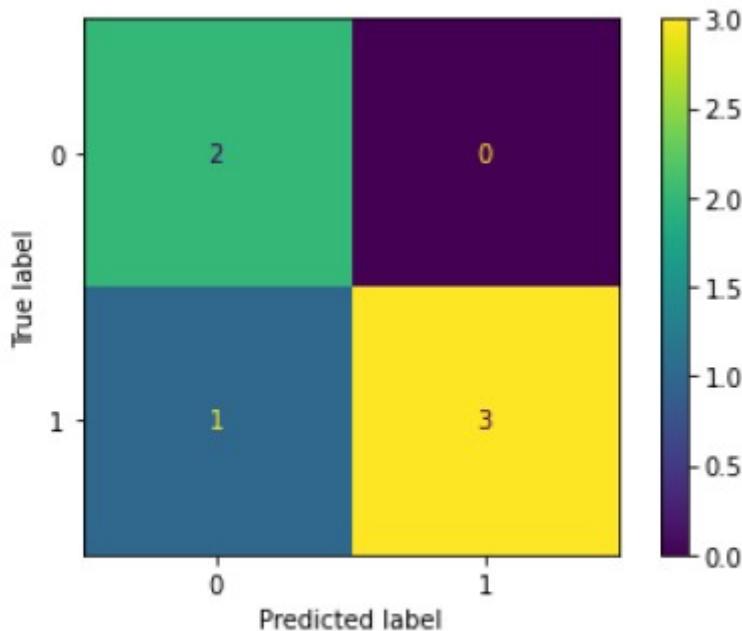


Figure 1: Confusion matrix

Options :

6406531512753. ✓ Precision: 0.833 , Recall: 0.875, Accuracy: 0.833

6406531512754. ✗ Precision: 0.875, Recall: 0.833, Accuracy: 0.833

6406531512755. ✗ Precision: 0.9783, Recall: 0.9281, Accuracy: 0.909

6406531512756. ✗ None of these

Question Number : 202 Question Id : 640653454874 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:

```
import numpy as np
data=np.array([ 2, 1, 2, 1])
from sklearn.preprocessing import PolynomialFeatures
poly= PolynomialFeatures(degree=2, interaction_only=False)
data = data.reshape(2,2)
poly.fit_transform(data)
```

which of the following could be the correct output?

Options :

6406531512761. ✓ array([[1., 2., 1., 4., 2., 1.], [1., 2., 1., 4., 2., 1.]])

6406531512762. ✗ array([[1., 2., 1., 4., 2., 1., 4., 4., 2., 1.], [1., 2., 1., 4., 2., 1., 4., 4., 2., 1.]])

6406531512763. ✗ array([[1., 2., 1., 2.], [1., 2., 1., 2.]])

6406531512764. ✗ array([[1., 2., 1., 4., 2., 1., 8., 4., 2., 1.], [1., 2., 1., 4., 2., 1., 8., 4., 2., 1.]])

Question Number : 203 Question Id : 640653454875 Question Type : MCQ Is Question

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

Time : 0

Correct Marks : 3

Question Label : Multiple Choice Question

Which of the following is likely to be the correct output of the code given below?

```
from sklearn import linear_model
clf = linear_model.Lasso(alpha=0.1)
clf.fit([[2,4], [3, 2], [1, 2]], [2,3,1])
linear_model.Lasso(alpha=0.1,max_iter=1000, tol=0.0001,
warm_start=True,fit_intercept=False)
print(clf.coef_)
```

Options :

6406531512765. ✓ [0.95238095, 0.]

6406531512766. ✗ [0.85,0.1,0.05]

6406531512767. ✗ [2,3,1]

6406531512768. ✗ [0.85,0]

6406531512769. ✗ There are some mistakes in the 3rd /4th line of code, hence it will produce error.

Question Number : 204 Question Id : 640653454879 Question Type : MCQ Is Question

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

Time : 0

Correct Marks : 3

Question Label : Multiple Choice Question

Which of the following is likely to be the correct output of the code given below?

```
from sklearn import linear_model
clf = linear_model.Ridge(alpha=1)
X= [[2,1], [2, 3], [3, 2]]
y= [5, 10, 5]
clf.fit(X, y)
linear_model.Ridge(alpha=1,max_iter=1000, tol=0.0001,fit_intercept=True)
clf.score(X,y)
```

Options :

6406531512780. ✘ 5

6406531512781. ✘ 99

6406531512782. ✘ Given code will produce an error

6406531512783. ✓ 0.82

Question Number : 205 Question Id : 640653454885 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

Mention whether the following statements are TRUE or FALSE .

Statement 1 : 'In order to extend ROC curve and ROC area to multi-label classification, it is necessary to binarize the output.'

Statement 2: In SGDClassifier, setting warm_start=True does not retain the weight values of the model after max_iter and hence does not produce different results for each execution.'

Options :

6406531512800. ✓ Statement 1 is correct but statement 2 is incorrect.

6406531512801. ✘ Statement 1 is incorrect but statement 2 is correct.

6406531512802. ✘ Statement 1 and statement 2 both are incorrect.

6406531512803. ✘ Statement 1 and statement 2 both are correct.

Question Number : 206 Question Id : 640653454888 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 options represents the correct option to set the regularization rate in RidgeClassifier?

Options :

6406531512810. ❌ logit_classifier = LogisticRegression(penalty='12')

6406531512811. ❌ SGDClassifier(loss="perceptron", eta0=1, learning_rate="constant",
penalty=None)

```
if solver == 'auto':  
    if return_intercept:  
        solver = "sag"  
    elif not sparse.issparse(X):  
        solver = "cholesky"  
    else:  
        solver = "sparse_cg"
```

6406531512812. ❌

6406531512813. ✓
from sklearn.linear_model import RidgeClassifier
ridge_classifier = RidgeClassifier(alpha=0.001)

Question Number : 207 Question Id : 640653454890 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 options represents the correct output for the following block of code?

```
text_data=['A metaverse is a network of 3D virtual worlds focused on connection.']
from sklearn.feature_extraction.text import CountVectorizer
c_vectorizer = CountVectorizer()
X_c = c_vectorizer.fit_transform(text_data)
print(X_c.shape)
```

Options :

6406531512819. ✘ (11, 1)

6406531512820. ✓ (1, 10)

6406531512821. ✘ (1, 11)

6406531512822. ✘ (10, 1)

Question Number : 208 Question Id : 640653454892 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 options represents the correct output for the following block of code?

```
from sklearn.feature_extraction.text import TfidfVectorizer
corpus = [ 'The name is Bond, James Bond!']
vectorizer = TfidfVectorizer()
X = vectorizer.fit_transform(corpus)
vectorizer.get_feature_names_out()
```

Options :

6406531512827. ✘ array(['is', 'name', 'the'], dtype=object)

6406531512828. ✘ array(['bond', 'is', 'james', 'name', 'the', '!'], dtype=object)

6406531512829. ✘ array(['bond', 'bond', 'is', 'james', 'name', 'the'], dtype=object)

6406531512830. ✓ array(['bond', 'is', 'james', 'name', 'the'], dtype=object)

Question Number : 209 Question Id : 640653454899 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 options represents the correct output of the following block of code?

```
from sklearn import svm
X = [[10, 10], [11, 11]]
y = [0, 2]
clf = svm.SVC()
clf.fit(X, y)
clf.predict([[0.5, 0.5]])
```

Options :

6406531512855. ✓ array([2])

6406531512856. ✗ array([1.5])

6406531512857. ✗ array([1])

6406531512858. ✗ array([0])

Question Number : 210 Question Id : 640653454902 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 following statements regarding a decision tree model?

1. It is a parametric model.
2. It has a tendency to overfit if allowed to grow unconditionally.
3. It can be used for regression, classification and clustering problems.
4. It is one of the easiest machine learning models to interpret.
5. For classification problems, decision tree's loss function is same as evaluation metric.

Choose the option with all the correct statements:

Options :

6406531512864. ✓ 2, 3

6406531512865. ✗ 1, 3, 5

6406531512866. ✗ 3, 5

6406531512867. ✗ 2, 4

6406531512868. ✗ 1, 3, 4

6406531512869. ✗ 1, 2, 4

Question Number : 211 Question Id : 640653454906 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

Lata wants to train a bagging regressor with following specifications:

1. Base estimator as linear regressor with default parameter values.
2. 80% of the data is used to create a bag of samples.
3. Create subsets of training set without replacement.
4. Take half the features when creating subsets of data.
5. Train the model on training set.
6. Print score on test set.

Which of the following code blocks will correctly execute Lata's task?

[Note: Assume necessary imports and variables for training and test sets.]

Options :

bagging = BaggingRegressor(LinearRegression(), max_samples = 0.8,
 bootstrap = False, max_features = 0.5)
bagging.fit(X_train, y_train)
6406531512883. ✓ print(bagging.score(X_test, y_test))

bagging = BaggingRegressor(LinearRegression(), max_samples = 80,
 bootstrap = True, max_features = 'all')
bagging.fit(X_train, y_train)
6406531512884. ✗ print(bagging.score(X_test, y_test))

bagging = BaggingRegressor(LinearRegression(), max_samples = 0.8,
 bootstrap = True, max_features = 'all')
bagging.fit(X_train, y_train)
6406531512885. ✗ print(bagging.score(X_test, y_test))

bagging = BaggingRegressor(LinearRegression(), max_samples = 80,
 bootstrap = False, max_features = 'all')
bagging.fit(X_train, y_train)
6406531512886. ✗ print(bagging.score(X_test, y_test))

```
bagging = BaggingRegressor(LinearRegression(), bag_size = 0.8,
                           bootstrap = True, all_features = True)
bagging.fit(X_train, y_train)
6406531512887. ✘ print(bagging.score(X_test, y_test))
```

Question Number : 212 Question Id : 640653454909 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 have a binary classification problem to solve with following information:

1. All features are numerical.
2. The dataset has size of 80 GB.
3. The data has to be preprocessed.
4. After data preprocessing, train the model with gradient descent or its variations.

Which of the following will perform the above task?

Options :

6406531512896. ✘ Perform one hot encoding followed by training with LogisticRegression model (sklearn.linear_model).

6406531512897. ✘ Perform standard scaling followed by a sklearn.neighbors.KNeighborsClassifier model.

6406531512898. ✘ Perform min-max scaling followed by a sklearn.tree.DecisionTreeClassifier model.

6406531512899. ✓ Perform max-absolute scaling followed by a sklearn.linear_model.SGDClassifier model.

Question Number : 213 Question Id : 640653454910 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

Let data-set X has 784 features and the trained using the following code:

```
mlp_clf = MLPClassifier(hidden_layer_sizes=(100,),  
                        alpha=1e-4,  
                        solver="sgd",  
                        learning_rate_init=0.2,  
                        max_iter=50,  
                        random_state=1)  
  
mlp_clf.fit(X, y)
```

What will be the output of following code:

```
w = mlp_clf.coefs_  
w = np.array(w[0])  
w.shape
```

Options :

6406531512900. ✓ (784, 100)

6406531512901. ✗ (784, 1)

6406531512902. ✗ (1, 784)

6406531512903. ✗ (100, 784)

Sub-Section Number : 7

Sub-Section Id : 64065365742

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 214 Question Id : 640653454886 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

Which of the following options represents the correct output of the following block of code?

```
from sklearn.metrics import log_loss
loss=log_loss(["spam", "ham"], [[.04,.16], [.16, .04]],eps=1e-15,
normalize=True, sample_weight=None, labels=None)
print(loss)
```

Options :

6406531512804. ✘ 1

6406531512805. ✘ 0

6406531512806. ✘ 0.110

6406531512807. ✓ 0.223

6406531512808. ✘ 0.533

Question Number : 215 Question Id : 640653454905 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

Brajesh wants to tune hyper parameters of an AdaBoost model for a classification problem with following specifications:

1. Base estimator as Decision trees classifier with max_depth=1.
2. Number of estimators range from 30 to 40 (both inclusive) at the interval of 1.
3. Cross validation = 5.
4. Learning rate must vary between 0.2 to 2.0 (both inclusive) at the intervals of 0.2.
5. Train the best model on the entire training set.
6. Print score on test set.

Which of the following code blocks will correctly execute Brajesh's task?

[Note: Assume necessary imports and variables for training and test sets.]

Options :

6406531512878. ✓

```
params = {'n_estimators':list(range(30,41)),
          'learning_rate': np.linspace(0.1,1.0, 10)}

ada = AdaBoostClassifier()
gs_ada = GridSearchCV(ada, cv = 5, refit = True, param_grid=params)
gs_ada.fit(X_train, y_train)

print(gs_ada.best_estimator_.score(X_test, y_test))

params = {'n_estimators':list(range(30,51,2)),
          'learning_rate': np.linspace(0.1,1.0, 10)}

ada = AdaBoostClassifier()
gs_ada = GridSearchCV(ada, cv = 5, refit=True, param_grid =params )
gs_ada.fit(X_train, y_train)

print(gs_ada.best_estimator_.score(X_test, y_test))

6406531512879. ✘ print(gs_ada.best_estimator_.score(X_test, y_test))

params = {'n_estimators':list(range(10,21,2)),
          'learning_rate': np.linspace(0.1,1.0, 10)}

ada = AdaBoostClassifier()
gs_ada = GridSearchCV(ada, cv = 5, fit_best=True, parameters =params )
gs_ada.fit(X_train, y_train)

print(gs_ada.best_estimator_.score(X_test, y_test))

6406531512880. ✘ print(gs_ada.best_estimator_.score(X_test, y_test))

params = {'n_estimators':list(range(10,21,2)),
          'learning_rate': np.linspace(0.1,1.0, 10)}

ada = AdaBoostClassifier()
gs_ada = GridSearchCV(base_estimator = ada,
                      cross_validation = 5, fit_best=True,
                      parameters =params)
gs_ada.fit(X_test, y_test)

print(gs_ada.best_estimator_.score(X_train, y_train))
```

6406531512881. ✘

6406531512882. ✘

```
params = {'n_estimators':list(range(10,21,2)),
          'learning_rate': np.linspace(0.1,1.0, 10)}

ada = AdaBoostClassifier()
gs_ada = GridSearchCV(base_estimator = ada,
                      cross_validation = 5,
                      parameters =params)

best_model = gs_ada.get_best_model().fit(X_train, y_train)

print(gs_ada.best_estimator_.score(X_test, y_test))
```

Sub-Section Number : 8

Sub-Section Id : 64065365743

Question Shuffling Allowed : No

Is Section Default? : null

Question Id : 640653454880 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 : (216 to 217)

Question Label : Comprehension

Go through the code snippet given below and answer the subquestions.

```
from sklearn.preprocessing import PolynomialFeatures
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import RidgeCV
from sklearn.datasets import load_diabetes
dataset = load_diabetes(as_frame=True)
X=dataset.data
y=dataset.target

lf= np.arange(0.005, 1, 0.05)

ridge_reg_pipeline = Pipeline([
    ("poly", PolynomialFeatures(degree=2)),
    ("feature_scaling", StandardScaler())
])

ridge= RidgeCV(alphas=lf,scoring="neg_mean_squared_error")
results = ridge.fit(X,y)
```

Sub questions

Question Number : 216 Question Id : 640653454881 Question Type : MSQ Is Question

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

Correct Marks : 3 Selectable Option : 0

Question Label : Multiple Select Question

Based on the given code,Select all option Which will be considered as input alpha value for our RidgeCV estimator.

Options :

6406531512784. ✓ 0.055

6406531512785. ✗ 0.05

6406531512786. ✓ 0.905

6406531512787. ✗ 0.1

6406531512788. ✗ 1

Question Number : 217 Question Id : 640653454882 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 could be the possible output of print(results.best_score_)?

Options :

6406531512789. ✘ 1

6406531512790. ✓ -2999.79

6406531512791. ✘ 0.528

6406531512792. ✘ 0.681

PDSA

Section Id :	64065329424
Section Number :	8
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	30
Number of Questions to be attempted :	30
Section Marks :	100
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 :	64065365744
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Number : 218 Question Id : 640653454911 Question Type : MCQ Is Question

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

Indian Institute of Technology, Madras - BS in Data Science and Applications

Notations :

- 1.Options shown in **green** color and with ✓ icon are correct.
- 2.Options shown in **red** color and with ✗ icon are incorrect.

Question Paper Name :

IIT M DIPLOMA ET1 EXAM QPD1 S2 30 Apr
2023

Subject Name :

2023 Apr30: IIT M DIPLOMA ET1 EXAM
QPD1

Creation Date :

2023-04-21 17:54:54

Duration :

90

Total Marks :

1145

Display Marks:

Yes

Share Answer Key With Delivery Engine :

Yes

Actual Answer Key :

Yes

Calculator :

Scientific

Magnifying Glass Required? :

No

Ruler Required? :

No

Eraser Required? :

No

Scratch Pad Required? :

No

Rough Sketch/Notepad Required? :

No

Protractor Required? :

No

Show Watermark on Console? :

Yes

Highlighter :

No

Auto Save on Console?

Yes

Change Font Color :

No

Correct Marks : 8

Question Label : Multiple Choice Question

What will be the output of the following command?

```
echo {a,b,c,c}{c,c,e,f} | tr ' ' '\n' | sort | uniq | awk '{  
    n += NR  
}  
END {  
    print n  
}  
'
```

Options :

6406531892394. ✓ 45

6406531892395. ✗ 78

6406531892396. ✗ 136

6406531892397. ✗ 55

MLP**Section Id :** 64065338407**Section Number :** 10**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

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 :	64065380971
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Number : 255 Question Id : 640653566235 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 :

6406531892406. ✓ YES

6406531892407. ✗ NO

Sub-Section Number :	2
Sub-Section Id :	64065380972
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 256 Question Id : 640653566236 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

_____ generates a bunch of normally-distributed clusters of points with specific mean and standard deviations for each cluster.

Options :

6406531892408. ❌ `sklearn.datasets.make_clusters()`

6406531892409. ❌ `sklearn.datasets.make_centers()`

6406531892410. ❌ `sklearn.datasets.make_normal_clusters()`

6406531892411. ✓ `sklearn.datasets.make_blobs()`

Question Number : 257 Question Id : 640653566238 Question Type : MCQ Is Question

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

Correct Marks : 2

Question Label : Multiple Choice Question

Why is data preprocessing necessary?

Options :

6406531892416. ❌ Some columns have values only between 0 and 1

6406531892417. ✓ The data is divided into multiple types of files i.e. html, csv, tsv, etc.

6406531892418. ❌ The data has only numbers in all the columns.

Question Number : 258 Question Id : 640653566241 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 is correct with respect to R2-score?

Options :

6406531892429. ❌ R2 score is always positive and it may go up to infinity.

6406531892430. ❌ R2 score is always positive, but it ranges between 0 and 1 only.

6406531892431. ✓ R2 score can be negative. That happens if our model is worse than the mean model.

6406531892432. ❌ R2 score can be negative. That happens if the mean model is worse than our model.

Question Number : 259 Question Id : 640653566242 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 options represent the output of the following code block?

```
from sklearn.metrics import confusion_matrix
y_true = ["Rainy", "Sunny", "Sunny", "Rainy", "Sunny", "Rainy"]
y_pred = ["Sunny", "Sunny", "Rainy", "Rainy", "Sunny", "Sunny"]
confusion_matrix(y_true, y_pred, labels=["Rainy", "Sunny"])
```

Options :

6406531892433. ❌ array([[2, 1], [2, 1]])

6406531892434. ❌ array([[2, 1], [1, 2]])

6406531892435. ✓ array([[1, 2], [1, 2]])

array([[1, 2],
[2, 1]])
6406531892436. ✘

Question Number : 260 Question Id : 640653566243 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Consider the following code block:

```
from sklearn.datasets import make_regression
X, y = make_regression(n_samples = 10, n_features = 3)

from sklearn.model_selection import LeavePOut
lpo = LeavePOut(p = 2)
count = 0
for train, test in lpo.split(X):
    print(train, test)
    count += 1
print(count)
```

What will be the value of the 'count'?

Options :

6406531892437. ✘ 20

6406531892438. ✘ 30

6406531892439. ✓ 45

6406531892440. ✘ 15

Question Number : 261 Question Id : 640653566244 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction

Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Consider the following code block:

```
X = ['a', 'b', 'c', 'd', 'e', 'f']
from sklearn.model_selection import RepeatedKFold
rkf = RepeatedKFold(n_splits = 3, n_repeats = 2, random_state = 10)
for train, test in rkf.split(X):
    print(train, test)
```

Which of the following may be the correct output of the above code?

Options :

[0 1] [3 4] [2 5]
[1 2] [4 5] [0 3]
[0 2] [3 5] [1 4]
[1 3] [4 5] [0 2]
[0 2] [3 4] [1 5]

6406531892441. ✘ [0 1] [2 5] [3 4]

[0 1 3 4] [3 5]
[1 2 4 5] [0 2]
[0 2 3 5] [1 5]
[1 3 4 5] [0 4]
[0 2 3 4] [1 0]

6406531892442. ✘ [0 1 2 5] [3 2]

[0 1 3 4] [2 5]
[1 2 4 5] [0 3]
[0 2 3 5] [1 4]
[1 3 4 5] [0 2]
[0 2 3 4] [1 5]

6406531892443. ✓ [0 1 2 5] [3 4]

6406531892444. ✘

[0 1 1 4] [2 5]
[1 2 5 5] [0 3]
[0 2 0 5] [1 4]
[1 3 1 5] [0 2]
[0 2 0 4] [1 5]
[0 1 5 5] [3 4]

[0 1] [3 4] [0 1]
[1 2] [1 2] [0 3]
[0 2] [3 5] [3 5]
[1 3] [4 5] [1 3]
[0 2] [3 4] [0 2]
[0 1] [0 1] [3 4]

6406531892445. ✘

Question Number : 262 Question Id : 640653566247 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 is a hyperparameter?

Options :

6406531892452. ✘ L1-ratio in elasticnet

6406531892453. ✘ Pruning parameter in a decision tree

6406531892454. ✘ Learning rate in SGDRegressor

6406531892455. ✓ All of these

Question Number : 263 Question Id : 640653566255 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 estimators can implement the `partial_fit` method ?

Options :

6406531892484. ❌ `DecisionTreeClassifier`

6406531892485. ❌ `RandomForestRegressor`

6406531892486. ❌ `LogisticRegressor`

6406531892487. ✓ `SGDRegressor`

Question Number : 264 Question Id : 640653566256 Question Type : MCQ Is Question

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

Correct Marks : 2

Question Label : Multiple Choice Question

Consider the following code

```
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression

pipe = [
    ('scaler', StandardScaler()),
    ('softmax', _____)
]
```

Which of the following options is True for blank space if I want to train the above pipeline as softmax regression ?

Options :

6406531892488. ❌ `LogisticRegression(solver = 'sag')`

6406531892489. ❌ `LogisticRegression(multi_class = 'ovr')`

6406531892490. ✘ LogisticRegression(solver = 'lbfgs')

6406531892491. ✓ LogisticRegression(multi_class = 'multinomial')

Question Number : 265 Question Id : 640653566258 Question Type : MCQ Is Question

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

Correct Marks : 2

Question Label : Multiple Choice Question

Consider the following code:

```
from sklearn.metrics import f1_score
y_true = [1,1,0,1,0,0,1,0,1]
y_pred = [0,1,0,1,0,1,1,1,1]
print(f1_score(y_true,y_pred))
```

What will be the output?

Options :

6406531892496. ✘ 0.66

6406531892497. ✓ 0.72

6406531892498. ✘ 0.80

6406531892499. ✘ 1.00

Question Number : 266 Question Id : 640653566259 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 are correct about sklearn.svm.SVC:

```
from sklearn.svm import SVC  
clf = SVC(C=1).fit(X_train, y_train)  
print(clf.support_)
```

What will be the output?

Options :

6406531892500. ✘ It will print number of support vectors

6406531892501. ✓ It will print an array of support vectors

6406531892502. ✘ It will print an array of probabilities representing distance from decision boundary with each data point.

6406531892503. ✘ It will print indices of the support vectors from the training set.

Question Number : 267 Question Id : 640653566260 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=1)  
neigh.fit(X_train)  
print(neigh.kneighbors(X_test[1:2,:]))
```

Assume X_train and X_test are of type numpy.ndarray .

Options :

6406531892504. ✘ It will print nearest neighbours from the test point

6406531892505. ✓ It will print indices of and distances to the neighbouring points (in training set) from test point

6406531892506. ✘ It will print indices, distance and nearest training point from the test point

6406531892507. ✘ It will throw an error

Sub-Section Number :	3
Sub-Section Id :	64065380973
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 268 Question Id : 640653566239 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 APIs only supports conjoining transformers and estimators in series (i.e. one after another)?

Options :

6406531892419.  Pipeline

6406531892420.  ColumnTransformer

6406531892421.  FeatureUnion

6406531892422.  All of these

Question Number : 269 Question Id : 640653566240 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 ML task/steps for a regression dataset:

1. Read the data from a file (named 'dataset.csv'). It has 7 columns. The last column is the target variable, all 6 features numerical.
2. Remove rows which have target values missing.
3. Fill the missing values in the features by KNN using 3 nearest neighbours.
4. Split the data into training and test sets. Take randomly the 70% of rows in the training set and

the rest of them into the test set.

5. Train a simple linear regression model, with intercept, on the training set.

6. Report R2 score on the test set.

Which of the following code snippets correctly accomplishes the above task? Assume necessary imports.

Options :

```
data = pd.read_csv('dataset.csv')
data = data.dropna()
X, y = data[data.columns[:-1]], data[data.columns[-1]]
X = X[~y.isna()]
y = y.dropna()
X_train, X_test, y_train, y_test = train_test_split(X,y,
test_size=0.7)
pipe = Pipeline([('imputer', KNNImputer(n_neighbors = 3)),
('estimator', LinearRegression())])

pipe.fit(X_train,y_train)
print(pipe.score(X_test, y_test))
```

6406531892423. ✓

```
data = pd.read_csv('dataset.csv')
data = data.dropna()
X, y = data[data.columns[:-1]], data[data.columns[-1]]
X = X[y.isna()]
y = y.dropna()
X_train, X_test, y_train, y_test = train_test_split(X,y,
test=0.3)
pipe = Pipeline([('imputer', KNNImputer(n_neighbors = 3)),
('estimator', LinearRegression())])

pipe.fit(X_train,y_train)
print(pipe.score(X_test, y_test))
```

6406531892424. ✘

6406531892425. ✘

```
data = pd.read_csv('dataset.csv')
data = data.dropna()
X, y = data[data.columns[:-1]], data[data.columns[-1]]
X = X[~y.isna()]
y = y[y.isna()]
X_train, X_test, y_train, y_test = train_test_split(X,y,
                                                    train_size=0.8)
pipe = Pipeline([('imputer', KNNImputer(n_neighbors = 3)),
                 ('estimator', LinearRegression())])

pipe.fit(X_train,y_train)
print(pipe.score(y_test, y_test))
```

```
data = pd.read_csv('dataset.csv')
data = data.dropna()
X, y = data[data.columns[:-1]], data[data.columns[-1]]
X = X[~y.isna()]
y = y.dropna()
X_train, X_test, y_train, y_test = train_test_split(X,y,
                                                    test=0.2)
pipe = Pipeline([('imputer', KNNImputer(n_neighbors = 2)),
                 ('estimator', LinearRegression(fit_intercept=False))])

pipe.fit(X_train,y_train)
print(pipe.score(X_test, X_test))
```

6406531892426. *

```
data = pd.read_csv('dataset.csv')
data = data.dropna()
X, y = data[data.columns[:-1]], data[data.columns[-1]]

X_train, X_test, y_train, y_test = train_test_split(X,y,
                                                    test_size=0.2)
pipe = Pipeline([('imputer', KNNImputer(n_neighbors = 3)),
                 ('estimator', LinearRegression())])

pipe.fit(X_train,y_train)
print(pipe.score(X_test, X_test))
```

6406531892428. * None of these

Question Number : 270 Question Id : 640653566246 Question Type : MCQ Is Question

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

Time : 0

Correct Marks : 3

Question Label : Multiple Choice Question

What will be the output of the following code?:

```
from sklearn.datasets import make_regression
from sklearn.preprocessing import PolynomialFeatures
X, y = make_regression(n_samples = 10, n_features = 3)
poly_transform = PolynomialFeatures(degree=2, interaction_only=True)
X_trans = poly_transform.fit_transform(X)
print(X_trans.shape)
```

Options :

6406531892447. ✘ (10, 3)

6406531892448. ✓ (10, 7)

6406531892449. ✘ (10, 8)

6406531892450. ✘ (10, 10)

6406531892451. ✘ (10, 11)

Question Number : 271 Question Id : 640653566249 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 statements about SGDClassifier:

1. It can be used to train a model on large dataset that doesn't fit in main memory
2. It can emulate a KNN model
3. It can emulate a decision tree model
4. It can emulate a perceptron

Choose the correction option(s)

Options :

6406531892460. ✓ 1 and 4

6406531892461. ✗ 1 and 2

6406531892462. ✗ 2 and 3

6406531892463. ✗ 3 and 4

Question Number : 272 Question Id : 640653566250 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 an image classification task: an image can have dogs, birds and trees. An image can have any combination of these three. The classifier is expected to report all these three for every sample. What kind of classification problem is this?

Options :

6406531892464. ✓ multi-label and multiclass problem.

6406531892465. ✗ multi-label and binary class problem.

6406531892466. ✗ multiclass problem.

6406531892467. ✗ binary class single label problem.

Question Number : 273 Question Id : 640653566254 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 following code snippets, assuming necessary imports.

```
model1 = KNeighborsClassifier(n_neighbors=2)
model2 = KNeighborsClassifier(n_neighbors=5)

model1.fit(X_train,y_train)
model2.fit(X_train,y_train)
```

Choose the correct options:

Options :

6406531892480. ✓ model1 will have smoother decision boundary compared to model model2

6406531892481. ✗ model2 will have smoother decision boundary compared to model model1

6406531892482. ✗ model1 will have same decision boundary compared as model model2

6406531892483. ✗ No comparison can be made between decision boundaries of model1 and model2

Question Number : 274 Question Id : 640653566257 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 options is true for the gamma parameter in a non-linear soft margin SVM?

Options :

6406531892492. ✓ For high values of gamma, the points need to be very close to each other in order to be considered in the same class

6406531892493. ✗ For low values of gamma, the points need to be very close to each other in order to be considered in the same class

6406531892494. ✗ Gamma doesn't affect the SVM model at all

6406531892495. ✗ None of these

Question Number : 275 Question Id : 640653566261 Question Type : MCQ Is Question

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

Time : 0

Correct Marks : 3

Question Label : Multiple Choice Question

Consider the following code:

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

dtc1 = DecisionTreeClassifier(ccp_alpha = 0.0)
dtc1.fit(X, y)
dtc2 = DecisionTreeClassifier(ccp_alpha = 0.06)
dtc2.fit(X, y)

dtc3 = DecisionTreeClassifier(ccp_alpha = 0.1)
dtc3.fit(X, y)

dtc4 = DecisionTreeClassifier(ccp_alpha = 0.03)
dtc4.fit(X, y)
```

Which model is likely to overfit the most?

Options :

6406531892508. ✓ dtc1

6406531892509. ✗ dtc2

6406531892510. ✗ dtc3

6406531892511. ✗ dtc4

Question Number : 276 Question Id : 640653566265 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

Following is the code to tune the n_estimators parameter of a Bagging Classifier model.

```
from sklearn.model_selection import GridSearchCV
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import BaggingClassifier

param_grid = [
{_____ : [200, 300, 400, 500, 600]}
]
pipeline = Pipeline(steps=[('scaler', StandardScaler()),
('bc', BaggingClassifier())])
grid_search = GridSearchCV(pipeline, param_grid, cv=5,
scoring='neg_mean_squared_error',
return_train_score=True)
grid_search.fit(X_train, y_train)
```

What should the blank space contain?

Options :

6406531892526. ✘ 'n_estimators'

6406531892527. ✘ 'bc_n_estimators'

6406531892528. ✓ 'bc__n_estimators'

6406531892529. ✘ 'bc___n_estimators'

6406531892530. ✘ 'bc.n_estimators'

Sub-Section Number :

4

Sub-Section Id :

64065380974

Question Shuffling Allowed :

Yes

Is Section Default? :

null

Question Number : 277 Question Id : 640653566262 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 DecisionTreeClassifier models will be trained internally?

```
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import GridSearchCV
param_grid = [ {'max_depth':range(1, 10, 2)}, {'min_samples_split': range(1, 10, 3)}]
gs = GridSearchCV(DecisionTreeClassifier(), param_grid, cv = 10)
gs.fit(X,y)
```

Options :

6406531892512. ✘ 20

6406531892513. ✘ 200

6406531892514. ✘ 8

6406531892515. ✘ 150

6406531892516. ✘ 15

6406531892517. ✓ 80

Question Number : 278 Question Id : 640653566264 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 two classifiers as shown in 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_wine(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)
clf1 = DecisionTreeClassifier(min_samples_split = 3, min_samples_leaf = 2,
                               random_state = 5)
clf1.fit(X_train, y_train)

clf2 = DecisionTreeClassifier(min_samples_split = 6, min_samples_leaf = 4,
                               random_state = 5)
clf2.fit(X_train, y_train)

```

What can we say about the depths of the classifiers clf1 and clf2?

Options :

6406531892522. ✓ depth(clf1) ≥ depth(clf2)

6406531892523. ✗ depth(clf1) ≤ depth(clf2)

6406531892524. ✗ depth(clf1) = depth(clf2)

6406531892525. ✗ Insufficient Information

Question Number : 279 Question Id : 640653566266 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:

```

from sklearn.tree import DecisionTreeClassifier
from sklearn.datasets import load_wine

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

dtc1 = DecisionTreeClassifier(ccp_alpha = 0.0)
dtc1.fit(X, y)

dtc2 = DecisionTreeClassifier(ccp_alpha = 0.03)
dtc2.fit(X, y)

dtc3 = DecisionTreeClassifier(ccp_alpha = 0.06)
dtc3.fit(X, y)

dtc4 = DecisionTreeClassifier(ccp_alpha = 0.1)
dtc4.fit(X, y)

d1 = dtc1.get_depth()
d2 = dtc2.get_depth()
d3 = dtc3.get_depth()
d4 = dtc4.get_depth()

```

What can we say about d1, d2, d3 and d4?

Options :

6406531892531. ❌ $d_1 < d_2 < d_3 < d_4$

6406531892532. ❌ $d_1 \leq d_2 \leq d_3 \leq d_4$

6406531892533. ❌ $d_1 > d_2 > d_3 > d_4$

6406531892534. ✓ $d_1 \geq d_2 \geq d_3 \geq d_4$

Question Number : 280 Question Id : 640653566268 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 that we have 10000 samples in a dataset. Suppose further we use the $K -$ means algorithm to find clusters in the dataset. Then, the statement that K-Means algorithm always converges with zero inertia (or zero Sums Square Error) for some value of K is

Options :

6406531892539. ✓ Always True

6406531892540. ✗ Always False

6406531892541. ✗ True, sometimes

Question Number : 281 Question Id : 640653566269 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 that we use k-means clustering for a dataset having 100 samples. The initial centroids for k clusters can be initialized in multiple ways. One such way is shown below

```
km = KMeans(n_clusters=20, init='random', n_init=10, random_state=42)
km.fit(X)
```

Choose the correct statements

Options :

6406531892542. ✓ 20 centroids are randomly initialized 10 times

6406531892543. ✗ 10 centroids are randomly initialized 20 times

6406531892544. ✗ 20 samples in the dataset are selected as initialization point such that they are at least 10 units away from each other

6406531892545. ✗ 10 samples in the dataset are selected as initialization point such that they are at least 20 units away from each other

Sub-Section Number : 5

Sub-Section Id : 64065380975

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 282 Question Id : 640653566237 Question Type : MSQ Is Question

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

Correct Marks : 2 Selectable Option : 0

Question Label : Multiple Select Question

We wish to load the wine dataset from sklearn. Which of the following will throw an error?

Options :

6406531892412. ✓ `from sklearn.datasets import load_wine
X, y = load_wine(load_X_y = True)`

6406531892413. ✓ `from sklearn.datasets import load_wine
data = load_wine(load_X_y = True)`

6406531892414. ✗ `from sklearn.datasets import load_wine
data = load_wine(return_X_y = True)`

6406531892415. ✗ `from sklearn.datasets import load_wine
X, y = load_wine(return_X_y = True)`

Question Number : 283 Question Id : 640653566248 Question Type : MSQ Is Question

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

Correct Marks : 2 Selectable Option : 0

Question Label : Multiple Select Question

Which of the following is correct?

Options :

6406531892456. ✗ SGDClassifier(loss="percept") is stochastic version of a perceptron model

6406531892457. ✓ SGDClassifier(loss="log_loss") is stochastic version of a logistic classifier model

6406531892458. ❌ SGDClassifier(loss="log_loss") is stochastic version of a SVM model

6406531892459. ✓ SGDClassifier(loss="hinge") is stochastic version of a SVM model

Question Number : 284 Question Id : 640653566251 Question Type : MSQ Is Question

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

Correct Marks : 2 Selectable Option : 0

Question Label : Multiple Select Question

Which of the following algorithms may get impacted by feature scaling?

Options :

6406531892468. ✓ LinearRegression

6406531892469. ❌ DecisionTree

6406531892470. ✓ SVM

6406531892471. ❌ BinomialNaiveBayes

Question Number : 285 Question Id : 640653566270 Question Type : MSQ Is Question

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

Correct Marks : 2 Selectable Option : 0

Question Label : Multiple Select Question

Which of the following class(es) is (are) used to instantiate a neural network in Sklearn.

Options :

6406531892546. ❌ SGDClassifier()

6406531892547. ✓ MLPClassifier()

6406531892548. ✖ NNClassifier()

6406531892549. ✓ MLPRegressor()

Sub-Section Number :	6
Sub-Section Id :	64065380976
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 286 Question Id : 640653566252 Question Type : MSQ Is Question

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

Correct Marks : 3 Selectable Option : 0

Question Label : Multiple Select Question

Following information about X_train is given:

- Shape of X_train is (100,6)
- 4 continuous features, 2 categorical features
- One categorical feature contains 3 categories/unique values
- Second categorical feature contains 4 categories/unique values

```
from sklearn.preprocessing import OneHotEncoder  
Ohe = OneHotEncoder()
```

```
Encoded_X_train = ohe.fit_transform(X_train)
```

```
Encoded_X_train.shape
```

Which of the following is(are) correct option(s) for above information ?

Options :

6406531892472. ✓ Encoded_X_train will have more number of columns than X_train

6406531892473. ✖ Encoded_X_train will have 11 columns

6406531892474. ✓ Encoded_X_train will have more than 11 columns

6406531892475. ✘ Encoded_X_train will have more number of rows than X_train

Question Number : 287 Question Id : 640653566253 Question Type : MSQ Is Question

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

Correct Marks : 3 Selectable Option : 0

Question Label : Multiple Select Question

Which of the following ways can help in feature selection?

Options :

6406531892476. ✓ Drop a Feature with many missing values

6406531892477. ✘ Drop a feature containing data with high standard deviation

6406531892478. ✓ Use SelectKBest or SelectKPercentile methods

6406531892479. ✘ Drop a feature which has high correlation with target variable

Sub-Section Number : 7

Sub-Section Id : 64065380977

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 288 Question Id : 640653566263 Question Type : MSQ Is Question

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

Correct Marks : 4 Selectable Option : 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 = 6, min_samples_leaf = 4,
                             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 done at node N?

Options :

6406531892518. ❌ Number of samples at node N = 15. If it is split, it will result in 9 nodes in the left child and 6 nodes in the right child.

6406531892519. ✓ Number of samples at node N = 5. If it is split, it will result in 4 nodes in the left child and 2 nodes in the right child.

6406531892520. ✓ Number of samples at node N = 7. If it is split, it will result in 4 nodes in the left child and 3 nodes in the right child.

6406531892521. ✓ Number of samples at node N = 12. If it is split, it will result in 3 nodes in the left child and 9 nodes in the right child.

Question Number : 289 Question Id : 640653566267 Question Type : MSQ Is Question

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

Correct Marks : 4 Selectable Option : 0

Question Label : Multiple Select Question

The code given below attempts to find the clusters in the dataset, X using the K-means algorithm.

```
X,y = make_blobs(n_samples=7,n_features=2,centers=2,random_state=42)
km = KMeans(n_clusters=8,init='random',n_init=1,random_state=42)
km.fit(X)
```

Select the true statements about the code upon execution. Assume necessary imports.
Note: There are no typos in the code, the argument names passed to the function are all correct

Options :

6406531892535. ✘ The code attempts to find 2 clusters in the given dataset

6406531892536. ✓ The dataset X can be visualized in the Euclidean space

6406531892537. ✓ The code raises an error upon execution

6406531892538. ✘ The code gets executed without an error upon execution

Question Number : 290 Question Id : 640653566271 Question Type : MSQ Is Question

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

Correct Marks : 4 Selectable Option : 0

Question Label : Multiple Select Question

The following line of code creates a neural network (assume necessary imports)

```
regr = MLPRegressor(hidden_layer_sizes= (3,5),
                     max_iter=5).fit(X_train, y_train)
```

Select the correct statements from the following list of statements

Options :

6406531892550. ✘ The neural network contains 3 hidden layers with 5 neurons in each hidden layer

6406531892551. ✘ The neural network contains 5 hidden layers with 3 neurons in each hidden layer

6406531892552. ✓ The neural network contains 2 hidden layers with 5 neurons in the second hidden layer

6406531892553. ✓ The neural network contains 2 hidden layers with 3 neurons in the first hidden layer

6406531892554. ✗ None of the given options are correct

Sub-Section Number : 8
Sub-Section Id : 64065380978
Question Shuffling Allowed : Yes
Is Section Default? : null

Question Number : 291 Question Id : 640653566245 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 block:

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

X, y = make_classification(n_samples = 100, n_features = 3,
                           n_informative = 2, n_redundant = 1)

param_grid = [{'max_depth': [2, 3, 4, 5, 6], 'min_samples_split': [2, 3, 4, 5, 6]},
              {'min_samples_leaf': [2, 3, 4, 5, 6]},
              {'min_impurity_decrease': [0.2, 0.3, 0.4, 0.5, 0.6]},
              {'ccp_alpha': [0.1, 0.2, 0.3, 0.4, 0.5, 0.6]}]

gscv = GridSearchCV(DecisionTreeClassifier(), param_grid, cv = 3)
gscv.fit(X, y)

print(gscv.best_params_)
```

How many parameter combinations will be tried by GridSearchCV?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

60

PDSA

Section Id :	64065338408
Section Number :	11
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	29
Number of Questions to be attempted :	29
Section Marks :	100
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 :	64065380979
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Number : 292 Question Id : 640653566272 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 : PROGRAMMING, DATA STRUCTURES AND ALGORITHMS USING PYTHON (COMPUTER BASED EXAM)"