Answers Type : Equal

Text Areas : PlainText

Possible Answers :

2

Question Number : 76 Question Type : SA

Correct Marks : 1.25

Question Label : Short Answer Question

Enter the correct answer for **d**

NOTE: Enter your answer to the nearest integer.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

6

MLT

Number of Questions :	18
Section Marks :	50

Question Number : 77 Question Type : MCQ

Correct Marks : 0

Question Label : Multiple Choice Question

Options :

A. ✔ YES

B. ✖ NO

Question Number : 78 Question Type : SA

Correct Marks : 2

Question Label : Short Answer Question

For the given data set

x	y
1	3
2	4
3	8
4	11

What is the mean squared error if the predicted model is given by $\hat{y} = 2x + 1$? Write your answer correct upto 2 decimal places, do not round up or off.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

1.48 to 1.52

Question Number : 79 Question Type : SA

Correct Marks : 3

Question Label : Short Answer Question

For the given dataset:

x	y
1	3
2	4
3	8

For fitting linear regression model, assume the initial weight vector is $w = [0,0]$ and learning rate $\alpha = 0.01$. Compute the weight vector after one iteration, let's call it w_1. Compute prediction for $x = 5$ correct upto two decimal points with new weight vector (i.e. w_1). Do not round up or off.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

1.88 to 1.92

Question Number : 80 Question Type : SA

Correct Marks : 3

Question Label : Short Answer Question

Ankita had a feature vector x and she wanted to apply polynomial transformation on x. She had written a code snippet as follows:

```
1 import itertools, functools
2 import numpy as np
3 x = np.array([[2, 3]])
4 degree= 3
5 x_t = x.transpose()
6 features = [np.ones(len(x))]
7 for degree in range(1, degree + 1):
8   for items in itertools.combinations_with_replacement(x_t, degree):
9     features.append(functools.reduce(lambda x, y: x * y, items))
10 output = np.sum(np.asarray(features).transpose())
11 print(int(output))
```

What is the output of the code snippet?

NOTE: Enter your answer to the nearest integer.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

90

Question Number : 81 Question Type : SA

Correct Marks : 3

Question Label : Short Answer Question

Suppose a line goes through the origin and is in the direction of the weight vector $w = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$. Given the data point $(10, 0.1)$ calculate the normal distance from the line to the data point. Write your answer correct upto 2 decimal places, do not round up or off.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

0.61 to 0.65

Question Number : 82 Question Type : SA

Correct Marks : 4

Question Label : Short Answer Question

For the given data set

X	y
2, 2	1
0.5, 0.5	-1
1, 1.5	1
-1, -1	-1

A perceptron model $(h_{\mathbf{w}}(\mathbf{x}) = sign(\mathbf{w}^T\mathbf{x}))$ is used to classify data points with parameter vector $\mathbf{w} = [-2, 2, 1]^T$. How many data points are correctly classified?

NOTE: Enter your answer to the nearest integer.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

4

Question Number : 83 Question Type : MCQ

Correct Marks : 2

Question Label : Multiple Choice Question

Consider following label vector for a movie data $\mathbf{D} = (\mathbf{X}, \mathbf{y})$

$$\mathbf{X} = \begin{bmatrix} 2019 & \text{'Uri: The Surgical Strike'} & 8.3 \\ 2009 & \text{'3 Idiots'} & 8.4 \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} \text{'Drama', 'Action', 'War'} \\ \text{'Comedy', 'Drama'} \end{bmatrix}$$

X has year of release, name of the movie and IMDb rating as features. And y has corresponding genres. Predicting genres of a movie is following type of problem:
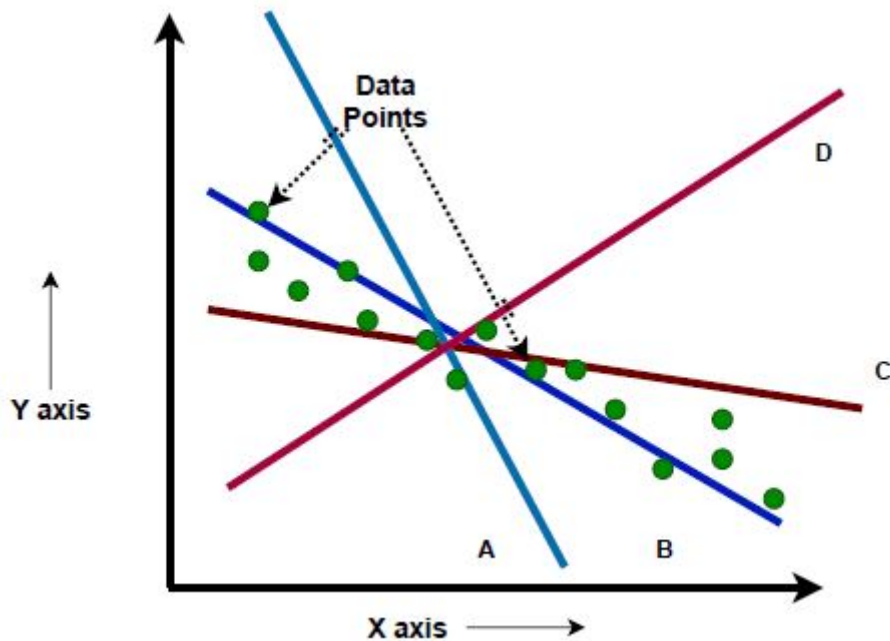
Options :

A. ✔ Multi label multi class classification problem

B. ✖ Single label multi class classification problem

C. ✖ Single label binary class classification problem

D. ✖ Multi label binary class classification problem

Question Number : 84 Question Type : MCQ

Correct Marks : 2

Question Label : Multiple Choice Question

By visually inspecting, which of the following models fits the data best?

Options :

A. ✖ Model A

B. ✔ Model B

C. ✖ Model C

D. ✖ model D

Question Number : 85 Question Type : MCQ

Correct Marks : 2

Question Label : Multiple Choice Question

Which of the following correctly computes the solution of linear regression problem via normal equation method? Assume necessary imports.

Options :

A. ✔
```
1  w = np.linalg.pinv(X)@y
```

B. ✗ ```
1 w = X@X.T@y
```

C. ✗ ```
1 w = X.T@np.linalg.inv(X)@y
```
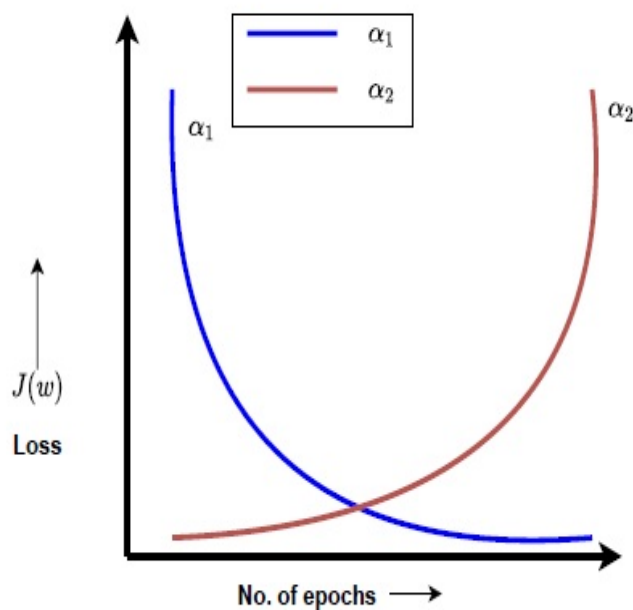
D. ✗ ```
1 w = X.T@np.linalg.pinv(X@y)@y
```

**Question Number : 86 Question Type : MCQ**

**Correct Marks : 2**

Question Label : Multiple Choice Question

Raj trains his linear regression model with two different values of learning rate, i.e. $\alpha_1$ and $\alpha_2$, for the same training set. He gets learning curves as displayed in the following chart.



Which of the following correctly explains relationship between $\alpha_1$ and $\alpha_2$?

**Options :**

A. ✗  $\alpha_1 > \alpha_2$

B. ✔  $\alpha_1 < \alpha_2$

C. ✗  $\alpha_1 = \alpha_2$

D. ✖ Can not be determined

**Question Number : 87 Question Type : MCQ**

**Correct Marks : 2**

Question Label : Multiple Choice Question

Rahul used cross-validation procedure and trained different lasso regression models with various regularization rate $\lambda$ values on some non-linear dataset. The plot of $\lambda$ vs cross-validation errors (RMSE) is given below.

By inspecting the plot, Rahul came to a set of conclusions as mentioned below:

i. The most appropriate value of $\lambda$ is 0.001.

ii. The models with $1e^{-06} \leq \lambda \leq 0.0001$ may also smoothly fit the data.

iii. The models with $0.01 < \lambda \leq 1$ do not fit the data.

What can you say about his conclusions?



$\lambda$ vs Cross validation error

**Options :**

A. ✖ Statement (i) is True. Statements (ii) and (iii) are False

B. ✖ Statement (i) is False. Statements (ii) and (iii) are True

C. ✔ Statements (i), (ii) and (iii) are True

D. ✖ Statements (i), (ii) and (iii) are False

**Question Number : 88 Question Type : MCQ**

**Correct Marks : 2**

Question Label : Multiple Choice Question

Suppose that a company wants to develop a model that recognizes all 26 alphabets in English .The company also requires the model to tell apart whether the recognised alphabet is a lower case or upper case.Therefore, they collected 10K samples that represent all the classes. Assume that the problem is cast as a multi-class classification problem and one-hot encoding is used for the labels. What is the dimension of the label matrix?

**Options :**

A. ✔ $10K \times 52$

B. ✖ $10K \times 2$

C. ✖ $10K \times 26$

D. ✖ $10K \times 54$

**Question Number : 89 Question Type : MCQ**

**Correct Marks : 2**

Question Label : Multiple Choice Question

Consider four datasets represented in following four figures respectively:

For which of the figures above, perceptron algorithm will enventually converge?

**Options :**

A. ✔ Figure 1

B. ✖ Figure 2

C. ✖ Figure 3

D. ✖ Figure 4

**Question Number : 90 Question Type : MSQ**

**Correct Marks : 2**

Question Label : Multiple Select Question

Which of the following can NOT be a linear regression model?

**Options :**

A. ✖ $y = \sum_{i=0}^{m} w_i x_i$

B. ✔ $y = \prod_{i=0}^{m} w_i x_i$

C. ✔ $y = \sum_{i=0}^{m} w_i^2 x_i$

D. ✔ $y^2 = \sum_{i=0}^{m} w_i^2 x_i$

**Question Type : COMPREHENSION**

**Question Numbers : (91 to 93)**

Question Label : Comprehension

Consider the following confusion matrix

|  |  | Predicted | |
|---|---|---|---|
|  |  | False | True |
| Actual | False | 167 | 153 |
|  | True | 85 | 95 |

Write your answer correct upto 3 decimal places, do not round up or off.

(Do not write your answer as percentage)

Based on the above data, answer the given subquestions.

**Sub questions**

**Question Number : 91 Question Type : SA**

**Correct Marks : 1**

Question Label : Short Answer Question

What is the precision metric?

**Response Type :** Numeric

**Evaluation Required For SA :** Yes

**Show Word Count :** Yes

**Answers Type :** Range

**Text Areas :** PlainText

**Possible Answers :**

0.380 to 0.386

**Question Number : 92 Question Type : SA**

**Correct Marks : 1**

Question Label : Short Answer Question

What is the recall metric?

**Response Type :** Numeric

**Evaluation Required For SA :** Yes

**Show Word Count :** Yes

**Answers Type :** Range

**Text Areas :** PlainText

**Possible Answers :**

0.524 to 0.530

**Question Number : 93 Question Type : SA**

**Correct Marks : 1**

Question Label : Short Answer Question

What is the F-1 Score metric?

**Response Type :** Numeric

**Evaluation Required For SA :** Yes

**Show Word Count :** Yes

**Answers Type :** Range

**Text Areas :** PlainText

**Possible Answers :**

0.440 to 0.447

**Question Type : COMPREHENSION**

**Question Numbers : (94 to 95)**

Question Label : Comprehension

Consider the following code snippet and answer the given subquestions.

```
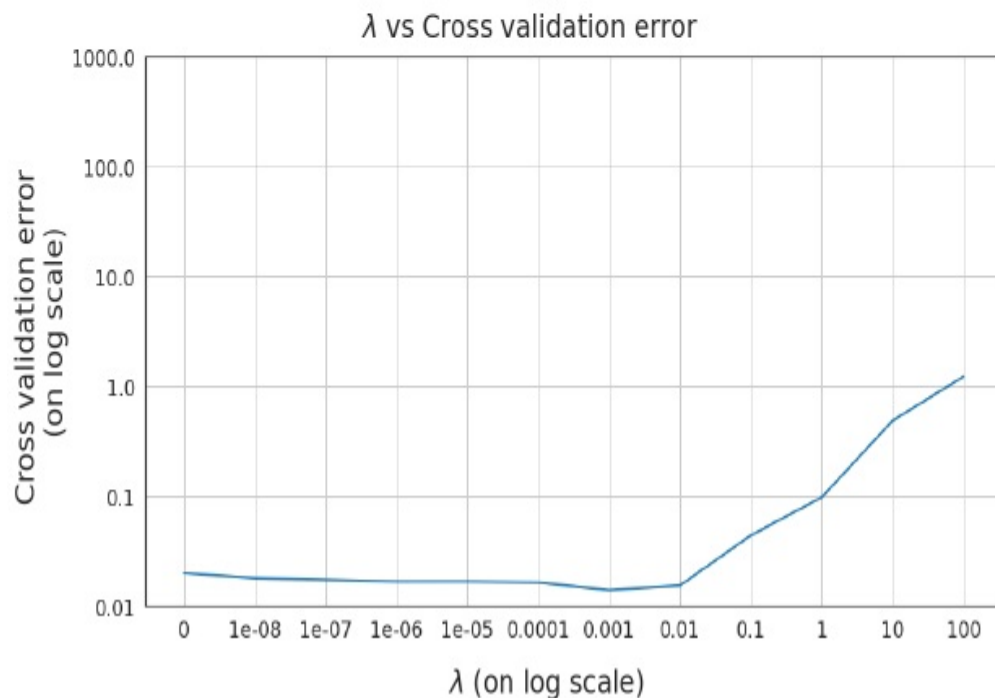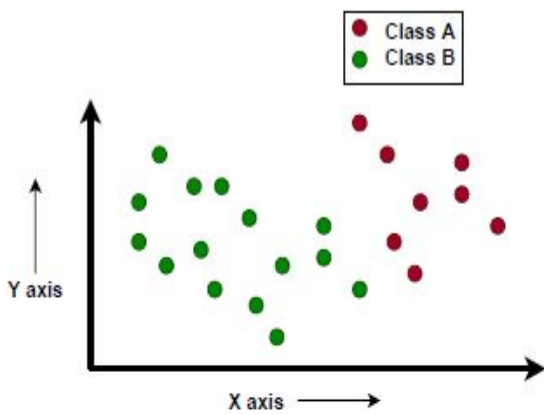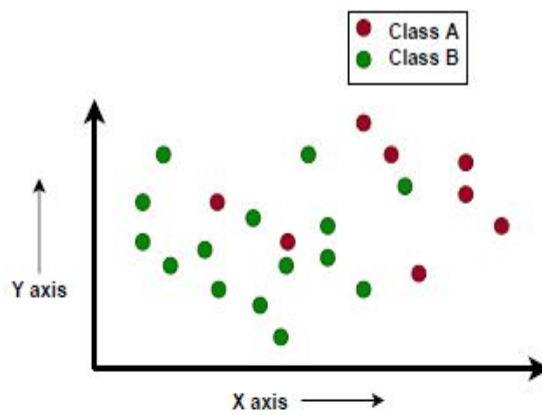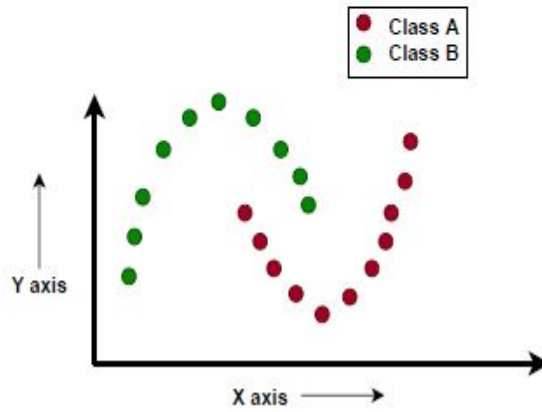1  import numpy as np
2  a=[[1,2,0],[0,2,1],[0,2,1]]
3  b=[[1,1,0],[1,2,3],[0,1,3]]
4  A=np.array(a)
5  B=np.array(b)
6  C=A@B
7  D=A*B
```

Sub questions

Question Number : 94 Question Type : SA

Correct Marks : 2

Question Label : Short Answer Question

What is the output of following code snippet?

```
1  print(C[1][0]+D[1][2])
```

NOTE: Enter your answer to the nearest integer.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

5

Question Number : 95 Question Type : SA

Correct Marks : 2

Question Label : Short Answer Question

What is the output of following code snippet?

```
1  print(C.sum().sum())
```

NOTE: Enter your answer to the nearest integer.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

46

Question Type : COMPREHENSION

Question Numbers : (96 to 97)

Question Label : Comprehension

Raj has a dataset that has a non-linear relationship between the features and the label. The dimension of training samples is 4×2 (excluding the additional all-ones feature for the bias term, which you should add), label vector is 4 × 1 and weight vector is 3 × 1.

X	y
[1 , 1]	1
[2 , 2]	1
[3 , 2]	2
[4 , 3]	2

w
0.1
0.2
0.3

Based on the above data, answer the given subquestions.

Sub questions

Question Number : 96 Question Type : MCQ

Correct Marks : 5

Question Label : Multiple Choice Question

Suppose, after fitting the linear regression model, Raj has computed the ridge loss value for this data as 0.3507, which of the following is the regularization parameter value chosen by Raj for performing ridge regression?

Options :

A. ✖ 0.001

B. ✖ 0.1

C. ✔ 0.01

D. ✖ 0.5

Question Number : 97 Question Type : MCQ

Correct Marks : 5

Question Label : Multiple Choice Question

For the same dataset, Suppose, after fitting the linear regression model, Raj has computed the lasso loss value as 0.38, which of the following is the regularization parameter value chosen by Raj for performing lasso regression?

Options :

A. ✖ 0.001

B. ✔ 0.1

C. ✖ 0.01

D. ✖ 0.5

Question Type : COMPREHENSION

Question Numbers : (98 to 99)

Question Label : Comprehension

Suppose that you have a non-linear dataset with a feature matrix $X_{100\times1}$. You used a 4th degree polynomial transformation to the features and trained a linear regression model with stochastic

gradient descent algorithm on the transformed features. During the training phase, you have fortunately obtained training and validation errors as zero (that is, model perfectly fits the data).

Based on the above data, answer the given subquestions.

Sub questions

Question Number : 98 Question Type : MCQ

Correct Marks : 1

Question Label : Multiple Choice Question

What will happen when you try to fit degree 8 polynomial on the same dataset?

Options :

A. ✔ Chances for overfitting is high

B. ✖ Chances for underfitting is high

C. ✖ No change in error values

D. ✖ None of these

Question Number : 99 Question Type : MCQ

Correct Marks : 1

Question Label : Multiple Choice Question

What will happen when you try to fit degree 2 polynomial on the same dataset?

Options :

A. ✖ Chances for overfitting is high

B. ✔ Chances for underfitting is high

C. ✖ No change in error values

D. ✖ None of these

Java

Number of Questions : 16

Section Marks : 50

Business Analytics

Number of Questions :	8
Section Marks :	50

Question Number : 172 Question Type : MCQ

Correct Marks : 0

Question Label : Multiple Choice Question

THIS IS QUESTION PAPER FOR THE SUBJECT "DIPLOMA LEVEL: BUSINESS ANALYTICS"

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 <u>TOP</u> FOR THE SUBJECTS REGISTERED BY YOU)

Options :

A. ✔ Yes

B. ✖ No

Question Type : COMPREHENSION

Question Numbers : (173 to 174)

Question Label : Comprehension

The linear demand response for a product-A is modelled as a simple linear regression represented as D(P) = 1500 – 20*P, where D(P) is the demand at price-P. Then, answer the given sub questions.

Sub questions

Question Number : 173 Question Type : SA

Correct Marks : 5

Question Label : Short Answer Question

What is the elasticity of this curve when price is Rs.50?

NOTE: Enter your answer to the nearest integer.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

2

Question Number : 174 Question Type : SA

Correct Marks : 2.5

Question Label : Short Answer Question

What is the satiating price for this curve?

NOTE: Enter your answer to the nearest integer.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

750

Question Number : 175 Question Type : MSQ

Correct Marks : 2.5

Question Label : Multiple Select Question

For a demand response curve which has constant elasticity, which of the following statements are true? (choose all that are applicable)

Options :

A. ✱ If the curve is for an inelastic product, the revenue is increased only by setting price close to zero

B. ✔ If the curve is for an inelastic product, the revenue is increased by simply increasing the prices

C. ✱ If the curve is for an elastic product, the revenue is increased by simply increasing the prices

D. ✔ If the curve is for an elastic product, the revenue is increased only by setting price close to zero

Question Number : 176 Question Type : MSQ

Correct Marks : 2.5

Question Label : Multiple Select Question

Which of the following distributions is/are not symmetric in nature (select all that are applicable)?

Options :

A. ✱ Standard normal distribution

B. ✱ Standard bernoulli distribution

C. ✱ Uniform distribution between [-1 to +1]

D. ✔ Poisson distribution

Question Number : 177 Question Type : MCQ

Correct Marks : 2.5

Question Label : Multiple Choice Question

The given figure communicates which message the best?

Options :

A. ✵ Message: "Mileages trend for different Models"

B. ✔ Message: "Relationship between Car Weight and Mileage"

C. ✵ Message: "Relationship between Models and their Car Weights"

D. ✵ Message: "Car Model wise the minimum and maximum Mileage and Car Weights"

E. ✵ Message: "Model wise relationship between Car Weight and Mileage"

F. ✵ Message: "Car weight wise Mileage trend for different Models"

G. ✵ Message: "Mileage wise relationship between Models and their Car Weights"

Question Type : COMPREHENSION

Question Numbers : (178 to 179)

Question Label : Comprehension

Given the below table, answer the subquestions.

Possible number of defects in an item	Actual number of items found which had the specified number of defects	Theoretically expected number of items found which had the specified number of defects
1	12	18
2	22	20
3	16	20
4+	32	12

Sub questions

Question Number : 178 Question Type : SA

Correct Marks : 7.5

Question Label : Short Answer Question

Considering a lot size of 100, what is value of test statistic (chi-square goodness of fit) that the table mentioned in main question follows a poisson distribution (Round to one decimal place)?

$$\{Hint: Chi\text{-}square = \sum_k \frac{(observed_k - Expected_k)^2}{Expected_k}\}$$

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

41.0 to 41.2

Question Number : 179 Question Type : MSQ

Correct Marks : 2.5

Question Label : Multiple Select Question

What is the null hypothesis that the test statistic mentioned in previous question is computed for? (choose all that is applicable)

Options :

A. ✳ Null: The sampling distribution is a poisson distribution

B. ✔ Null: The population distribution is a poisson distribution

C. ✳ Null: The sampling distribution is not a poisson distribution

D. ✖ Null: The population distribution is not a poisson distribution

Question Type : COMPREHENSION

Question Numbers : (180 to 182)

Question Label : Comprehension

You are given the following contingency table based on a sample data where different cities and their brand preferences are provided. Perform a chi-squared test of independence to make inferences about the population from this sample and answer the given subquestions.

Note: Round off to two decimal values at all the intermediate iterations and the final step.

	Brand A	Brand B	Total
Chennai	288	124	412
Mumbai	622	204	826
Total	910	328	1238

Sub questions

Question Number : 180 Question Type : SA

Correct Marks : 5

Question Label : Short Answer Question

From the given contingency table, find the expected frequency of Chennai people preferring brand B?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

109 to 110

Question Number : 181 Question Type : SA

Correct Marks : 5

Question Label : Short Answer Question

What is the calculated value of chi-squared?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

3.9 to 4.4

Question Number : 182 Question Type : MCQ

Correct Marks : 5

Question Label : Multiple Choice Question

At the significance level 0.05, chi-squared tabular value is 3.84. What do you conclude?

Options :

A. ✔ Reject the null hypothesis and conclude that the categorical variables are not independent

B. ✖ Fail to reject the null hypothesis and conclude that the categorical variables are not independent

C. ✖ Fail to reject the null hypothesis and conclude that the categorical variables are independent

D. ✖ Reject the null hypothesis and conclude that the categorical variables are independent

Question Number : 183 Question Type : SA

Correct Marks : 10

Question Label : Short Answer Question

Suppose a factory manufactures products on two machines A and B. Suppose 65% of total output comes from machine A, 30% of total output comes from machine B and 5% of total output comes from machine C. From the past data, it is known that 1% of products by machine A are defectives, 2% of products by machine B are defectives and 10% of products by machine C are defectives. What is the probability that the product has come from machine B given that it is a defective?

NOTE: Enter your answer in two decimal places.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

0.32 to 0.36

MLP

Number of Questions :	21
Section Marks :	50

Question Number : 184 Question Type : MCQ

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 :

A. ✔ Yes

B. ✖ No

Question Number : 185 Question Type : MCQ

Correct Marks : 1

Question Label : Multiple Choice Question

Which of the following is/are NOT a metric for regression?

Options :

A. ✹ Mean Absolute Error

B. ✹ R^2 score

C. ✔ $F1$ score

D. ✹ Mean Squared Error

Question Number : 186 Question Type : MCQ

Correct Marks : 1

Question Label : Multiple Choice Question

Consider following code:

```
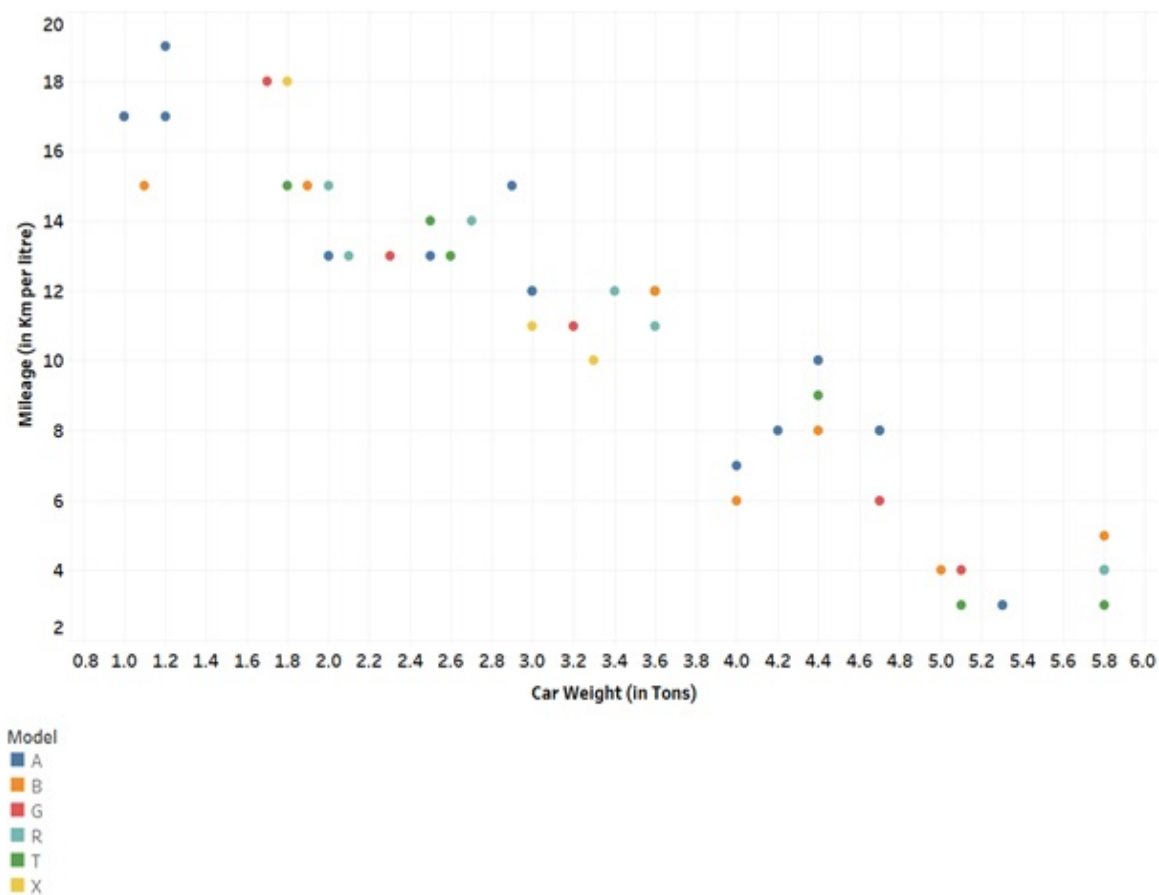from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LinearRegression
from sklearn.pipeline import Pipeline
pipe = Pipeline([('feature_scaling', StandardScaler()),
                 ('regressor', LinearRegression())])
```

Which of the following commands can be used to print the number of steps in a pipeline?

Options :

A. ✹ ((print step_pipe))

B. ✹ print(steps(pipe.len))

C. ✔ print(len(pipe.named_steps))

D. ✹ print(len[pipe_steps])

Question Number : 187 Question Type : MCQ

Correct Marks : 2

Question Label : Multiple Choice Question

Consider the following code:

```
from sklearn.datasets import load_diabetes
data = load_diabetes()
```

The feature matrix containing the samples will be present in:

Options :

A. ✖ data.feature_matrix

B. ✖ data.features

C. ✖ data.matrix

D. ✔ data.data

Question Number : 188 Question Type : MCQ

Correct Marks : 2

Question Label : Multiple Choice Question

Consider following code snippet:

```
from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split
X, y = load_iris(return_X_y = True)
print(X.shape)
print(y.shape)
train_X, test_X, train_y, test_y = train_test_split(X, y,
                                                    test_size=0.2,
                                                    random_state=42)
```

The output of the above code is the following:

```
(150, 4)
(150,)
```

What will be the output of the following code?

```
print(train_X.shape)
print(test_X.shape)
print(train_y.shape)
print(test_y.shape)
```

A. ✶ (130, 3) (20, 1) (130, 3) (20, 1)

B. ✔ (120, 4) (30, 4) (120,) (30,)

C. ✶ (130, 4) (20, 4) (130,) (20,)

D. ✶ (120, 3) (30, 1) (120, 3) (30, 1)

Question Number : 189 Question Type : MCQ

Correct Marks : 2

Question Label : Multiple Choice Question

What is the output of the following code?

```
import numpy as np
from sklearn.preprocessing import FunctionTransformer
transformer = FunctionTransformer(np.exp)
X = np.array([[1,2],
              [3,4]])
transformer.transform(X)
```

Options :

A. ✔
```
array([[ 2.71828183,  7.3890561 ],
       [20.08553692, 54.59815003]])
```

B. ✶
```
array([[ 1.71828183,  6.3890561 ],
       [19.08553692, 53.59815003]])
```

C. ✶
```
array([[ 147.4131591 ,  402.42879349],
       [1095.63315843, 2979.95798704]])
```

D. ✶
```
array([[0.69314718, 1.09861229],
       [1.38629436, 1.60943791]])
```

Question Number : 190 Question Type : MCQ

Correct Marks : 2

Question Label : Multiple Choice Question

Based on the following statements, select the correct option.

I. For a very large dataset, it is better to employ LinearRegressor over SGDRegressor.

II. SGDRegressor solves the normal equation to calculate the value of unknown parameters.

Options :

A. ✖ Statement I is true, Statement II is true, Statement II is a correct explanation for statement I.

B. ✖ Statement I is true, Statement II is true, Statement II is not a correct explanation for statement I.

C. ✖ Statement I is true, Statement II is false.

D. ✔ Statement I and II are false.

Question Number : 191 Question Type : MCQ

Correct Marks : 2

Question Label : Multiple Choice Question

Consider the following 6 lines of code:

```
from sklearn.linear_model import SGDRegressor
from sklearn.model_selection import GridSearchCV
estimator = SGDRegressor(learning_rate='constant').fit(X,y)
param_grid = [{'eta0':[0.1,0.001]}]
gs = GridSearchCV(estimator,param_grid)
gs.fit(X,y)
```

Assuming that the input dataset is loaded to the variable (X, y). The shapes of X and y are $(100, 10)$ and $(100, 1)$, respectively. Does the code raise an error upon execution? If yes, which line of the code raises the error first?

Options :

A. ✖ Error in the third line of code

B. ✖ Error in the fourth line of code

C. ✖ Error in the fifth line of code

D. ✖ Error in the sixth line of code

E. ✔ It won't raise any error

Question Number : 192 Question Type : MCQ

Correct Marks : 2

Question Label : Multiple Choice Question

Consider a grid search for various learning rate values in SGDRegressor

```
from sklearn.linear_model import SGDRegressor
from sklearn.model_selection import GridSearchCV
param_grid = [{'eta0':[0.1,0.2,0.3,0.4,0.5]}]
gs = GridSearchCV(SGDRegressor(),param_grid,cv=5)
gs.fit(X,y)
```

Upon execution of GridSearchCV(), how many times the fit() method of the estimator would have been called internally?

Options :

A. ✖ 1

B. ✔ 25

C. ✖ 5

D. ✖ 125

Question Number : 193 Question Type : MCQ

Correct Marks : 3

Question Label : Multiple Choice Question

Fill the name of the missing attribute in the following code:

```
from sklearn.datasets import make_blobs
import matplotlib.pyplot as plt
X, y = make_blobs(n_samples = 100, n_features=2, _____=2)
plt.scatter(X[:, 0], X[:, 1], marker="o", c=y, s=25, edgecolor="k")
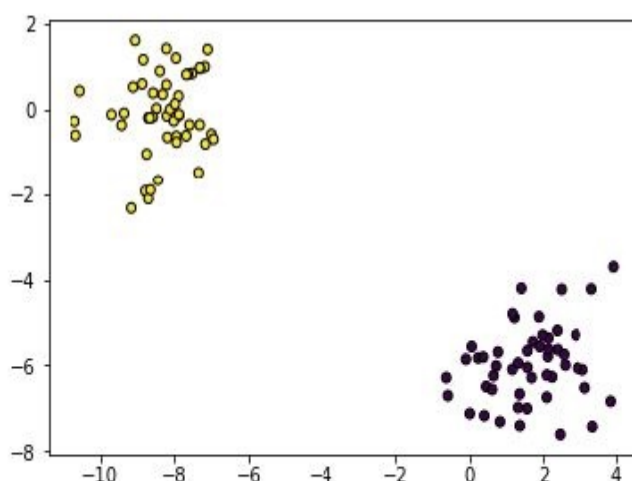```

The output plot is shown in Figure 1:



Figure 1

**Options :**

A. �ö clusters

B. ✖ labels

C. ✔ centers

D. ✖ classes

**Question Number : 194 Question Type : MCQ**

**Correct Marks : 3**

Question Label : Multiple Choice Question

What is the output of the following code?

```python
import numpy as np
from sklearn.impute import SimpleImputer
imp_mean = SimpleImputer(missing_values=np.nan, strategy='mean')
X=imp_mean.fit_transform([[42, 22, 19],
                          [np.nan, np.nan, 96],
                          [1200, 5, np.nan],
                          [42, np.nan, 92]])
print(X)
```

Options :

A. ✔

$$\begin{bmatrix} 42. & 22. & 19. \\ 428. & 13.5 & 96. \\ 1200. & 5. & 69. \\ 42. & 13.5 & 92. \end{bmatrix}$$

B. ✖

$$\begin{bmatrix} 42 & 22 & 19 \\ 2 & 3 & 96 \\ 10 & 100 & 69 \\ 42 & 13.5 & 92 \end{bmatrix}$$

C. ✖

$$\begin{bmatrix} 42 & 22 & 19 \\ 3 & np.nan & np.nan \\ 10 & 100 & 92 \\ 1200 & 5 & 42 \end{bmatrix}$$

D. ✖

$$\begin{bmatrix} 42. & 13.5 & 91. \\ 428. & 22 & 96. \\ 1200. & 13.5. & 96. \\ 42. & 236 & 92. \end{bmatrix}$$

Question Number : 195 Question Type : MCQ

Correct Marks : 4

Question Label : Multiple Choice Question

Consider the following code and select the correct option(s).

```
from sklearn.linear_model import SGDRegressor
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import StandardScaler
sgd = Pipeline([('feature_scaling', StandardScaler()),
                ('sgd_regressor', SGDRegressor())])

sgd.fit(X, y)
```

where X and y are training data.

Options :

A. ✴ It transforms the original feature vector X into a new feature vector so that all values fall within range [0, 1].

B. ✴ It transforms the original feature vector X into a new feature vector so that all values fall within range [−1, 1].

C. ✴ It will fit the linear regression model on the transformed data using normal equation.

D. ✔ It will fit the linear regression model on the transformed data using iterative optimization.

Question Number : 196 Question Type : MSQ

Correct Marks : 2

Question Label : Multiple Select Question

Which of the following list of functions are used for hyper-parameter tuning in sklearn?

Options :

A. ✔ GridSearchCV()

B. ✔ RandomizedSearchCV()

C. ✳ MinMaxScaler()

D. ✳ StandardScaler()

Question Number : 197 Question Type : MSQ

Correct Marks : 2

Question Label : Multiple Select Question

Which of the following statements are true about hyper-parameter tuning?

Options :

A. ✳ It is only applicable for applications that require gradient calculations.

B. ✳ The tuning is done by using a portion of the test data set.

C. ✔ It is used to select the best model's various settings of hyper-parameters.

D. ✔ More than one hyper-parameter of a model can be tuned simultaneously in sklearn.

Question Number : 198 Question Type : MSQ

Correct Marks : 3

Question Label : Multiple Select Question

Which of the following statements is/are correct?

Options :

A. ✔ Oversampling refers to replicating some points from the minority class in order to increase the cardinality of the minority class.

B. ✔ Oversampling may consist of either replicating or generating synthetic data for the minority class.

C. ✳ SMOTE is a popular technique for undersampling.

D. ✔ FeatureUnion and Pipeline can be used to create complex transformers.

Question Number : 199 Question Type : MSQ

Correct Marks : 3

Question Label : Multiple Select Question

Consider the following box plots of features of iris data and select the correct options.

Box Plot
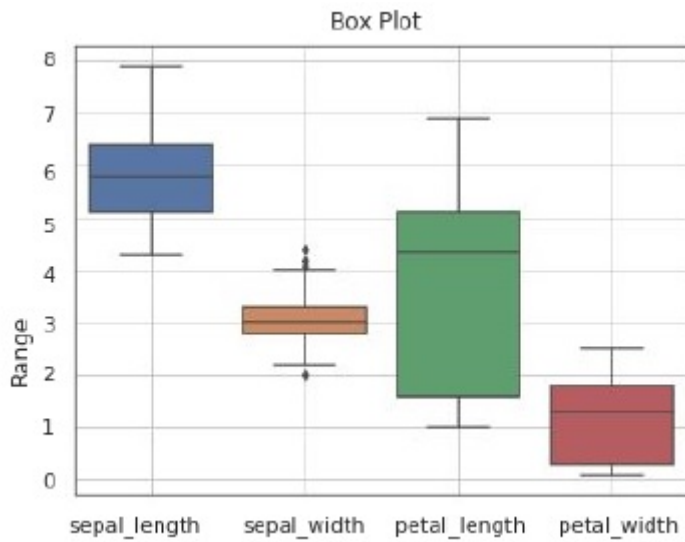
Options :

A. ✔ Petal length of data is left-skewed

B. ✖ Petal length of data is right-skewed

C. ✔ Median of petal width lies between 1 and 2

D. ✖ Median of petal length lies between 1 and 2

Question Number : 200 Question Type : MSQ

Correct Marks : 3

Question Label : Multiple Select Question

Select the correct options regarding SGDRegressor:

Options :

A. ✖ With every SGD iteration, the loss steadily decreases.

B. ✔ More iterations require more computational time.

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

D. ✔ Training error might actually increase while performing SGD iterations.

Question Number : 201 Question Type : MSQ

Correct Marks : 3

Question Label : Multiple Select Question

Consider the following code and select the correct options.

```python
from sklearn.linear_model import SGDRegressor
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import StandardScaler
estimator = SGDRegressor(penalty='elasticnet',
                         l1_ratio=0.2,
                         learning_rate='invscaling')
sgd = Pipeline([('feature_scaling', StandardScaler()),
                ('sgd_regressor', estimator )])

sgd.fit(X, y)
```

Options :

A. ✔ Regularization rate for L2 penalty will be 0.8

B. ✖ Regularization rate for L2 penalty will be 0.2

C. ✔ Learning rate decreases with every iteration

D. ✖ Learning rate increases with every iteration

Question Number : 202 Question Type : MSQ

Correct Marks : 3

Question Label : Multiple Select Question

Consider the following code:

```python
from sklearn.linear_model import SGDRegressor
SGDRegressor = SGDRegressor(loss='squared_error',
                            early_stopping=True,
                            max_iter=500,
                            tol=1e-3,
                            validation_fraction=0.2,
                            n_iter_no_change=5)
```

When is SGDRegressor likely to stop?

Options :

A. ✔ validation score does not improve by at least 10^{-3} for 5 consecutive epochs.

B. ✖ validation score does not improve by at least 10^{-3} for the first epoch.

C. ✔ after at most 500 iterations.

D. ✖ after exactly 5 iterations.

Question Number : 203 Question Type : MSQ

Correct Marks : 3

Question Label : Multiple Select Question

Consider the following code and select the correct options. Assume that all the necessary libraries are imported.

```python
from sklearn.preprocessing import PolynomialFeatures
from sklearn.linear_model import SGDRegressor
from sklearn.pipeline import Pipeline
poly_transform = PolynomialFeatures(degree=2,
                                    interaction_only= True)
estimator = SGDRegressor(penalty='elasticnet',
                         l1_ratio=0.3)
poly_model = Pipeline([('polynomial_transform', poly_transform),
                       ('elasticnet', estimator)])
poly_model.fit(X, y)
```

Options :

A. ✔ It will first transform the features into degree 2 such that power of 2 and higher of the same input feature are excluded.

B. ✖ It will first transform the features into degree 2 such that only power of 2 the same input feature are included.

C. ✖ It will not perform any regularization.

D. ✔ It will apply both lasso and ridge penalties to the model.

Question Number : 204 Question Type : MSQ

Correct Marks : 4

Question Label : Multiple Select Question

Consider the following code and select the correct options.

```python
from sklearn.datasets import make_classification
X, y = make_classification(n_features=5,
                           n_informative=3,
                           n_redundant = 1,
                           n_classes=2,
                           n_clusters_per_class=1,
                           random_state=10,
                           n_samples=50)
```

Options :

A. ✖ The shapes of X and y are $(5, 2)$ and $(5,)$, respectively.

B. ✔ The shapes of X and y are $(50, 5)$ and $(50,)$, respectively.

C. ✔ X has one feature such that it can be written as linear combination of other features.

D. ✖ X has three features such that they can be written as linear combination of other features.