

Linear Regression Intuition

MCQ Questions:

1.

Error Measure

[Send Feedback](#)

Which of the following is common error measure ?

Options

Attempts left: **1/2**

This problem has only one correct answer

- ☐ Sensitivity
- ☐ Median absolute deviation
- ☐ Specificity
- ☒ All of the Mentioned
- ☒ Hurray! Correct Answer

Solution Description

Sensitivity and specificity are statistical measures of the performance of a binary classification test, also known in statistics as classification function.

2.

Linear Regression

[Send Feedback](#)

A process by which we estimate the value of dependent variable on the basis of one or more independent variables is called.

Options

Attempts left: **1/2**

This problem has only one correct answer

- ☐ Correlation
- ☒ Regression
- ☐ Residual
- ☐ Slope
- ☒ Hurray! Correct Answer

Solution Description

Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

3.

Linear Regression

[Send Feedback](#)

The method of least squares calculates square of deviations of the points from the line chosen using Linear Regression. Our target is that this error should be -

Options

Attempts left: **1/2**

This problem has only one correct answer

- ☐ Maximum
- ☒ Minimum
- ☐ Zero
- ☐ Positive
- ☒ Hurray! Correct Answer

Solution Description

The method of least squares calculates the square of deviations of the points from the line chosen using Linear Regression. Our target is that this error should be minimum.

4.

Linear Regression
[Send Feedback](#)

All data points falling along a straight line is called:

Options
Attempts left: 1/2
This problem has only one correct answer
☒ Linear relationship
☐ Non linear relationship
☐ Residual
☐ Scatter diagram
☒ Hurray! Correct Answer

Solution Description
All data points falling along a straight line is called Linear relationship.

5.

Linear Regression
[Send Feedback](#)

In simple linear regression, the numbers of unknown constants are:

Options
Attempts left: 1/2
This problem has only one correct answer
☐ One
☒ Two
☐ Three
☐ Four
☒ Hurray! Correct Answer

Solution Description
In simple linear regression, the numbers of unknown constants are two.

The equation for simple linear regression is :
$$y = mx + c$$

where m and c are unknown constants.

6.

Linear Regression
[Send Feedback](#)

In simple regression equation, the numbers of variables involved are:

Options
Attempts left: 1/2
This problem has only one correct answer
☐ 0
☐ 1
☒ 2
☐ 3
☒ Hurray! Correct Answer

Solution Description
In simple linear regression, the numbers of variables involved are two.

The equation for simple linear regression is :
$$y = mx + c$$

where y and x are variables.

7.

Linear Regression

[Send Feedback](#)

If the value of any regression coefficient is zero, then two variables are:

Attempts left: **1/2**

Options

This problem has only one correct answer

- ☐ Qualitative
- ☐ Correlation
- ☐ Dependent
- ☒ Independent
- ☒ Hurray! Correct Answer

Solution Description

If the value of any regression coefficient is zero, then two variables are independent.

8.

Linear Regression

[Send Feedback](#)

In Which of the following case , A Straight line slope will be upward given linear equation is $Y = a + bX$.

Attempts left: **1/2**

Options

This problem has only one correct answer

- ☐ $b = 0$
- ☐ $b < 0$
- ☒ $b > 0$
- ☐ $b \neq 0$
- ☒ Hurray! Correct Answer

Solution Description

A Straight line slope will be upward when the value of $b > 0$.

9.

Linear Regression

[Send Feedback](#)

The straight line graph of the linear equation $Y = a + bX$, slope will be downward If:

Attempts left: **1/2**

Options

This problem has only one correct answer

- ☐ $b > 0$
- ☒ $b < 0$
- ☐ $b = 0$
- ☐ $b \neq 0$
- ☒ Hurray! Correct Answer

Solution Description

The slope will be downward when $b < 0$.

Analysis of LR using dummy data

MCQ Questions:

1.

Numpy loadtxt Function

[Send Feedback](#)

What all statements are correct about numpy loadtxt() function:

Options

This problem may have one or more correct answers

- ☒ file path is input to this function✓
- ☒ It loads data from text file✓
- ☒ By-default delimiter is space✓
- ☐ By-default delimiter is comma

✓ Hurray! Correct Answer

Solution Description

1. file path is input to this function.
2. It loads data from the text file
3. By-default delimiter is space

Attempts left: **1/2**

2.

Is Statement Correct

[Send Feedback](#)

Consider below code:

```
import numpy as np
data = np.loadtxt("data.csv",delimiter=",")
X = data[:,0]
Y = data[:,1]
```

Statement: X has all the rows of 0th Col and Y has all the rows of 1st Col

Options

This problem has only one correct answer

- ☒ Statement is Correct
- ☐ Statement is Incorrect
- ☐ It will produce error

✓ Hurray! Correct Answer

Solution Description

With this line, X = data[:,0]
X will have all the rows of 0th Col

With this line, Y = data[:,1]
Y will have all the rows of 1st Col

That's why the statement is true

Attempts left: **1/2**

3.

Numpy Shape Function

[Send Feedback](#)

Consider below code:

```
import numpy as np
X = np.array([1,2,3,4])
X.shape
```

What will be the output ?

Options

This problem has only one correct answer

- ☒ (4,)
- ☐ (.4)
- ☐ (4,4)

✓ Hurray! Correct Answer

Solution Description

np.array([1,2,3,4]) is a one dimensional array. The shape of this is (4,).

Attempts left: **1/2**

Read : Using Random State Notes

Question: 1

Find coefficient and intercept

Problem

Result

Find coefficient and intercept

[Send Feedback](#)

For the given dataset "FuelEconomy.csv"

Create a Linear Regressor and fit the dataset in this. After that, print the coefficient and intercept.

Note:

1. Inside the function "train_test_split", use the "random_state = 42".
2. Split the dataset in the ratio of 70:30 into the training and testing datasets.

To know more about randomstate please read the document given regarding randomstate.

Output

Print the coefficient and intercept in separate lines rounded off to 2 decimal places.

```
1 ## Open and read data file as specified in the question
2 ## Print the required output in given format
3 import numpy as np
4
5 data = np.loadtxt("FuelEconomy.csv", delimiter = ",")
6
7 x = data[:,0].reshape(-1,1)
8 y = data[:,1].reshape(-1,1)
9
10 from sklearn import model_selection
11
12 x_train, x_test, y_train, y_test = model_selection.train_test_split(x,y, test_size = 0.30)
13
14 from sklearn.linear_model import LinearRegression
15
16 alg1 = LinearRegression()
17
18 alg1.fit(x_train, y_train)
19
20
21
22 coefficient = np.round(alg1.coef_[0], 2)
23 intercept = np.round(alg1.intercept_, 2)
24
25 print(coefficient[0])
26 print(intercept[0])
```

Coefficient of Determination

Question: 2

Find the testing and training score

Problem

Result

Run Output

0.913
0.906

Submissions

You will get score only if your output matches with the correct answer.

Select Submission
Submission 1

Yay, Correct answer [View submission](#)

Score 50.0% **Submitted:** Sep 12, 2023, 11:49 AM

Penalty: 50%
Deadline crossed on Jul 24, 2023, 1:56 PM

Output
0.913
0.906

0.988 s

```
1 ## Open and read data file as specified in the question
2 ## Print the required output in given format
3 import numpy as np
4
5 data = np.loadtxt("FuelEconomy.csv", delimiter = ",")
6
7 x = data[:,0].reshape(-1,1)
8 y = data[:,1].reshape(-1,1)
9
10 from sklearn import model_selection
11
12 x_train, x_test, y_train, y_test = model_selection.train_test_split(x,y, test_size = 0.30, random_state=42)
13
14 from sklearn.linear_model import LinearRegression
15
16 alg1 = LinearRegression()
17
18 alg1.fit(x_train, y_train)
19
20
21 score_test = alg1.score(x_test, y_test)
22 score_train = alg1.score(x_train, y_train)
23
24 score_test = np.round(score_test, 3)
25 score_train = np.round(score_train, 3)
26
27 print(score_test)
28 print(score_train)
29
```

Cost Function

MCQ Questions:

1.

Objective Way to Define Best Fit Line

[Send Feedback](#)

What is the best objective way to define the best fit line ?

Options

Attempts left: **1/2**

This problem has only one correct answer

- ☒ Minimization of Error function
- ☐ Minimization of Sum function
- ☐ Minimization of Multiplication Function
- ☒ Hurray! Correct Answer

Solution Description

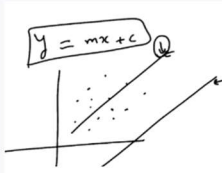
The best objective way to define the best fit line is to the Minimization of the Error function.

2.

Error in Line

[Send Feedback](#)

Which line has more error as taught in video:



Options

Attempts left: **1/2**

This problem has only one correct answer

- ☐ Upper Line
- ☒ Lower Line
- ☒ Hurray! Correct Answer

Solution Description

The lower line has more error, as most of the data points are far from the lower line than from the upper line.

3.

Mean Squared Error

[Send Feedback](#)

For simple linear equation $y=mx+b$ we can calculate Mean Square Error as:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - (mx_i + b))^2$$

Where

- N is the total number of observations (data points)
- $\frac{1}{N} \sum_{i=1}^N$ is the mean
- y_i is the actual value of an observation and $mx_i + b$ is our prediction

Options

Attempts left: **1/2**

This problem has only one correct answer

- ☒ True
- ☐ False
- ☒ Hurray! Correct Answer

Solution Description

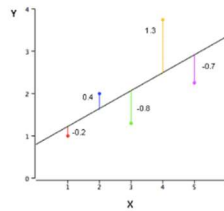
The formula described here is for Mean Square Error.

4.

Sum of Squared Error

[Send Feedback](#)

The graph below represents a regression line predicting Y from X. The values on the graph shows the residuals for each predictions value. Use this information to compute the Sum of Squared Error.



Options

Attempts left: 0/2

This problem has only one correct answer

- ☒ 3.02
- ☐ 0.75
- ☐ 1.01
- ☐ None of these
- ☒ Hurray! Correct Answer

Solution Description

SSE is the sum of the squared errors of prediction, so $SSE = (-.2)^2 + (.4)^2 + (-.8)^2 + (1.3)^2 + (-.7)^2 = 3.02$

5.

M and C of Linear Regression

[Send Feedback](#)

Choose the correct equation for m and c once you derive m and c of linear regression:

Equation-1:

$$y.mean() - m * x.mean()$$

Equation-2:

$$\frac{(x*y).mean() - x.mean()*y.mean()}{(x*x).mean() - x.mean()*x.mean()}$$

Note: mean() function in above equation is used to calculate the mean of value, y.mean() will give you the mean of y(y being 1D array) and similarly x.mean() will give you mean of x(x being 1D array)

Options

Attempts left: 1/2

This problem has only one correct answer

- ☒ Equation 1 represent c and Equation 2 represent m
- ☐ Equation 1 represent m and Equation 2 represent c
- ☒ Hurray! Correct Answer

Solution Description

Equation 1 represent c and Equation 2 represent m

Optimal Coefficients:

--> "please refer to notes"

Linear Regression from Scratch

Question: 3

Coding Linear Regression from Scratch

Problem

Result

Coding Linear Regression from Scratch

[Send Feedback](#)

For the given dataset "FuelEconomy.csv"

You need to write four functions namely:

1. `fit(x_train, y_train)`
2. `predict(x, m, c)`
3. `score(y_truth, y_pred)`
4. `cost(x, y, m, c)`

The output is going to print the training and testing score and the cost of the regressor trained using the fit function.

Source Code:

```
## Open and read data file as specified in the question
## Print the required output in given format
import numpy as np
from sklearn.model_selection import train_test_split

data = np.loadtxt('FuelEconomy.csv', delimiter=',')
X = data[:, 0]
y = data[:, 1]

X_train, X_test, Y_train, Y_test = train_test_split(X, y, test_size = 0.3, random_state=42)
```



```
#####  
##### Start from here #####  
#####
```

```
def fit(x_train, y_train):
```

```
    ## Write code for fit function
```

```
    n = len(x_train)
```

```
    # Calculate the mean of x and y
```

```
    mean_x = np.mean(x_train)
```

```
    mean_y = np.mean(y_train)
```

```
    # Calculate the slope (m) and intercept (c) of the regression line
```

```
    numerator = np.sum((x_train - mean_x) * (y_train - mean_y))
```

```
    denominator = np.sum((x_train - mean_x) ** 2)
```

```
    m = numerator / denominator
```

```
    c = mean_y - m * mean_x
```

```
    return m, c
```

```
def predict(x, m, c):
```

```
    ## Write code for predict function
```

```
    # Predict the values using the regression line equation:  $y = mx + c$ 
```

```
    return m * x + c
```

```
def score(y_truth, y_pred):
```

```
    ## Write code for score function
```

```
    # Calculate the R-squared score
```

```
    mean_y = np.mean(y_truth)
```

```
    ss_total = np.sum((y_truth - mean_y) ** 2)
```

```
ss_residual = np.sum((y_truth - y_pred) ** 2)
```

```
r_squared = 1 - (ss_residual / ss_total)
```

```
return r_squared
```

```
def cost(x, y, m, c):
```

```
    ## Write code for cost function
```

```
    # Calculate the mean squared error
```

```
    y_pred = predict(x, m, c)
```

```
    mse = np.mean((y - y_pred) ** 2)
```

```
    return mse
```

```
#####
```

```
#####
```

```
m, c = fit(X_train, Y_train)
```

```
y_test_pred = predict(X_test, m, c)
```

```
y_train_pred = predict(X_train, m, c)
```

```
print(round(score(Y_test, y_test_pred), 2))
```

```
print(round(score(Y_train, y_train_pred), 2))
```

```
print(round(cost(X_train, Y_train, m, c ), 2))
```

Linear Regression from Scratch:

MCQ Questions:

1.

Identify Linear Regression Function

[Send Feedback](#)

In the last video we discussed how to write down code for Linear Regression algorithm. So which out of the four functions discussed is this ?

```
def fun(x_train, y_train)
    num = (x_train*y_train).mean() -x_train.mean() *
    y_train.mean()
    den = (x_train**2).mean() -x_train.mean()**2
    m =num/den
    c = y_train.mean() -m * x_train.mean()
    return m,c
```

Options

Attempts left: **1/2**

This problem has only one correct answer

- ☐ predict
- ☒ fit
- ☐ cost
- ☐ score

✓ Hurray! Correct Answer

Solution Description

This function is used to describe the "fit" function.

2.

Identify Linear Regression Function

[Send Feedback](#)

In the last video we discussed how to write down code for Linear Regression algorithm. So which out of the four functions discussed is this ?

```
def fun(x,m,c)
    return m*x + c
```

Options

Attempts left: **1/2**

This problem has only one correct answer

- ☐ score
- ☒ predict
- ☐ fit
- ☐ cost

✓ Hurray! Correct Answer

Solution Description

This function is used to describe the "predict" function.

3.

Identify Linear Regression Function

[Send Feedback](#)

In the last video we discussed how to write down code for Linear Regression algorithm. So which out of the four functions discussed is this ?

```
def fun(y_truth, y_pred)
    u = ((y_truth - y_pred)**2).sum()
    v = ((y_truth-y_truth.mean())**2).sum()
    return 1-u/v
```

Options

Attempts left: **1/2**

This problem has only one correct answer

- ☐ predict
- ☒ score
- ☐ cost
- ☐ fit

✓ Hurray! Correct Answer

Solution Description

This function is used to describe the "score" function.

4.

Identify Linear Regression Function

[Send Feedback](#)

In the last video we discussed how to write down code for Linear Regression algorithm. So which out of the four functions discussed is this ?

```
def fun(x,y,m,c)
    return ((y-m*x-c)**2).mean()
```

Options

This problem has only one correct answer

☐ predict

☐ score

☐ fit

☒ cost

✓ Hurray! Correct Answer

Solution Description

This function is used to describe the "cost" function.

Attempts left: **1/2**

Read: Linear Regression Notes

Assignment

Linear Regression - Diabetes Dataset

Linear Regression - Diabetes Dataset

[Send Feedback](#)

Diabetes dataset is one of the datasets available in sklearn. The diabetes dataset consists of 10 physiological variables (age, sex, weight, blood pressure) measure on 442 patients, and an indication of disease progression after one year.

You are given a Training dataset csv file with X train and Y train data. As studied in lecture, your task is to come up with Linear Regression training algorithm and thus predictions for the test dataset given.

Read Instructions carefully -

1. Use Linear Regression(in scikit learn) as a training algorithm and submit results predicted by that.
2. Files are in csv format, use genfromtxt function in numpy to load data from csv file. Similarly you can use savetxt function to save data into a file.
3. Submit a csv file with only predictions for X test data. File should not have any headers and should only have one column i.e. predictions. Also prediction values in file should be upto **5** decimal places.
4. Upon submission of predictions, the score you will get is based on coefficient of determination.

Data

test.csv 

train.csv 

[Upload Predictions](#) Predictions file should be in .csv format

[Upload Notebook](#) Notebook file should be in .ipynb format

Note :

1. You need to upload both predictions and notebook file to submit.
2. Please make sure that your submissions are genuine and you haven't cheated/copied from anywhere. Also only use techniques taught till this stage for submissions.
3. Your batchmates and TAs can view your submissions, and report them for use of invalid means if that happens. Your score will be reduced to 0 in that case.

Source code: → **check the repo..**