

Linear Regression and Introduction to Dimensional Reduction

The screenshot displays the Internshala Trainings web application. The interface includes a sidebar on the left with a 'Progress report' section and a list of modules and topics. The main content area shows two multiple-choice questions. The first question, 'QUESTION 1', asks to choose the correct statement based on MSE and RMSE values for two models. The second question, 'QUESTION 2', asks which statement is true about a linear regression model that perfectly fits the training data. Both questions have four options, and the correct answers are highlighted with green borders and green checkmarks.

Module: 5. Linear Regression

Topics: 3. Understanding Gradient Descent, 4. Assumptions of Linear Regression, 5. Implementing Linear Regression, 6. Feature Engineering, 7. Assignment, Module test

QUESTION 1
Multiple choice question (2/2 MARKS)

Consider that we have built two linear regression models over the same data set. MSE of model 1 is 10. RMSE of model 2 is 10.

Choose the correct statement.

- ☒ A. Model 1 is better
- ☐ B. Model 2 is better
- ☐ C. Both models are equally good
- ☐ D. Insufficient information

QUESTION 2
Multiple choice question (1/1 MARKS)

Let's say, a "Linear regression" model perfectly fits the training data (train error is zero). Now, which of the following statements is true?

- ☐ A. You will always have test error zero
- ☐ B. You can not have test error zero
- ☐ C. Prediction function will throw an error
- ☒ D. Cannot be determined

Excellent! Correct answer.

QUESTION 3

Multiple choice question (4/4 MARKS)

In a linear regression problem, we are using "R-squared" to measure goodness-of-fit. We add a feature in the linear regression model and retrain the same model. Which of the following options is true?

- ☐ A. If R squared increases, this variable is significant.
- ☐ B. If R squared decreases, this variable is not significant.
- ☒ C. Individually R squared cannot tell about variable importance. We can't say anything about it right now. ✓
- ☐ D. None of the above

✓ Bravo! Correct answer.

Q & A Forum

< Prev. Next >

QUESTION 4

Multiple choice question (1/1 MARKS)

Which one of the statements is true regarding residuals in regression analysis?

- ☒ A. Mean of residuals is always zero. ✓
- ☐ B. Mean of residuals is always less than zero.
- ☐ C. Mean of residuals is always greater than zero.
- ☐ D. There is no such rule for residuals.

✓ Excellent! Correct answer.

Q & A Forum

< Prev. Next >

QUESTION 5

Multiple choice question (1/1 MARKS)

Which of the following is true about heteroskedasticity?

- ☒ A. Linear regression with varying error terms
- ☐ B. Linear regression with constant error terms
- ☐ C. Linear regression with zero error terms
- ☐ D. None of the above

Well done! Correct answer.

Q & A Forum

Prev. Next

QUESTION 6

Multiple choice question (2/2 MARKS)

Which of the following indicates a fairly strong relationship between X and Y?

- ☐ A. Correlation coefficient = 0.7
- ☒ B. Correlation coefficient = -0.9
- ☐ C. Correlation coefficient = 0
- ☐ D. Insufficient information

Bravo! Correct answer.

Q & A Forum

Prev. Next

QUESTION 7

Multiple choice question (4/4 MARKS)

Which of the following assumptions do we make while deriving linear regression parameters?

- ☒ A. The true relationship between dependent y and predictor x is linear. ✓
- ☒ B. The model errors are statistically independent. ✓
- ☒ C. The errors are normally distributed with a 0 mean and constant standard deviation. ✓
- ☒ D. The predictor x is non-stochastic and is measured error-free. ✓

✓ Excellent! Correct answer.

Q & A Forum

< Prev. Next >

QUESTION 8

Multiple choice question (1/1 MARKS)

To test linear relationships of y (dependent) and x (independent) continuous variables, which of the following plots is best suited?

- ☒ A. Scatter plot ✓
- ☐ B. Bar chart
- ☐ C. Histogram
- ☐ D. None of the above

✓ Perfect! You got this right.

Q & A Forum

< Prev. Next >

QUESTION 9

Multiple choice question (1/1 MARKS)

Generally, which of the following method(s) is used for predicting continuous dependent variable?

☒ A. Linear regression



☐ B. Logistic regression

🟢 Bravo! Correct answer.

Solution

Correct answer : A

Your answer : A

Explanation

Logistic regression is used for classification problems. The regression term is misleading here.

Q & A Forum

< Prev. Next >

QUESTION 10

Multiple choice question (1/1 MARKS)

A correlation between age and health of a person found to be -1.09. On the basis of this, you would tell the doctors that:

☐ A. The age is a good predictor of health

☐ B. The age is a poor predictor of health

☐ C. Correlation and prediction have no association

☒ D. None of the above



🟢 Perfect! You got this right.

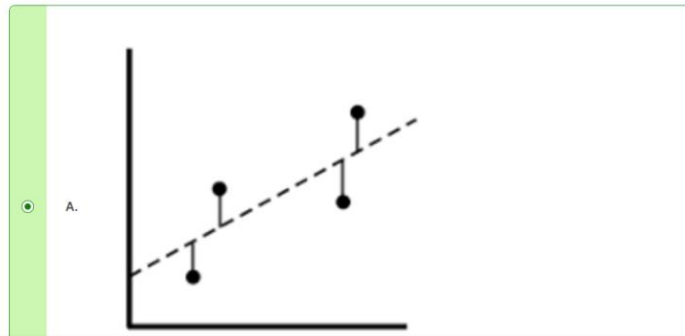
Q & A Forum

< Prev. Next >

QUESTION 11

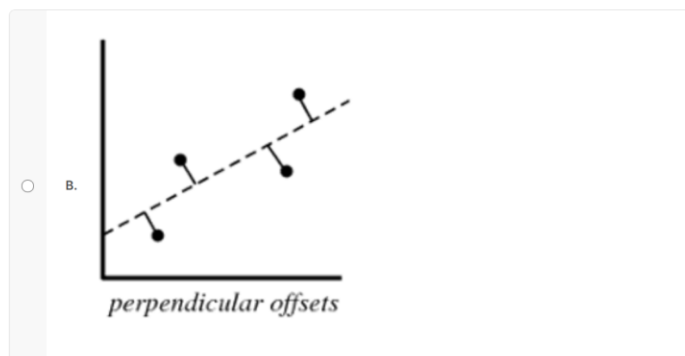
Multiple choice question (4/4 MARKS)

Which of the following offsets, do we use in case of least square line fit? Suppose the horizontal axis is an independent variable and vertical axis is dependent variable.



Q & A Forum

Prev. Next >



Well done! Correct answer.

Solution

Correct answer : A

Q & A Forum

Prev. Next >

QUESTION 12

Multiple choice question (2/2 MARKS)

Suppose we have generated the data with help of polynomial regression of degree 3 (degree 3 will perfectly fit this data). Now, consider below points and choose the option based on these points.

- ☒ A. Simple linear regression will have high bias and low variance.
- ☐ B. Simple linear regression will have low bias and high variance.
- ☐ C. Polynomials of degree 3 will have low bias and high variance.
- ☒ D. Polynomials of degree 3 will have low bias and Low variance.

Perfect! You got this right.

Q & A Forum

Prev. Next >

QUESTION 13

Multiple choice question (2/2 MARKS)

Which of the following is/are feature selection method(s)?

- ☐ A. Binning
- ☒ B. Forward feature selection
- ☐ C. Missing value ratio
- ☐ D. None of the above

Excellent! Correct answer.

Solution

Q & A Forum

Prev. Next >

QUESTION 1 OF 8

Multiple choice question

Which of the following can we infer from the residual plot of simple mean regression?

- ☒ A. "Spread of data" of the target variable ✓
- ☒ B. Possible "outliers" in target variable ✓
- ☐ C. We can infer nothing, it is for visualisation purposes only
- ☐ D. None of the above

✓ 2/2 are correct. Well done.

Q & A Forum

< Prev. Next >

Multiple choice question

Why do we make the mean regression model?

- ☒ A. It acts as the baseline/benchmark model ✓
- ☐ B. It is a compulsory step
- ☐ C. It gives an idea how good the independent variable is a predictor of the target variable
- ☐ D. None of the above

✓ Correct Answer.

Explanation

Q & A Forum

< Prev. Next >

Internshala Trainings

trainings.internshala.com/learn/machine-learning/screen/21316

Multiple choice question

"Mean Squared Error" and "Root Mean Squared Error" for the mean regression model is equivalent to which of the following options?

- ☐ A. Standard deviation and variance of the target variable
- ☐ B. Variance and covariance of the target variable
- ☒ C. Variance and standard deviation of the target variable
- ☐ D. None of the above

Correct Answer.

Explanation

Q & A Forum

Prev. Next >

Internshala Trainings

trainings.internshala.com/learn/machine-learning/screen/21316

What are residuals?

- ☐ A. It is the difference between the variance of the predicted values and the actual values.
- ☐ B. It is the difference between the standard deviation of the predicted values and the actual values.
- ☐ C. It is the absolute difference between the actual values and the predicted values.
- ☒ D. It is the simple difference between the predicted values and the simple values.

Correct Answer.

Explanation

Residuals is generally defined as the simple difference between the predicted and the actual values. Absolute difference is not used because it removes the interpretability of the predictions that determine whether they are overestimated or underestimated.

Q & A Forum

Prev. Next >

Internshala Trainings

trainings.internshala.com/learn/machine-learning/screen/21316

Multiple choice question

What happens when we calculate 'mean error' of any given model?

- ☐ A. The result is 'inf' (infinite).
- ☐ B. The result is a simple real number.
- ☐ C. The result is 'nan' (not a number).
- ☒ D. The result is '0' (zero).

Correct Answer.

Explanation

Q & A Forum

Prev. Next

Internshala Trainings

trainings.internshala.com/learn/machine-learning/screen/21316

Progress report

Module

5. Linear Regression

Topics

- 1. Introduction to Linear Regression
- 2. Introduction to Linear Regression
- 3. Understanding Gradient Descent
- 4. Assumptions of Linear Regression
- 5. Implementing Linear Regression
- 6. Feature Engineering
- 7. Assignment

Multiple choice question

What is the result if we calculate the R2 of the mean regression model?

- ☒ A. 0
- ☐ B. nan
- ☐ C. 1
- ☐ D. .5

Correct Answer.

Explanation

Q & A Forum

Prev. Next

Internshala Trainings

Multiple choice question

What is a linear regression?

- ☐ A. A simple model that linearly separates the target categorical variable
- ☒ B. A simple model that linearly traces the target continuous variable
- ☐ C. It is an error metric for the regression models
- ☐ D. None of the above

Correct Answer.

Explanation

Linear regression is used to best represent the continuous target variable. It predicts the continuous target variable.

Q & A Forum

Prev. Next

Internshala Trainings

Multiple choice question

Consider the following example:

We want to build a linear regression model for an employer who wants to predict the salary that a newly hired employee should get based on the past data. The employer only discloses the experience and the academic qualification attribute of the new employee.

What are the number of parameters in this model?

- ☐ A. 1
- ☐ B. 2
- ☒ C. 3
- ☐ D. Cannot say

Correct Answer.

Q & A Forum

Prev. Next

Quiz

QUESTION 4 OF 7

Fill in the blanks

In a linear regression model, there are fundamentally two types of parameters: m and c . The m parameter governs the slope of the regression line and the c represents the _____ of the line.

Type your answer here

f

✖ Incorrect Answer.

Solution

intercept / Intercept

Q & A Forum

< Prev. Next >

Multiple choice question

Cost function curve is a plot between:

- ☐ A. Model predictions and error of the model
- ☐ B. Model predictions and the parameter of the model
- ☒ C. Cost function and parameters of the model
- ☐ D. None of the above

✔ Bravo! Correct answer. (38% of students get this answer correct in their first attempt)

Explanation

Q & A Forum

< Prev. Next >

Internshala Trainings

Progress report

Module

5. Linear Regression

Topics

4. Assumptions of Linear Regression

6m 53s

Quiz

5. Implementing Linear Regression

6. Feature Engineering

7. Assignment

Module test

© Copyright Internshala 2022

Q & A Forum

Prev. Next

Well done! Correct answer. (35% of students get this answer correct in their first attempt)

Explanation

Options C and D follow a strict linear regression.

Multiple choice question

What is/are the assumption(s) of linear regression?

A. There should be a linear and additive relationship between dependent (response) variable and independent (predictor) variable(s).

B. There should be no correlation between the residual (error) terms.

C. The independent variables should not be correlated.

D. The error terms must have constant variance and must be normally distributed.

4/4 are correct. Well done.

Q & A Forum

Prev. Next

Internshala Trainings x Internshala Trainings x Internshala Trainings x +

trainings.internshala.com/learn/machine-learning/screen/21334

INTERNSHALA TRAININGS

Quiz

QUESTION 2 OF 5

Multiple choice question

What does the correlation coefficient tell us?

- ☒ A. Strength between two variables
- ☒ B. Direction of two variables
- ☐ C. None of the above

Perfect! You got this right. (63% of students get this answer correct in their first attempt)

Q & A Forum Prev. Next >

INTERNSHALA TRAININGS

Quiz

QUESTION 3 OF 5

Multiple choice question

What are the methods to check for multicollinearity?

- ☒ A. Correlation
- ☒ B. Scatter plot
- ☒ C. VIF
- ☐ D. Line plot

3/3 are correct. Well done.

Q & A Forum Prev. Next >

Internshala Trainings

trainings.internshala.com/learn/machine-learning/screen/21334

Multiple choice question

Which of the following can be used to check the normality of error terms?

- ☐ A. Count plot
- ☒ B. Residual plot
- ☐ C. Bar plot
- ☐ D. None of the above

Correct Answer.

Explanation

Residual plots are the best method to check the Normality of the error terms. The plot-based residual plot is effective.

Q & A Forum

Prev. Next >

Internshala Trainings

trainings.internshala.com/learn/machine-learning/screen/21334

Multiple choice question

Which one is true about heteroskedasticity?

- ☒ A. Linear regression with varying error terms.
- ☐ B. Linear regression with constant error terms.
- ☐ C. Linear regression with zero error terms.

Correct Answer.

Explanation

The presence of non-constant variance in the error terms results in heteroskedasticity. Generally, non-constant variance arises because of presence of outliers or extreme leverage values.

Q & A Forum

Prev. Next >

Internshala Trainings

Progress report

Module

6. Introduction to Dimensionality Reduction

Topics

1. Common Dimensionality Reduction Techniques

2. Advanced Dimensionality Reduction Techniques

3. Assignment

Module test

QUESTION 1

Multiple choice question (1/1 MARKS)

Imagine, you have 1000 input features and 1 target feature in a machine learning problem. You have to select 100 most important features based on the relationship between input features and the target features. Do you think this is an example of dimensionality reduction?

☒ A. Yes

☐ B. No

That's right!

Solution

Correct answer : A

Your answer : A

Explanation

QUESTION 2

Multiple choice question (1/1 MARKS)

It is not necessary to have a target variable for applying dimensionality reduction algorithms.

☒ A. True

☐ B. False

Bravo! Correct answer.

Solution

Correct answer : A

Your answer : A

Explanation

LDA is an example of a supervised dimensionality reduction algorithm.

QUESTION 3

Multiple choice question (2/2 MARKS)

I have 4 variables in the dataset such as - A, B, C & D. I have performed the following actions:

Step 1: Using the above variables, I have created two more variables, namely $E = A + 3 * B$ and $F = B + 5 * C + D$.

Step 2: Then using only the variables E and F, I have built a random forest model.

Could the steps performed above represent a dimensionality reduction method?

☒ A. Yes

☐ B. No

Bravo! Correct answer.

Solution

Correct answer : A

Your answer : A

QUESTION 4

Multiple choice question (1/1 MARKS)

Which of the following techniques would perform better for reducing dimensions of a data set?

☒ A. Removing columns which have too many missing values

☐ B. Removing columns which have high variance in data

☐ C. Removing columns with dissimilar data trends

☐ D. None of the above

Excellent! Correct answer.

QUESTION 5

Multiple choice question (1/1 MARKS)

Dimensionality reduction algorithms are one of the possible ways to reduce the computation time required to build a model.

☒ A. True

☐ B. False

That's right!

Solution

Correct answer : A

Your answer : A

Explanation

Reducing the dimension of data will take less time to train a model.

Prev. Next

QUESTION 6

Multiple choice question (2/2 MARKS)

Why do we remove variables with a high missing value ratio?

☒ A. Variables that have high missing values contain less learnable information.

☐ B. Variables with missing values have high variance.

☐ C. Missing values add generalization to the model.

☐ D. Missing values change the distribution of data.

Perfect! You got this right.

Prev. Next

QUESTION 7

Multiple choice question (1/1 MARKS)

Which of the following variables is more important in the dataset?

- Income (in Rs.) with a variance of 2000.
- Age (in years) with a variance of 10.

☐ A. Income variable

☐ B. Age variable

☒ C. Can't decide as variables are of different scale

Perfect! You got this right.

Solution

QUESTION 8

Multiple choice question (4/4 MARKS)

Why do we use the low variance filter?

☒ A. Variables with low variance do not contain any significant information

☒ B. They reduce the performance of the predictive model

☒ C. They decrease the effect of other independent variables

☐ D. None of the above

Perfect! You got this right.

Internshala Trainings

trainings.internshala.com/learn/machine-learning/module-test/1093

QUESTION 9

Multiple choice question (1/1 MARKS)

Which function is used to find the null values in a dataset?

- ☐ A. not_null
- ☒ B. isnull
- ☐ C. low_variance
- ☐ D. check_variance

✓ Bravoi! Correct answer.

QUESTION 10

Multiple choice question (2/2 MARKS)

How do we apply a high correlation filter?

- ☐ A. Eliminate the independent variable that is highly correlated with the target variable.
- ☐ B. Eliminate the target variable.
- ☐ C. If the two independent variables are highly correlated, eliminate the variable which has a higher correlation with the target variable.
- ☒ D. If the two independent variables are highly correlated, eliminate the variable which has a lower correlation with the target variable.

✓ Perfect! You got this right.