Today, Agenda! www.aiquest.org



1. Dataset splitting

2. Overfit and underfit concepts

3. Gradient descent

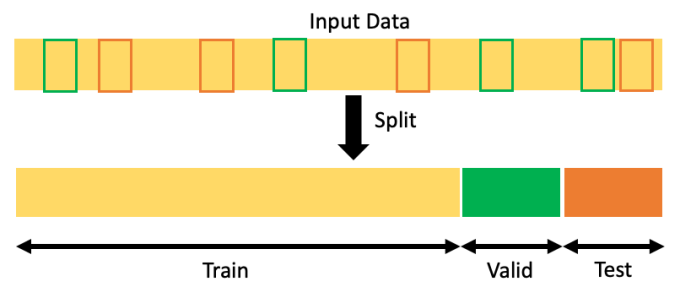
4. R2 Value

5. Concept of multi-variable LR

1

Splitting Dataset www.aiquest.org



100% 

70% 15% 15%2

Splitting Dataset 

Splitting a dataset generally refers to the process of dividing the data into separate subsets for training, validation, and testing during the development of a predictive model. This is a critical practice in machine learning and statistics to evaluate the performance of models. Here's what each subset is used for:

1. Training Set: This subset is used to train the model, meaning that the model learns the patterns from this data. Typically, the training set is the largest portion of the data.

2. Validation Set: After the model is trained, you use this subset to tune the model's hyperparameters. The validation set is used to evaluate the model's performance during the training process and to prevent overfitting. It acts as a proxy for the test set during this tuning phase.

3. Test Set: This subset is used to evaluate the final model's performance after it has been trained and validated. The test set is crucial because it is used to estimate how well the model is expected to perform on unseen data. It should only be used once for this purpose to avoid biasing the model's performance metrics.

The Common starting point is 70% for training, 15% for validation, and 15% for testing, or 80% for training and 20% for testing if a separate validation set is not used.

www.aiquest.org

3

Overfitting and Underfitting 

4

www.aiquest.org

Overfitting and Underfitting Overview

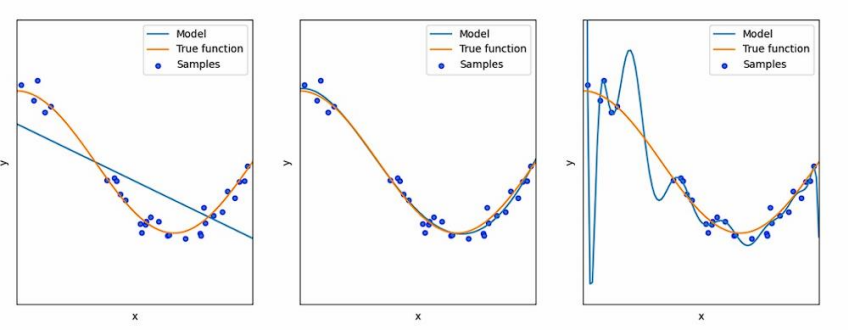
Training Performance: Low (Gorib)

Testing Performance: Low (Gorib)

Training Performance: Good Testing Performance: Good



Training Performance: High (Rich) Testing Performance: Low (Gorib)

High Bias, Low Variance Balanced Bias-Variance Low Bias, High VarianceFig 01: Underfitting Fig 02: Best fitting Fig 03: Overfitting

5

www.aiquest.org

Overfitting and Underfitting 

Why?

There are several reasons why a model might get overfitted, and it's helpful to understand the specifics of the situation to give you the most accurate answer. However, I can share some general causes and solutions:



www.aiquest.org

6

Overfitting and Underfitting 

Solution?



7

www.aiquest.org

Linear Regression using Gradient Descent 

Overview

Update Parameter (weights):

8

www.aiquest.org

Linear Regression using Gradient Descent 

Overview



The training curve could be used to monitor the effect of the

learning rate value and observe the convergence. We plot the

training loss against the number of training iterations.

9

www.aiquest.org

Linear Regression using Gradient Descent 

Overview

10 www.aiquest.org

Linear Regression using Gradient Descent 

Mathematics

11 www.aiquest.org

Linear Regression using Gradient Descent 

Mathematics

12 www.aiquest.org

Linear Regression using Gradient Descent Algorithm 

Step: 01

Gradient (m) = 0

Intercept (c) = 0

Learning Rate (L) = ~0.0001

Step: 02

Calculate the partial derivative of the Cost function with respect to m. Let the partial derivative of the Cost function with respect to m be Dm.

13 www.aiquest.org

Linear Regression using Gradient Descent 

Algorithm

Step: 03

Similarly, let’s find the partial derivative with respect to c. Let the partial derivative of the Cost function with respect to c be Dc.

14 www.aiquest.org

Linear Regression using Gradient Descent 

Calculus

15 www.aiquest.org

Linear Regression using Gradient Descent 

Algorithm

Step: 04 Update the value of the gradient and intercept.



16 www.aiquest.org

Linear Regression using Gradient Descent Algorithm

Repeat the steps!

1000 times

17 www.aiquest.org

Linear Regression with Multiple Variables 

Mathematical Representation

18 www.aiquest.org

Linear Regression with Single Vs. Multiple Variables Mathematical Representation 

Single 

Multiple

19 www.aiquest.org

R Squared Value / Model Accuracy 

Mathematical Calculation



▪ Residual sum of squared errors of our regression model (SSres)

▪ Total sum of squared errors (SStot)

20 www.aiquest.org

R Squared Value / Model Accuracy 

Mathematical Calculation

21 www.aiquest.org

R Squared Value / Model Accuracy 

Python Implementation

Way no: 01

reg.score(xtest, ytest)

Way no: 02

y\_pred = reg.predict(xtest) **#Predicted y**

from sklearn.metrics import r2\_score

Score = r2\_score(ytest, y\_pred)

22 www.aiquest.org

Measures for Classification: MCE & ACC 

Mathematical Calculation



23 www.aiquest.org