Class Overview 

▪ Objective Functions

▪ Regularization

▪ L1 Regularization

▪ L2 Regularization

▪ Polynomial Regression

▪ Multi Variable Linear Regression

▪ Regression Model Evaluation

▪ SSR

▪ SST

▪ R2 value

1

www.aiquest.org

Linear/Non-Linear Relationship

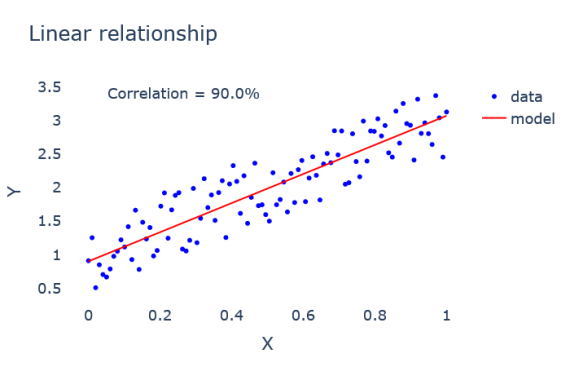
in Machine Learning

www.aiquest.org

2

Relationship www.aiquest.org

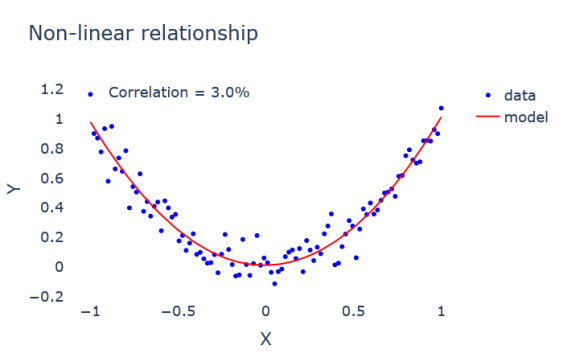


Fig: Correlation between X and Y

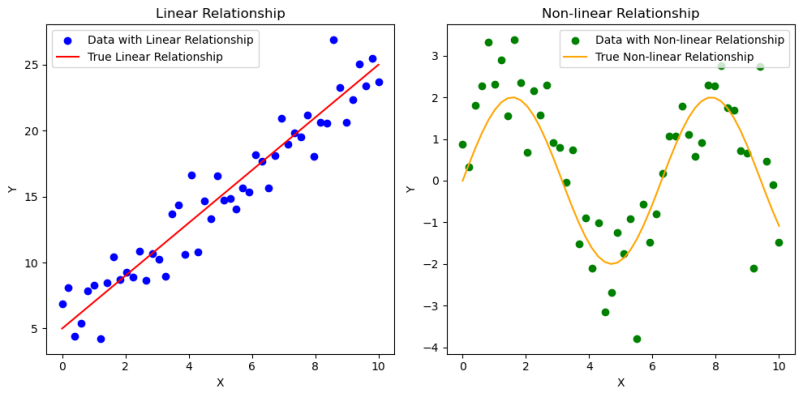
3

Relationship www.aiquest.org



Fig: Correlation between X and Y

4

Relationship 

5

www.aiquest.org

Correlation 

To calculate the correlation coefficient between two variables X and Y, you can use Pearson's correlation coefficient formula.



However, it's important to note that Pearson correlation measures only linear relationships and may not capture other types of relationships.

www.aiquest.org

6

Correlation

Example Mathematics:

www.aiquest.org

7

Correlation 

8

www.aiquest.org

Correlation 

9

www.aiquest.org

Correlation 

So, the correlation between x and y is 1.

10 www.aiquest.org

Correlation 

To calculate the correlation coefficient between two variables X and Y, you can use Pearson's correlation coefficient formula. **Another representation:**

****

11 www.aiquest.org

Correlation 

12 www.aiquest.org

Multiple Variable Regression 

Mathematical concepts

13 www.aiquest.org

Polynomial Regression 

Mathematical concepts

14 www.aiquest.org

Polynomial Regression 

Mathematical concepts

The "degree" of the polynomial refers to the **highest power of the input variable(s)** in the polynomial equation. The degree indicates the level of complexity of the model, with higher degrees allowing for more complex curves to fit the data.

15 www.aiquest.org

Polynomial Regression 

Mathematical concepts



16 www.aiquest.org

Polynomial Regression 

Mathematical concepts

The formula to calculate the total number of features after applying polynomial features of degree *d* to an original set of *n* features is given by:

17 www.aiquest.org

Polynomial Regression 

Mathematical concepts



After performing polynomial regression:

❖ With a degree of 2, the feature count will increase to 10. ❖ With a degree of 3, the feature count will increase to 20.

18 www.aiquest.org

Polynomial Regression Pros and Cons 

The "best" degree for polynomial regression depends on your specific dataset and the **complexity** of the **underlying relationship** that you're trying to **model between the features and the target variable**. Here are some considerations to help decide between a degree of 2 and a degree of 3, or any other degree:

**Degree 2 (Quadratic):**

• **Pros:**

• Captures relationships that are not merely linear, such as **parabolic trends**.

• Increases model complexity moderately, which can improve fitting for slightly more complex patterns without going too far into overfitting.

• **Cons:**

• May still be **too simple** for some datasets, failing to capture higher-order relationships.

19 www.aiquest.org

Polynomial Regression

Pros and Cons

**Degree 3 (Cubic):**

• **Pros:**

• Can model **more complex** patterns than quadratic, including S-shaped curves, which can be more representative of certain natural or economic phenomena.

• Useful for datasets where the relationship between **features and target changes direction more than once.**

• **Cons:**

• Increases the risk of overfitting, especially if the dataset is not very large, as the model becomes significantly more complex.

• Requires more data to train effectively compared to quadratic models.

20 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

21 www.aiquest.org

Polynomial Regression General Considerations: 

▪ **Model Complexity vs. Overfitting**: Higher degrees can capture more complex relationships but also increase the risk of overfitting. Overfitting happens when the model learns the noise in the training data instead of the actual underlying pattern, leading to poor performance on unseen data.

▪ **Cross-Validation**: To find the best degree for your polynomial regression, you can use cross-validation. This involves training the model with different degrees and evaluating their performance on a validation set or through k-fold cross-validation.

▪ **Domain Knowledge**: Sometimes, domain knowledge can guide the choice of degree. If you know the underlying relationship should have a specific shape or form, you can choose the degree accordingly.

22 www.aiquest.org



Regression Model Evaluation

23 www.aiquest.org

Regression Performances 

The coefficient of determination

R2 is a statistical measure that represents the proportion of the variance in the dependent variable that is predictable from the independent variable(s) in a regression model. Mathematically, it is defined as follows:



24 www.aiquest.org

Regression Performances 

The coefficient of determination

25 www.aiquest.org



Objective Functions

26 www.aiquest.org

Objective Function Basics 

The objective function, or the cost function or loss function, is a mathematical function that a machine learning algorithm seeks to minimize during the training process. It represents a measure of the error or discrepancy between the predicted values of the model and the actual target values in the training data.

The primary goal of a machine learning algorithm is to learn a set of parameters that minimizes this objective function, thereby improving the model's ability to make accurate predictions on new, unseen data.

The objective function guides the optimization algorithm, which iteratively adjusts the model parameters to find the optimal values. or loss function is a mathematical function that a machine learning algorithm seeks to minimize during training loss function, is a mathematical function that a machine learning algorithm seeks to minimize during training loss function, is a mathematical function that a machine learning algorithm seeks to minimize during training

27 www.aiquest.org

Objective Function 

Types / Situations

❖ **Regression Problems:** The mean squared error (MSE) is a common objective function. It measures the average squared difference between the predicted and actual values.

❖ **Classification Problems:** The cross-entropy loss is often used. It quantifies the dissimilarity between the true class labels and the predicted probabilities.

28 www.aiquest.org

Objective Function 

The Mean Squared Error

The Mean Squared Error (MSE) is a commonly used objective function for regression problems. It quantifies the average squared difference between the predicted values of a model and the actual target values in the training data. The formula for MSE is as follows:

29 www.aiquest.org

Objective Function 

The Mean Absolute Error

The Mean Absolute Error (MAE) is a metric used to evaluate the performance of a regression model. It measures the average absolute difference between the actual and predicted values of a set of data points.



30 www.aiquest.org

Objective Function 

The Root Mean Squared Error

The Root Mean Squared Error (RMSE) is another common metric used to evaluate the performance of regression models, like Mean Absolute Error (MAE). However, RMSE gives more weight to larger errors because it squares the differences between predicted and actual values before taking the square root of the average.



31 www.aiquest.org

Objective Function MSE vs. MAE vs. RMSE 

**MSE Characteristics**:

❖ MSE is often used when you want to penalize **larger errors** more heavily, which might be appropriate in certain contexts where larger errors are considered more critical.

❖ It is also commonly used in situations where the assumption of normally distributed errors holds, as minimizing MSE is equivalent to maximum likelihood estimation under this assumption. Maximizing the likelihood also provides a measure of how well the model fits the observed data. A higher likelihood value indicates a better fit between the model and the data.

❖ However, MSE is **sensitive to outliers** due to the squaring of errors, so if your data contains outliers, it might not be the best choice.

32 www.aiquest.org

Objective Function MSE vs. MAE vs. RMSE 

**MAE Characteristics**:

❖ MAE is preferred when you want a metric that is **robust to outliers** since it treats all errors equally. ❖ It is easier to interpret than MSE because it is in the same units as the original data.

❖ MAE might be preferred in situations where interpretability is crucial or when the distribution of errors is **not assumed to be normal**.

**RMSE Characteristics**:

❖ RMSE combines the advantages of MSE (sensitivity to larger errors) with the **interpretability** of the original scale of the data (like MAE).

❖ It is commonly used when you want a metric that is both interpretable and penalizes larger errors.

❖ RMSE is useful when the errors are **normally distributed**, and you want a metric that aligns with maximum likelihood estimation.

33 www.aiquest.org



Regularization (L1 & L2)

34 www.aiquest.org

Regularization Basics 

Regularization is a technique used in machine learning to prevent overfitting and improve the generalization of a model. Overfitting occurs when a model learns the training data too well, including its noise and outliers, to the extent that it performs poorly on new, unseen data.

Regularization methods add a penalty term to the model's loss function, discouraging the learning algorithm from assigning excessive importance to any one feature or allowing the model to become too complex.

❖ L1 Regularization (Lasso)

❖ L2 Regularization (Ridge)

35 www.aiquest.org

Regularization 

Lasso / L1

36 www.aiquest.org

Regularization 

Lasso / L1 with Linear Regression

L1 regularization, also known as Lasso (Least Absolute Shrinkage and Selection Operator) regularization, has the property of inducing sparsity in models. When you apply L1 regularization to linear regression (or other linear models), it adds a penalty term to the cost function that is proportional to the absolute values of the coefficients. This penalty encourages the optimization process to drive some of the coefficients to exactly zero. The L1 regularization term is typically added to the mean squared error (MSE) cost function for linear regression, resulting in the Lasso cost function::

37 www.aiquest.org

Regularization 

Ridge / L2

38 www.aiquest.org

Regularization 

Ridge / L2 with Linear Regression

Linear regression with an L2 regularization term is often referred to as Ridge regression. The L2 regularization term adds the sum of the squared values of the coefficients to the linear regression cost function. The complete cost function for Ridge regression is given by:

39 www.aiquest.org

Regularization 

L1 Vs. L2 and Applications

40 www.aiquest.org



Let’s do it with Python

41 www.aiquest.org