Linear regression

> Linear regression is a supervised machine learning algorithm used to find a linear relationship between a dependent (y) and one or more independent (y) variables, hence called linear regression Types of linear regression:

1. Simple Linear Regression
2. Multiple Linear Regression
3. Polynomial Linear Regression

Simple linear regression

> If a single independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Simple Linear Regression. OR

> It is a linear regression model with only one dependent and one inedependent variable. y = mx + c

multiple linear regression

> If more than one independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Multiple Linear Regression. OR

> It is a linear regression model with multiple independent and one dependent variable.

> y = m1x1 + m2x2 + ... + mNxN + c

y = Dependent Variable (Target Variable)

x = Independent Variable (predictor Variable)

c = intercept of the line (Gives an additional degree of freedom)

m = Linear regression coefficient (scale factor to each input value).

Assumptions of the Linear regression model

1. Linearity: Relationship between the independent and dependent variables to be linear.

2. No Multicollinearity (Independence): Observations are independent of each other.

3. Normality of Residual

4. Homoscedasticity: The variance of residual is the same for any value of X.

if these assumptions get violated

1. The validity of the results may be compromised: If the assumptions underlying a study are violated, the results may not accurately reflect reality, leading to potentially misleading conclusions.

2. Increased likelihood of bias: Violating assumptions can introduce bias into the analysis, since the data may not be representative of the population or may not be randomly sampled.

3. Decreased generalizability: If assumptions are violated, it may be difficult to generalize the findings of a study to a larger population or to a different context.

4. Inaccurate predictions: Violating assumptions can lead to inaccurate predictions or forecasts, as the models used may not accurately capture the underlying relationships in the data.

5. Difficulty in interpreting results: Violating assumptions can make it challenging to interpret the results of a study, as the relationships between variables may be distorted or unclear.

6.Increased risk of Type I or Type II errors: Violating assumptions can increase the risk of making Type errors (incorrectly rejecting a true null hypothesis)/Type II errors (failing to reject a false null hypothesis).

assumption of Linearity

The assumption of linearity in linear regression states that the relationship between the independent variables and the dependent variable is linear.

In other words, the change in the dependent variable is directly proportional to changes in the independent variables.

to check the Linearity:

1. Coefficient of correlation

2. Scatter Plot

3. Correlation matrix

to Handle the Linearity if get violated:

Apply a nonlinear transformation to the independent and/or dependent variable

1. Log transformation

2. Square root transformation

3. Reciprocal transformation.

What is the assumption of homoscedasticity? How to check homoscedasticity? How to prevent heteroscedasticity?

Homoscedasticity:

Residuals have constant variance at every level of x. This is known as homoscedasticity. When this is not the case, the residuals are said to suffer from heteroscedasticity.

When heteroscedasticity is present in a regression analysis, the results of the analysis become hard to trust

to Check the Homoscedasticity:

1. Scatter plot between fitted value and residual plot.

to prevent Homoscedasticity :

1. Transform the dependent variable(Y): log transformation of the dependent variable
2. Redefine the dependent variable
3. Use weighted regression: This type of regression assigns a weight to each data point based on the variance of its fitted value.

assumption of normality

1. The assumption of normality refers to the assumption that the data being analyzed comes from a population that follows a normal distribution.

2. It is a common assumption in many statistical tests and models, such as regression analysis and analysis of variance (ANOVA).

3. The assumption of normality is important because many statistical tests rely on the assumption that the data is normally distributed in order to provide accurate results.

4. Violation of the assumption of normality can lead to biased estimates and incorrect conclusions.

5. If the data does not follow a normal distribution, non-parametric statistical tests or data transformations may be used as alternatives.

to the check normality:

1. Graphs for Normality test:
    - Distribution curve, Histogram (sns. distplot, sns.kdeplot)
    - Q-Q or Quantile-Quantile Plot
2. Statistical Tests for Normality(Hypothesis Testing):
    - Shapiro-Wilk test
    - Kolmogorov-Smirnov test
    - D'Agostino's K-squared test

to the handle Normality:

1. Check and remove outliers
2. Apply a nonlinear transformation to the independent and/or dependent variable

multicollinearity mean

1. Multicollinearity refers to a situation where two or more independent variables in a regression model are highly correlated with each other.

2. Multicollinearity can make it challenging to identify the true effect of each independent variable on the dependent variable.

3. It can lead to a loss of statistical power and accuracy in the regression model.

4. Techniques for dealing with multicollinearity include removing one of the correlated variables, using dimension reduction methods, or combining the variables into a composite variable.

check Multicollinearity

1. VIF (Variable Inflation Factors): The VIF score of an independent variable represents how well the variable is explained by other independent variables. VIF = 1 → No correlation VIF = 1 to 5 → Moderate correlation VIF >10 → High correlation

2. Correlation matrix / Correlation plot

3. Scatter plots

VIF ? What is the best value of VIF?

-Variance inflation factor (VIF) is a measure used in regression analysis to assess how much multicollinearity is present among the independent variables.

-VIF measures how much the variance of an estimated regression coefficient is increased due to multicollinearity.

Best value:

-The formula for calculating the VIF VIF= 1/(1−R2) Where R2 is the coefficient of determination of the regression model.

-The best value of VIF (Variance Inflation Factor) is typically considered to be 1.

-This indicates that there is no multicollinearity between the independent variables in a regression model.

-A VIF between 5 and 10 may also indicate moderate multicollinearity, and values below 5 are generally considered acceptable.

-Ultimately, the goal is to minimize multicollinearity as much as possible to ensure reliable and accurate results in regression analysis.

feature selection methods in Linear Regression

Feature selection is the process of selecting the independent variables over the dependent variable from the given dataset.

-Filter methods: It is used to filter out the correlation of dependent and independent variables common filter methods include correlation analysis, mutual information, etc.

-LASSO Regression (L1 Regularization): Applies a penalty to the absolute values of the regression coefficients, encouraging some coefficients to become exactly zero. Features with zero coefficients are effectively excluded from the model.

feature scaling Is it required in Linear Regression.

Feature scaling is a technique used to standardize the range of independent variables.

It is done to ensure that all variables have the same scale and are equally weighted in the model. Common methods of feature scaling include normalization and standardization.

In linear regression, feature scaling is not always required. However, it can be beneficial in cases where the independent variables have different scales.

This is because variables with larger scales can dominate the learning process and lead to biased results. By scaling the features, we can improve the performance of the model and make the coefficients more interpretable.

find the best fit line in a linear regression model

To find the best fit line in a linear regression model, you can use the least squares method. Here are the steps to follow:

1. Gather your data: Collect a set of data points that represent a relationship between two variables.

2. Calculate the mean of both variables: Calculate the mean of the independent variable (x) and the dependent variable (y).

3. Calculate the covariance and variance of the data: Calculate the covariance between x and y, and the variance of x.

4. Calculate the slope of the line (m): m = covariance(x, y) / variance(x).

5. Calculate the intercept of the line (b): b = mean(y) - m * mean(x).

6. Formulate the equation of the best fit line: y = mx + b.

Why do we square the error instead of using modulus?

One reason for squaring the error instead of using the modulus is that squaring the error penalizes larger errors more than smaller errors,which can be useful in certain applications.

Squaring the error simplifies mathematical calculations, such as differentiation and optimization, making it easier to work with in various statistical and machine learning algorithms.

Squaring the error leads to a smoother, differentiable loss function, which can help in optimizing the model parameters more efficiently during the training process

What techniques are adopted to find the slope and intercept of the linear regression line of the model?

1.Ordinary Least Squares (OLS): We find the slope and intercept by minimizing the total squared distance between each data point and the line. This method is like adjusting the line until it fits the data as closely as possible.

2.Gradient Descent: We iteratively adjust the slope and intercept in the direction that reduces the error between the data points and the line, kind of like taking small steps downhill until we reach the lowest point.

3.Closed-form Solution: For simpler cases, there is a direct mathematical formula to find the slope and intercept without needing to do lots of calculations. 4. Matrix Inversion: We can use some math tricks involving matrices to find the slope and intercept efficiently.

What is the cost Function in Linear Regression?

The cost function in Linear Regression, often referred to as the Mean Squared Error (MSE), is a measure of how well the regression line fits the data points. It is defined as the average of the squared differences between the predicted and actual values.

Minimizing the MSE corresponds to finding the line that best fits the data points in the feature space.

Fomulation of MSE is:-

MSE= 1/n(sum(Yact-Ypred)**2)

where, n is the number of observations Yact is actual Y value Ypred is predicted Y value

the gradient descent algorithm

Gradient descent is an iterative optimization algorithm that aims to find the optimal values for a model’s parameters.

Gradient descent iterates until the model converges to an optimal solution to find Best slope(m) & intercept(c) values and reduce cost function(MSE).

-Procedure :

- Model will select any random values of m&c to find new m&c values

- It uses partial derivatives to find new m&c values.

- We can get best m&c values from infinite number of possibilities.

- Step size become smaller and smaller till we reach global minima point

- Global Minima Point where we achieve Best m&c values an least MSE

evaluate regression models

To evaluate regression models, various metrics and techniques can be used:

Mean Squared Error (MSE): Measures the average squared difference between actual and predicted values.

Mean Absolute Error (MAE): Computes the average absolute difference between actual and predicted values, providing a more interpretable metric compared to MSE. .>R-squared (R2): Indicates the proportion of the variance in the dependent variable that is predictable from the independent variables. A higher R2 value indicates a better fit of the model to the data. Root Mean Square Error (RMSE): The square root of MSE, providing a measure of the average magnitude of errors in the predicted values. Adjusted R-squared:

Similar to R2 but adjusted for the number of predictors in the model, penalizing excessive complexity.

Residual Analysis:

Plotting the residuals (difference between observed and predicted values) to assess the model isassumptions and performance.

Cross-Validation: Splitting the data into subsets to test the model is performance on different samples. Helps avoid bias and assess generalization.

Prediction Intervals: They give a range of values within which future observations are expected to fall with a certain confidence level.

Information Criteria: These help balance model fit with complexity and are used for comparing different models.

Comparing Models: By using these metrics and techniques, we can compare different models to choose the best one for our data.

Which evaluation technique should you prefer to use for data with many outliers in it?

Outliers are data points that are far from other data points. When dealing with data that has many outliers, principal component analysis (PCA) is an advanced multivariate method that can be used to detect outliers.

Residual? How is it computed?

Residuals are the error difference between actual data points and predicted data points.

It is calculated by formula residuals = Ya -Yp where, Ya = Actual Data Point. Yp = Predicted Data Point.

What are SSE, SSR, and SST? and What is the relationship between them?

> -SSE,SSR and SST are the evolution parameters in Linear Regression.

> -SSE (Sum of sqaured error) SSE = sum[(Ya - Yp)^2]

> -SSR (Sum of squared error due to regression) SSR = sum[(Yp - Ymean)^2]

> -SST (Total Error) SST = SSE + SSR = sum[(Ya - Ymean)^2]

The relationship between them can be expressed as:

> SST = SSR + SSE

Intuition behind R-Squared

> R-squared is a statistical measure that represents the goodness of fit of a regression model.

> The value of R-square lies between 0 to 1.

> Where we get R-square equals 1 when the model perfectly fits the data and there is no difference between the predicted value and actual value.

> However, we get R-square equals 0 when the model does not predict any variability in the model and it does not learn any relationship between the dependent and independent variables.

> The goodness of fit of regression models can be analyzed on the basis of the R-square method. The more the value of the r-square near 1, the better the model is.

coefficient of determination

> The coefficient of determination, often denoted as R-squared, explains how well the independent variables in a regression model predict the variation in the dependent variable.

> Explanation: R-squared measures the proportion of variation in the dependent variable that is explained by the independent variables in the model.

> Interpretation: For example, an R-squared of 0.8 means that 80% of the variability in the dependent variable is explained by the independent variables in the model.

> Higher R-squared: A higher R-squared indicates that the model does a better job of predicting the dependent variable is variation.

> Lower R-squared: Conversely, a lower R-squared suggests that the model's predictions may not be capturing much of the variability in the dependent variable.

Can R2 be negative?

Yes it can be negative,

> if R2 is -ve then (SSE/SST)>1, Negative R2 values indicate that the model's predictions have a larger error than those of the simplest possible model.

> It suggests that the model does not capture the variability in the data and might be a result of overfitting or fitting a poorly chosen model.

> R2 = 0 > SSE == SST (Data points are far away from BFL) # worst score

> R2 = -ve >> SSE > SST (data points not linear)

flaws in R-squared

> 1. Depends on Sample Size: R-squared can go up just because you have more data, even if the model is not necessarily better.

> 2. Can not Handle Nonlinear Relationships: It only measures linear relationships between variables, so it might miss important patterns in the data if they are not linear.

> 3. Affected by Outliers: Outliers can greatly change R-squared, making it misleading about how well the model fits the majority of the data.

> 4. Does not Predict Well: R-squared does not tell you how well the model predicts future data. A high R-squared does not guarantee accurate predictions.

> 5. Multicollinearity Confusion: If variables are highly correlated, R-squared might overestimate the model is performance.

> 6. Assumes Equal Variance: R-squared assumes that the variation of the errors is the same across all values of the independent variables. If this is not true, R-squared can be misleading.

> 7. Not Good for Comparing Different Models: It is not ideal for comparing models with different sets of predictors because adding more predictors can make R-squared higher, even if they are not actually helping the model.

adjusted R2?

- Adjusted R-squared tells us the proportion of variation explained by only the independent variables that actually affect the dependent variable. The adjusted R-Squared will penalize for adding independent variables that do not fit the model. Adjusted R-squared will always be less than or equal to R-squared.
- The formula for adjusted R-Squared is :- Adjusted R2 = 1 – [(1-R2)*(n-1)/(n-k-1)] where, R2: The R2 of the model

n: The number of observations k: The number of predictor variables

-Interpretation: Adjusted R-squared ranges from 0 to 1, similar to R-squared. A higher value indicates a better fit of the model to the data, while considering the model's complexity.

- Comparison: When comparing models, it is often preferable to use Adjusted R- squared, as it accounts for the number of predictors and provides a more accurate

measure of model fit, especially in situations where additional predictors may not significantly improve the model is performance.

Coefficient of Correlation: Definition, Formula

Coefficient of correlation is denoted by R-Value. It is used to find corelation between two variable(Input and Output variable)

Formula for finding R - value is , R_value = Covarience/Standard deviation. = Sum[(Xi - Xmean)*(Yi - Ymean)]/sqrt(sum[(Xi - Xmean)^2 * (Yi - Ymean)^2]) -r=1 indicates a perfect positive linear relationship, meaning that as one variable increases, the other variable also increases in a linear fashion. -r=−1 indicates a perfect negative linear relationship, meaning that as one variable increases, the other variable decreases in a linear fashion. -r=0 indicates no linear relationship between the variables.

best values for correlation

- Perfect Positive Correlation (Best): When (r = 1), it means the variables move together perfectly in the same direction.
- Perfect Negative Correlation (Also Good): When (r = -1), it means the variables move perfectly in opposite directions.
- No Correlation (Neutral): When (r = 0), it means there is no clear relationship between the variables.

difference between Correlation and covariance?

-Covariance: It measures how two variables change together. If one goes up while the other goes up, the covariance is positive. If one goes up while the other goes down, it is negative. But the actual numbers can be hard to interpret because they depend on the units of the variables.

-Correlation: It is like covariance, but it is scaled to always be between -1 and 1. A correlation of 1 means the variables move together perfectly in the same direction. A correlation of -1 means they move perfectly in opposite directions. And a correlation of 0 means there is no clear relationship.

Relationship between R-Squared and Adjusted R-Squared?

-R-Squared (R2): Measures how well the model fits the data, but it does not consider the number of predictors. -Adjusted R-Squared: Similar to R-squared, but it adjusts for the number of predictors in the model. -Relationship: Adjusted R-squared will be lower than R-squared if there are multiple predictors. It penalizes adding unnecessary predictors, giving a more accurate picture of model fit. -Use: Adjusted R-squared is preferred when comparing models with different numbers of predictors because it accounts for model complexity.

Difference between overfitting and underfitting?

-Overfitting: In Overfitting, training accuracy is high and testing accuracy is low,so we can say that model learned well from training data but not performing well on testing data. -Underfitting: In Underfitting, training accuracy is low and testing accuracy is also low, so we can say that model never learned from training dataset and not perfoming well on testing dataset

How to identify if the model is overfitted or under fitted? Explain in terms of Bias and Variance

- Overfitting: A model suffers from overfitting when it performs well on the training data but poorly on unseen data. It typically exhibits low bias and high variance.

- In terms of Bias and Variance:- In overfitting, the model has learned the noise in the training data, resulting in low bias (i.e., the model is too flexible). However, this flexibility causes the model to be highly sensitive to variations in the training data, leading to high variance.
- Underfitting: An underfitted model performs poorly on both the training and unseen data. It typically exhibits high bias and low variance.

How to interpret a Q-Q plot in a Linear regression model?

- Q-Q Plot: It is like comparing two sets of numbers to see if they match up. In linear regression, we use it to check if our data follows a normal (bell-shaped) distribution.
- How it Works: We plot the residuals (the differences between actual and predicted values) from our regression model against what we would expect if they were normally distributed.
- Interpretation of Q-Q plot: -Straight Line: If the points form a straight line, it means our data is close to normal. -Curves or Bends: If the points curve upwards or downwards, or bend away from the straight line, it suggests our data is not perfectly normal. -Outliers: If points are far from the line, it suggests outliers or extremes in our data.

Advantages and disadvantages of Linear Regression? What is Regularization in Machine Learning?

Advantages:

- Easy to Understand: Linear regress ion gives us clear, easy-to-understand results. We can see how each independent variable affects the dependent variable.
- Simple to Use: It is straightforward to use and does not require advanced statistical knowledge. You can quickly apply it to various problems.
- Widespread Use: Linear regression can be used in many situations, from predicting sales to understanding relationships in research studies.
- Assumption Clarity: We know the assumptions it makes (like the relationship being linear), so we can check if those are met and adjust if needed.
- Identifying Important Factors: It helps us see which factors are most important in predicting the outcome.

Disadvantages:

- Assumption Limits: Linear regression assumes certain things about the data, like a linear relationship between variables. If these assumptions are not met, the results may be inaccurate.

- Can not Handle Complexity: It is not good at capturing complex relationships between variables. If the relationship is not linear, it might not give accurate results.
- Sensitive to Outliers: It can be thrown off by outliers, which are extreme values in the data. They can skew the results and make them less reliable.
- Risk of Overfitting or Underfitting: Sometimes the model may be too simple (underfitting) or too complex (overfitting), leading to inaccurate predictions.
- Multicollinearity Issues: When independent variables are highly correlated, it can confuse the model and make it less accurate.

Lasso Regression(L1 Regularization) in Details

-Regularization in machine learning is a technique used to prevent overfitting by adding a penalty term to the model is objective function. The objective is to discourage the model from learning complex patterns in the training data that may not generalize well to new, unseen data.

- Common types of regularization techniques include L1 regularization (Lasso), L2 regularization (Ridge), each of which applies a different penalty to the model is coefficients. These techniques help to strike a balance between bias and variance, leading to models that perform well on both training and unseen data.
- Regularization is particularly important in scenarios where the training data is limited or noisy, as it helps to prevent models from memorizing the training examples and instead encourages them to capture the underlying patterns in the data.
- Preventing Overfitting: Regularization helps prevent overfitting, where the model learns noise or irrelevant details from the training data.
- Penalty Term: Regularization adds a penalty term to the model is loss function, which penalizes large coefficients or complex models.
- Types of Regularization: - L1 Regularization (Lasso): Adds the absolute value of the coefficients as the penalty term. It encourages sparsity in the coefficients, effectively performing feature selection by shrinking some coefficients to zero. - L2 Regularization (Ridge): Adds the square of the coefficients as the penalty term. It penalizes large coefficients without forcing them to zero, leading to smaller but non-zero coefficients.
- Control Parameter: Regularization introduces a hyperparameter (e.g., λ for L1 or L2 regularization) that controls the strength of regularization. The choice of this parameter affects the balance between model simplicity and accuracy.
- Trade-off: Regularization introduces a trade-off between model complexity and generalization performance. By penalizing complex models, regularization helps improve the model is ability to generalize to new data.
- Application: Regularization is commonly used in linear regression, logistic regression, neural networks, and other machine learning algorithms to improve model performance and stability.

Ridge Regression(L2 Regularization) in Details

-Lasso Regression, also known as L1 regularization, is a linear regression technique used for variable selection and regularization. It adds a penalty term to the standard linear regression objective function, which is the sum of squared residuals, to control the model's complexity and prevent overfitting.

-- L1 regularization is particularly useful when working with high-dimensional data since it enables one to choose a subset of the most important attributes. This lessens the risk of overfitting and also makes the model easier to understand. The size of a penalty term is controlled by the hyperparameter lambda, which regulates the L1 regularization regularization strength. As lambda rises, more parameters will be lowered to zero, improving regularization.

-Advantages of Lasso Regression: -Automatic feature selection. -Handles multicollinearity well by shrinking correlated variables. -Helps prevent overfitting by controlling model complexity.

-Disadvantages of Lasso Regression: -Less stable than Ridge Regression when features are highly correlated. -Not suitable for situations where all features are necessary for prediction. -May not perform well if the number of predictors is much larger than the number of observations.

1. Objective Function: -In Lasso Regression, the objective function is modified by adding a penalty term proportional to the absolute values of the regression coefficients. -The objective function to be minimized becomes: [\text{Objective} = \sum_{i=1}^{n} (y_i - \hat{y}i)^2 + \lambda \sum{j=1}^{p} | \beta_j |] -Here, (\lambda) is the regularization parameter that controls the strength of the penalty term. A higher (\lambda) value results in more regularization and vice versa.
2. Shrinking Coefficients: -The penalty term in Lasso Regression encourages sparsity in the regression coefficients by shrinking some coefficients to exactly zero. -This property makes Lasso Regression useful for feature selection, as it effectively eliminates irrelevant or less important features from the model.
3. Variable Selection: -Lasso Regression performs automatic variable selection by driving some coefficients to zero. -Features with non-zero coefficients are considered important predictors of the outcome, while features with zero coefficients are effectively ignored by the model.
4. Trade-off between Fit and Sparsity: -The choice of the regularization parameter (\lambda) in Lasso Regression involves a trade-off between model fit and sparsity. -A larger (\lambda) leads to more regularization and sparser models, while a smaller (\lambda) allows the model to fit the training data more closely.
5. Application: -Lasso Regression is commonly used in scenarios where feature selection and interpretation of coefficients are important, such as in high-dimensional datasets with many predictors.

Ordinary least square(OLS) method in Machine Learning

- Linear Regression Model: OLS is primarily applied in linear regression. In linear regression, the goal is to find the best-fitting straight line through the data points.
- Minimization of Residuals: OLS minimizes the sum of the squared differences between the observed and predicted values (residuals). It does this by adjusting the coefficients of the linear regression model to minimize the overall error.
- Mathematical Optimization: OLS employs mathematical optimization techniques to find the coefficients that minimize the sum of squared residuals. This optimization is typically achieved using calculus and linear algebra.
- OLS Assumptions: OLS relies on several assumptions, including linearity, independence of errors, homoscedasticity (constant variance of errors), and normality of errors.
- Application: OLS is widely used in various fields, including economics, finance, engineering, and social sciences, where linear relationships between variables are common.
- Evaluation: After fitting the model using OLS, its performance is evaluated using metrics such as R-squared, which indicates the proportion of the variance in the dependent variable that is predictable from the independent variables.

Advantages: -Simplicity: It is easy to understand and implement. Interpretability: The coefficients have clear meanings, showing how each variable affects the outcome. -Efficiency: It provides efficient and unbiased estimates of the coefficients under certain conditions.

Disadvantages: -Sensitivity to Outliers: Outliers can greatly affect the line. Assumption Violations: If the data does not meet the assumptions, the estimates may be biased. -Limited to Linear Relationships: It only works well when the relationship between variables is linear.

MSE, RMSE, and MAE in detail

-Mean Squared Error (MSE): -Definition: MSE is a measure of the average squared difference between the predicted values and the actual values in a regression problem. -Calculation: It is calculated by taking the average of the squared differences between each predicted value ((\hat{y}*i)) and the corresponding actual value ((y_i)): [\text{MSE} = \frac{1}{n} \sum*{i=1}^{n} (y_i - \hat{y}_i)^2] -Interpretation: MSE gives higher weight to larger errors, making it sensitive to outliers. It is always non-negative, with lower values indicating better model performance.

Root Mean Squared Error (RMSE): -Definition: RMSE is the square root of the MSE, providing a measure of the average magnitude of the errors in the same units as the dependent variable. -Calculation: It is calculated as the square root of the MSE: [\text{RMSE} = \sqrt{\text{MSE}}] -Interpretation: RMSE gives us an interpretable measure of the average error. Like MSE, lower values indicate better model performance, with the advantage of being in the same units as the dependent variable.

Mean Absolute Error (MAE): -Definition: MAE is a measure of the average absolute difference between the predicted values and the actual values. -Calculation: It is calculated by taking the average of the absolute differences between each predicted value ($\hat{y}_i$) and the corresponding actual value (y_i): $$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$ -Interpretation: MAE is less sensitive to outliers compared to MSE because it treats all errors equally. It provides a robust measure of the average prediction error.

Compare the Robustness of MAE, MSE, and RMSE

- Sensitivity to Outliers:

> MAE: MAE is less sensitive to outliers since it considers the absolute differences between predicted and actual values.

> MSE/RMSE: MSE and RMSE are more sensitive to outliers due to the squaring operation, penalizing large errors disproportionately.

- Interpretability:

> MAE: MAE provides an easily interpretable measure of average prediction error.

> MSE/RMSE: MSE and RMSE are less interpretable as they involve squaring errors and taking square roots, respectively.

- Impact of Large Errors:

> MAE: Large errors have a linear impact on MAE.

> MSE/RMSE: Large errors have a quadratic impact on MSE/RMSE, which may inflate the metric significantly.

- Performance Evaluation:

> MAE: Useful when predicting outliers is important and equally penalizing all errors.

> MSE/RMSE: Useful when small errors are preferable, but larger errors can have a significant impact.

-Non-Negativity:

> All three are always non-negative, meaning lower values are better. This helps in comparing models.

-Non-Negativity:

> All three are always non-negative, meaning lower values are better. This helps in comparing models.

In [ ]: