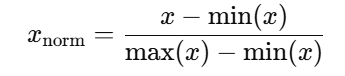
1. **What is Normalization & Standardization and how is it helpful**?

Ans:

**Normalization and Standardization** are two crucial techniques in data preprocessing, particularly beneficial for multiple linear regression models. They address the issue of features having different scales, which can significantly impact the model's performance.

-Normalization: Normalization scales the values of a feature to a specific range, typically [0, 1] or [-1, 1]. It is often done using the **min-max scaling formula** as below:



Ensures all features contribute equally, especially when they have different units or ranges.

Particularly useful when features have minimum and maximum values that vary widely.

- Standardization: Standardization scales the values of a feature to have a mean of 0 and a standard deviation of 1. This is achieved using the **z-score formula**:



Where, μ = Mean of the feature, σ = Standard deviation of the feature.

**How is it helpful?**

* **Improved Convergence:** Gradient descent, a common optimization algorithm used in linear regression, converges faster when features are on a similar scale. Normalization and standardization can help the algorithm reach the optimal solution more quickly.
* **Better Model Interpretability:** When features have vastly different scales, the coefficients of the regression model can be difficult to interpret. Normalization and standardization can make the coefficients more comparable, providing a clearer understanding of the relative importance of each feature.
* **Improved Model Performance:** In some cases, normalization and standardization can lead to improved model accuracy and generalization.

2. What techniques can be used to address multi collinearity in multiple linear regressions?

Ans:

Multicollinearity in multiple linear regression occurs when two or more predictor variables are highly correlated with each other. This can lead to unstable and unreliable coefficient estimates, making it difficult to interpret the individual effects of the predictors on the response variable.

Multicollinearity techniques:

* Feature Removal:
* **Identify and remove one of the highly correlated variables:** If two or more variables are highly correlated, remove one of them from the model. This can be done based on domain knowledge, variable importance, or other criteria.
* **Use a feature selection method:** Techniques like stepwise regression or regularization methods (e.g., LASSO) can automatically select the most important variables and remove those that are highly correlated
* Data Transformation:
* **Combine correlated variables:** Create a new variable that is a combination of the correlated variables, such as a ratio or a weighted average.
* **Principal Component Analysis (PCA):** Transform the original variables into a new set of uncorrelated variables called principal components. These components are linear combinations of the original variables and can be used as predictors in the regression model
* Regularization Techniques:
* **Ridge Regression:** Adds a penalty term to the regression equation that shrinks the coefficients of the predictors, reducing their impact and stabilizing the model.
* **LASSO Regression:** Similar to Ridge Regression, but it can also set some coefficients to zero, effectively performing feature selection.
* Increase Sample Size:
* In some cases, increasing the sample size can help to reduce the impact of multicollinearity. With more data, the model can better estimate the true relationships between the predictors and the response variable.