**Q1** What is normalization and standardization and how is it helpful ?

**Normalization** and **standardization** are two techniques used to transform features (variables) in a dataset so that they are on a comparable scale. They are especially important in machine learning and data analysis to ensure that features with different units or scales do not disproportionately affect model performance.

1. **Normalization** (Min-Max Scaling)

Normalization, also known as **Min-Max Scaling**, transforms the data into a specific range, usually between 0 and 1.

**Benefits of Normalization:**

* Helps when the features are on different scales and you want them to be within the same range.
* Particularly useful in algorithms that rely on distance metrics, like k-Nearest Neighbors (KNN), Support Vector Machines (SVM), and neural networks.
* Ensures that no feature dominates others due to its scale.

**2. Standardization (Z-score Normalization)**

Standardization, also known as **Z-score normalization**, transforms data so that it has a mean of 0 and a standard deviation of 1.

**Benefits of Standardization:**

* Useful when the data has varying scales but is not strictly bound within a specific range.
* Helps in algorithms that assume the data follows a normal distribution, like linear regression, logistic regression, and principal component analysis (PCA).
* Often preferred when the features are not bounded (e.g., income or age).

Q2 What techniques can be used to address multicollinearity in multiple linear regression ?

Multicollinearity occurs when two or more independent variables in a multiple linear regression model are highly correlated, which can lead to unreliable estimates of regression coefficients. There are several techniques to address multicollinearity:

**1 Standardize Variables**: Standardizing variables (e.g., transforming them into z-scores) can help with multicollinearity in some cases, especially if the variables have different scales. However, this does not always resolve the issue if strong collinearity exists.

2 **Ridge Regression (L2 Regularization)**: Ridge regression adds a penalty term to the regression equation that shrinks the coefficients of correlated predictors. This reduces the impact of multicollinearity by making the model more stable.

3 **Lasso Regression (L1 Regularization)**: Similar to ridge regression, lasso regression adds a penalty term but it can also force some coefficients to be exactly zero, effectively removing highly correlated predictors from the model.

4 **Principal Component Regression (PCR)**: This approach involves transforming the original predictor variables into a smaller set of uncorrelated components using PCA. Then, a linear regression is performed on the transformed components.