Interview question

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

Normalisation and standardisation are two commonly used data preparation techniques for scaling numerical characteristics. They make data analysis easier, particularly for algorithms that utilise data range or distribution, such as linear regression, neural networks, and k-nearest neighbours.

Normalization scales data to a fixed range, usually between 0 and 1, or -1 and 1, depending on the chosen method. This is often done with **Min-Max Scaling.**

Standardization transforms data to have a mean of 0 and a standard deviation of 1, following a **z-score transformation.**

Improved Model Convergence: On scaled data, several optimisation methods (such as gradient descent) converge more quickly.
Scale Consistency: Algorithms may be skewed when variables have disparate scales. Scaling guarantees that all features contribute equally.
Reduced Model Sensitivity: Because scaled data minimises sensitivity to initialisation and prevents complications with exploding/vanishing gradients, some models, including neural networks, perform better with it.

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

Multicollinearity occurs in multiple linear regression when two or more predictor variables are highly correlated, making it difficult to estimate the individual effect of each predictor on the dependent variable. Multicollinearity can lead to inflated standard errors for regression coefficients, making them unstable and potentially non-significant.

Multicollinearity can be removed by:

Identifying and removing highly correlated features using Correlation matrix, Variance inflation factors.

Feature engineering and Principal Component Analysis can be used to for reducing the multicollinearity.

Regularizing methods like Lasso and Ridge regression optimises multicollinearity in the model.