Assignment ML Regression (Module 3)

In [4]:
```
## Q1 What is simple Linear Regression?

## Ans: Simple linear regression is a linear regression model that uses a single independent variable to predict a continuous
# dependent variable.

## Q2 What are the key assumptions of Simple Linear Regression?

## Ans: Simple linear regression assumes that the relationship between the independent variable and the dependent variable
# is linear,
# that is, the relationship can be described by a straight line.

## Q3 What does the coefficient m represent in the equation Y = mX + c?

## Ans: The coefficient m represents the slope of the regression line.
# It indicates the change in the dependent variable for a one-unit change in the independent variable.

## Q4 What does the intercept c represent in the equation Y = mX + c?

## Ans: The intercept c represents the point where the regression line intersects the y-axis.
#  It is the value of the dependent variable when the independent variable is zero.

## Q5 How do we calculate the slope m in Simple Linear Regression?

## Ans: The slope m is calculated by finding the slope of the regression line using the formula
# m = (Sum of (X - Xmean) * (Y - Ymean)) / (Sum of (X - Xmean)^2).

## Q6 What is the purpose of the least squares method in Simple Linear Regression?

## Ans: The least squares method is used to find the best-fit line that minimizes
# the sum of the squared residuals between the predicted and actual values of the dependent variable.

## Q7 How is the coefficient of determination (R2) interpreted in simple linear regression?

## Ans: The coefficient of determination (R2) is a measure of the goodness of fit of a regression model.
# It ranges from 0 to 1, where a higher value indicates a better fit.
```

```
## Q8 What is multiple linear regression?

## Ans: Multiple linear regression is a regression model that uses
# more than one independent variable to predict a continuous dependent variable.

## Q9 What is the main difference between Simple and Multiple Linear Regression?

## Ans: Simple linear regression uses a single independent variable to predict a continuous dependent variable,
# while multiple linear regression uses more than one independent variable.

## Q10 What are the key assumptions of multiple linear regression?

## Ans: Multiple linear regression assumes that the relationship between
# the independent variables and the dependent variable is linear, that is, the relationship can be
#  described by a straight line.

## Q11 What is the hetroscedasticity, and how does it affect the results of a Multiple Linear Regression model?

## Ans: Heteroscedasticity refers to the situation where the variance of
# the errors (residuals) is not constant across all levels of the independent variables.
# In multiple linear regression, heteroscedasticity can lead to inefficient estimates and biased standard errors,
# which may result in invalid statistical tests and confidence intervals.

## Q12 How can you improve a multiple linear regression model with high multicollinearity?

## Ans: To improve a multiple linear regression model with high multicollinearity, you can:
## 1. Remove one of the highly correlated independent variables.
## 2. Use a different regression model, such as ridge regression or lasso regression.
## 3. Use a different set of independent variables that are less correlated.

## Q13 What are some common techniques for transforming categorical variables for use in regression models?

## Ans: Some common techniques for transforming categorical variables
# for use in regression models include one-hot encoding, dummy coding, or label encoding.

## Q14 What is the role of interaction terms in Multiple Linear Regression?

## Ans: Interaction terms in multiple linear regression are terms that
# combine two or more independent variables to create a new term
# that has a non-linear relationship with the dependent variable.
```

```
## Q15 How can the interpretation of intercept differ between Simple and Multiple Linear Regression?

## Ans: The interpretation of the intercept differs between simple and multiple linear regression:
## In simple linear regression, the intercept represents
# the expected value of the dependent variable
# when the independent variable is zero.
## In multiple linear regression, the intercept represents
#  the expected value of the dependent variable
# when all independent variables are zero.

## Q16 What is the significance of the slope in regression analysis, and how does it affect perdictions?

## Ans: The slope in regression analysis represents the change
# in the dependent variable for a one-unit change in the independent variable.
# The significance of the slope affects the predictions:
## A positive slope indicates that
# as the independent variable increases, the dependent variable also increases.
## A negative slope indicates that as
#  the independent variable increases, the dependent variable decreases.

## Q17 How does the intercept in a regression model provide context for the relationship between variables?

## Ans: The intercept in a regression model provides
# context for the relationship between variables:
## It indicates the expected value of the dependent
#  variable when all independent variables are zero.
## It provides a baseline for the relationship
# between the independent variables and the dependent variable.

## Q18 What are the limitations of using R2 as a role measure of model performance?

## Ans: Using R2 as a role measure of model performance has some limitations:
## It is sensitive to the scale of the dependent variable.
## It is not appropriate for models with many independent variables.
## It is not appropriate for models with high multicollinearity.

## Q19 How would you interpret a large standard error for a regression coefficient?

## Ans: A large standard error for a regression
# coefficient indicates that the coefficient is unstable
```

```
# and may be sensitive to changes in the data.

## Q20 How can heteroscedasticity be identified in residual plots, and why is it important to address it?

## Ans: Heteroscedasticity can be identified in residual plots
# by looking for patterns where the spread (variance) of the residuals
# increases or decreases as the fitted values increase.
# If the residuals form a funnel shape or show non-constant variance, this indicates heteroscedasticity.
# It is important to address heteroscedasticity because
# it can lead to inefficient estimates, biased standard errors,
# and unreliable hypothesis tests in regression analysis.

## Q21 What does it mean if a Multiple Linear Regression model has a high R2 but low adjusted R2?

## Ans: A Multiple Linear Regression model with
# a high R2 but low adjusted R2 indicates that the model is overfitting the data,
# which can occur when the model is too complex for the available data.

## Q22 Why is it important to scale variables in a Multiple Linear Regression?

## Ans: Scaling variables in a Multiple Linear Regression is important
#  because it can help to normalize the data,
# which can improve the stability and interpretability of the model.

## Q23 What is polynomial regression?

## Ans: Polynomial regression is a regression model that uses
#  a polynomial function of the independent variable(s) to predict the dependent variable.

## Q24 How does polynomial regression differ from linear regression?

## Ans: Polynomial regression differs from linear regression
# in that it uses a polynomial function of the independent variable(s)
# instead of a straight line.

## Q25 When is polynomial regression used?

## Ans: Polynomial regression is used when
#  the relationship between the independent variable(s) and
# the dependent variable is non-linear.
```

```
## Q26 What is the general equation of polynomial regression?

## Ans: The general equation of polynomial regression
# is Y = a0 + a1X + a2X^2 + a3X^3 + ... + anX^n.

## Q27 Can polynomial regression be applied to multiple variables?

## Ans: Yes, polynomial regression can be applied to
# multiple variables by including the polynomial terms
# for each independent variable separately.

## Q28 What are the limitations of polynomial regression?

## Ans: Polynomial regression has some limitations:
## It is not appropriate for models with many independent variables.
## It is not appropriate for models with high multicollinearity.
## It can be sensitive to outliers in the data.

## Q29 What methods can be used to evaluate model fit when selecting the degree of a polynomial?

## Ans: Methods for evaluating model fit when selecting
# the degree of a polynomial include:
## Cross-validation.
## Leave-one-out cross-validation.
## K-fold cross-validation.
## Residual analysis.

## Q30 Why is visualization important in polynomial regression?

## Ans: Visualization is important in polynomial regression because
#  it allows for a better understanding of
# the relationship between the independent variable(s) and the dependent variable.

## Q31 How is Polynomial regression implemented in Python?

## Ans: Polynomial regression in Python can be
#  implemented using libraries such as scikit-learn, statsmodels, or statsmodels.stats.api.
```