**Assignment Code: D-AG-008**

# Supervised Learning: Regression Models and Performance Metrics | **Solution**

**Instructions:** Carefully read each question. Use Google Docs, Microsoft Word, or a similar tool to create a document where you type out each question along with its answer. Save the document as a PDF, and then upload it to the LMS. Please do not zip or archive the files before uploading them. Each question carries 20 marks.

**Total Marks**: 200

**Question 1** : What is Simple Linear Regression (SLR)? Explain its purpose.

**Answer:**

Simple Linear Regression (SLR) is a statistical technique that models the relationship between two variables — one independent variable (predictor) and one dependent variable (response). It aims to find the best-fitting straight line that predicts the dependent variable based on the independent variable.
**Purpose:** The purpose of SLR is to predict the value of one variable using another and to determine the strength and direction of their relationship.

**Question 2**: What are the key assumptions of Simple Linear Regression?

**Answer:**

The key assumptions are:

1. **Linearity:** The relationship between the dependent and independent variable is linear.
2. **Independence:** The observations are independent of each other.
3. **Homoscedasticity:** The variance of residuals (errors) is constant across all levels of the independent variable.
4. **Normality:** The residuals are normally distributed.
5. **No Multicollinearity:** (In multiple regression) predictors are not highly correlated — not applicable for SLR as it has one predictor.

**Question 3**: Write the mathematical equation for a simple linear regression model and explain each term.

**Answer:**

The equation is:

$$Y=\beta 0+\beta 1X+\varepsilon$$

Where:

- Y= Dependent variable (predicted outcome
- X = Independent variable (predictor
- β0= Intercept (value of Y when X = 0)
- β1 = Slope (rate of change in Y for a unit change in X)
- ε = Error term (difference between observed and predicted values)

**Question 4**: Provide a real-world example where simple linear regression can be applied.

**Answer:**

An example is predicting a student's exam score based on the number of study hours. Here, "study hours" is the independent variable and "exam score" is the dependent variable. SLR helps estimate how much the score changes for each additional hour of study.

**Question 5**: What is the method of least squares in linear regression?

**Answer:**

The **method of least squares** finds the best-fitting line by minimizing the sum of the squares of the residuals (differences between observed and predicted values). It ensures that the total error between the data points and the regression line is as small as possible.

**Question 6**: What is Logistic Regression? How does it differ from Linear Regression?

**Answer:**

**Logistic Regression** is used for predicting categorical outcomes (e.g., yes/no, 0/1) rather than continuous ones. It uses the **logistic (sigmoid)** function to estimate probabilities between 0 and 1.

**Difference:**

- Linear regression predicts continuous values, while logistic regression predicts probabilities or binary outcomes.
- Linear regression uses a straight-line equation, while logistic regression uses a logistic curve.

**Question 7**: Name and briefly describe three common evaluation metrics for regression models.

**Answer:**

**Mean Absolute Error (MAE):** Average of the absolute differences between predicted and actual values.

**Mean Squared Error (MSE):** Average of squared differences between predicted and actual values.

**Root Mean Squared Error (RMSE):** Square root of MSE, gives error in same units as the target variable.

**Question 8**: What is the purpose of the R-squared metric in regression analysis?

**Answer:**

**R-squared** measures the proportion of variance in the dependent variable that is explained by the independent variable(s).

It ranges from 0 to 1:

- R2=1R^2 = 1R2=1 means perfect prediction.
- R2=0R^2 = 0R2=0 means the model explains none of the variance.
  It helps evaluate how well the regression model fits the data

**Question 9**: Write Python code to fit a simple linear regression model using scikit-learn and print the slope and intercept.
(Include your Python code and output in the code box below.)

**Answer:**

```python
# Importing necessary libraries
from sklearn.linear_model import LinearRegression
import numpy as np

# Sample data
X = np.array([1, 2, 3, 4, 5]).reshape(-1, 1)
y = np.array([2, 4, 5, 4, 5])

# Create and train the model
model = LinearRegression()
model.fit(X, y)

# Print slope and intercept
print("Slope (Coefficient):", model.coef_[0])
print("Intercept:", model.intercept_)

Output
Slope (Coefficient): 0.6
Intercept: 2.2
```

**Question 10**: How do you interpret the coefficients in a simple linear regression model?

**Answer:**

The **intercept ($\beta_0$)** indicates the predicted value of Y when X = 0.

The **slope ($\beta_1$)** shows how much Y changes for a one-unit increase in X.
If $\beta_1$ is positive, Y increases as X increases; if $\beta_1$ is negative, Y decreases as X increases.