Assignment Code: D-AG-008

Supervised Learning: Regression Models and Performance Metrics

---

Question 1:

What is Simple Linear Regression (SLR)? Explain its purpose.

Answer:
Simple Linear Regression (SLR) is a statistical technique used to study the relationship between two variables — one independent variable (X) and one dependent variable (Y). It helps in predicting the value of Y based on X by fitting a straight line through the data points.
The main purpose of SLR is to model and understand how the dependent variable changes when the independent variable changes.

Question 2:

What are the key assumptions of Simple Linear Regression?

Answer:
The key assumptions of Simple Linear Regression are:

1. **Linearity:** The relationship between X and Y is linear.
2. **Independence:** The residuals (errors) are independent of each other.
3. **Homoscedasticity:** The variance of residuals is constant across all values of X.
4. **Normality:** The residuals are normally distributed.
5. **No Multicollinearity:** Not applicable in SLR since only one independent variable is used.

Question 3:

Write the mathematical equation for a simple linear regression model and explain each term.

Answer:
The equation is:
[
$Y = \beta_0 + \beta_1 X + \varepsilon$
]
Where:

- **Y** = Dependent variable (predicted output)
- **X** = Independent variable (input feature)
- β_0 = Intercept (value of Y when X = 0)
- β_1 = Slope (change in Y for a one-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:
A real-world example is predicting **house prices** based on the **size of the house (in square feet)**.
Here, the house size is the independent variable (X), and the price is the dependent variable (Y). The regression line shows how the price changes as the size increases.

Question 5:

What is the method of least squares in linear regression?

Answer:
The **method of least squares** is a technique used to estimate the best-fitting regression line by minimizing the **sum of the squared differences** between observed and predicted values.
In simple terms, it finds the line that has the smallest possible total error when predicting Y from X.

Question 6:

What is Logistic Regression? How does it differ from Linear Regression?

Answer:
Logistic Regression is a statistical method used for binary classification problems, where the dependent variable is categorical (e.g., Yes/No, 0/1).
Difference:

- Linear Regression predicts **continuous values**, while Logistic Regression predicts **probabilities or classes**.
- Logistic Regression uses the **sigmoid function** to map predictions between 0 and 1.

Question 7:

Name and briefly describe three common evaluation metrics for regression models.

Answer:

1. **Mean Absolute Error (MAE):** Average of the absolute differences between predicted and actual values.
2. **Mean Squared Error (MSE):** Average of the squared differences between predicted and actual values.
3. **R-squared (R^2):** Indicates how well the model explains the variance in the dependent variable (ranges from 0 to 1).

Question 8:

What is the purpose of the R-squared metric in regression analysis?

Answer:
The **R-squared** metric measures how much of the variance in the dependent variable is explained by the independent variable(s).
It indicates the **goodness of fit** of the model.

- An R^2 value close to **1** means a strong relationship and a good fit.
- An R^2 value close to **0** means the model does not explain the variation well.

Question 9:

Write Python code to fit a simple linear regression model using scikit-learn and print the slope and intercept.

Answer:

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

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

# Creating and fitting the model
model = LinearRegression()
model.fit(X, y)

# Printing slope and intercept
print("Slope (β₁):", model.coef_[0])
print("Intercept (β₀):", model.intercept_)
```

Output Example:

Slope (β_1): 0.6
Intercept (β_0): 2.2

Question 10:

How do you interpret the coefficients in a simple linear regression model?

Answer:

- **Intercept (β_0):** Represents the expected value of Y when X = 0.
- **Slope (β_1):** Represents how much Y changes when X increases by one unit. If β_1 is positive, Y increases with X; if negative, Y decreases with X.