Assignment Regression

Simple Linear Regression

Q1: What is Simple Linear Regression?
Simple Linear Regression models the relationship between a dependent variable (Y) and a single independent variable (X) using a linear equation:

Y=mX+c

where mmm is the slope, and ccc is the intercept.

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

1. **Linearity** – The relationship between X and Y is linear.
2. **Independence** – Observations are independent.
3. **Homoscedasticity** – The variance of residuals is constant.
4. **Normality** – Residuals are normally distributed.
5. **No multicollinearity** – Not applicable here since only one predictor is used.

Q3: What does the coefficient mmm represent in Y=mX+c?
It represents the **slope**, indicating the change in Y for a one-unit increase in X

Q4: What does the intercept ccc represent in Y=mX+c?

The value of Y when X=0.

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

$$m = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}$$

Q6: What is the purpose of the least squares method in Simple Linear Regression?
It minimizes the sum of squared differences between observed and predicted values.

Q7: How is the coefficient of determination (R^2) interpreted in Simple Linear Regression?
It measures the proportion of variance in Y explained by X. Values range from 0 (no fit) to 1 (perfect fit).

Multiple Linear Regression

Q8: What is Multiple Linear Regression?
It extends Simple Linear Regression to multiple predictors:

$$Y = b_0 + b_1X_1 + b_2X_2 + ... + b_nX_n$$

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

- **Simple**: One independent variable.
- **Multiple**: Two or more independent variables.

Q10: What are the key assumptions of Multiple Linear Regression?

1. **Linearity**
2. **Independence**
3. **Homoscedasticity**
4. **Normality of residuals**
5. **No multicollinearity**

Q11: What is heteroscedasticity, and how does it affect results?
Heteroscedasticity occurs when residuals have **unequal variance**, leading to unreliable coefficients and hypothesis tests.

Q12: How can you improve a Multiple Linear Regression model with high multicollinearity?

- Remove correlated variables
- Use **Principal Component Analysis (PCA)**
- Use **Regularization techniques** (Lasso, Ridge)

Q13: What are common techniques for transforming categorical variables?

- **One-Hot Encoding**
- **Label Encoding**
- **Ordinal Encoding**

Q14: What is the role of interaction terms in Multiple Linear Regression?
They capture effects when two variables interact and influence Y together.

Q15: How can the interpretation of intercept differ between Simple and Multiple Linear Regression?
In **Simple Regression**, it is the expected Y when X=0.
In **Multiple Regression**, it is Y when **all** Xs are 0, which may not be meaningful.

Q16: What is the significance of the slope in regression analysis?
It shows the expected change in Y for a one-unit change in X.

Q17: How does the intercept provide context for variable relationships?
It sets the baseline value of Y when predictors are 0.

Q18: What are the limitations of using R2 alone?

- It doesn't indicate **causality**.
- It **doesn't penalize overfitting**.

Q19: How would you interpret a large standard error for a regression coefficient?
A large standard error indicates **high variability**, meaning the coefficient estimate is unstable.

Q20: How can heteroscedasticity be identified in residual plots?
Look for a **funnel-shaped** pattern in residual vs. fitted value plots.

Q21: What does it mean if a model has high R2 but low adjusted R2?
It suggests **too many predictors**, leading to overfitting.

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

- Prevents some variables from **dominating**
- Helps in **convergence** for optimization

Polynomial Regression

Q23: What is polynomial regression?
A form of regression where the relationship is modeled as an nnn-degree polynomial.

Q24: How does polynomial regression differ from linear regression?
Polynomial regression captures **non-linear** relationships.

Q25: When is polynomial regression used?
When the data shows **curvature** instead of a straight-line trend.

Q26: What is the general equation for polynomial regression?

$$Y = b_0 + b_1 X + b_2 X^2 + ... + b_n X^n$$

Q27: Can polynomial regression be applied to multiple variables?
Yes, but it increases model complexity.

Q28: What are the limitations of polynomial regression?

- Prone to **overfitting**
- Difficult to **interpret coefficients**

Q29: What methods evaluate model fit for polynomial regression?

- **R^2 and Adjusted R^2**
- **Mean Squared Error (MSE)**
- **Cross-validation**

Q30: Why is visualization important in polynomial regression?
It helps detect **overfitting and curvature** in the data.

Q31: How is polynomial regression implemented in Python?

```
import numpy as np

import matplotlib.pyplot as plt

from sklearn.preprocessing import PolynomialFeatures

from sklearn.linear_model import LinearRegression

X = np.array([1, 2, 3, 4, 5, 6, 7, 8, 9, 10]).reshape(-1, 1)  # Independent variable

y = np.array([2, 5, 9, 15, 23, 35, 49, 67, 89, 115])  # Dependent variable


# Transform X for Polynomial Features (Degree = 2)

poly = PolynomialFeatures(degree=2)

X_poly = poly.fit_transform(X)

# Train Polynomial Regression Model

model = LinearRegression()

model.fit(X_poly, y)


# Predictions

y_pred = model.predict(X_poly)
```

```
# Plot Results

plt.scatter(X, y, color='red', label='Actual Data')

plt.plot(X, y_pred, color='blue', label='Polynomial Fit')

plt.xlabel('X')

plt.ylabel('Y')

plt.legend()

plt.show()


# Print Model Coefficients

print("Intercept:", model.intercept_)

print("Coefficients:", model.coef_)
```