Linear Regression Scenario-Based Questions

April 04, 2025

1 Multiple Choice Questions (MCQs)

1. **Predicting House Prices**
   Scenario: A real estate agent predicts house prices based on size using linear regression with a slope of 50 and an intercept of 20,000. What's the predicted price for a 2,000 sq ft house?

   A) 100,000

   B) 120,000

   C) 150,000

   D) 200,000

   Answer: **B) 120,000**
   Explanation: Using $Y = MX + B$, where $M = 50$, $X = 2,000$, and $B = 20,000$, the price is $50 \times 2,000 + 20,000 = 120,000$. The slope represents dollars per square foot.

2. **Overfitting Analysis**
   Scenario: A data scientist splits the bamboo growth data into training (70%), validation (15%), and test (15%) sets. She fits a degree-8 polynomial regression model, achieving $R^2 = 1.0$ on training but $R^2 = 0.2$ on test data. What does this indicate?

   A) Underfitting due to insufficient model complexity

   B) Overfitting due to excessive model complexity

   C) Optimal fit with good generalizability

D) Incorrect data splitting proportions

Answer: **B) Overfitting due to excessive model complexity**
Explanation: High-degree polynomials (e.g., 8) can overfit, perfectly fitting training data ($R^2 = 1.0$) but failing on test data ($R^2 = 0.2$), as warned in the document's "Overfitting" section.

3. **Store Sales Forecast**
Scenario: A store manager's linear regression for sales has a slope of 20 and an intercept of 500. What's the predicted sales for 10 customers?

A) 600
B) 700
C) 800
D) 900

Answer: **B) 700**
Explanation: $Y = 20 \times 10 + 500 = 700$. Each customer adds \$20 to a base of \$500.

4. **Quiz Score Prediction**
Scenario: A teacher's linear regression for quiz scores has a slope of 15 and an intercept of 25. What's the predicted score for 3 revision hours?

A) 60
B) 70
C) 80
D) 90

Answer: **B) 70**
Explanation: $Y = 15 \times 3 + 25 = 45 + 25 = 70$. Each hour adds 15 points to a base of 25.

5. **Car Rental Cost**
Scenario: A car rental company's linear regression has a slope of 30 and an intercept of 50. What's the cost for a 5-day rental?

A) 150

B) 180

C) 200

D) 230

Answer: C) 200
Explanation: $Y = 30 \times 5 + 50 = 150 + 50 = 200$. The slope is \$30/day, added to a \$50 fee.

6. **Bamboo Growth Prediction**
Scenario: A botanist fits a cubic polynomial regression model to the bamboo growth data using this Python code:

```
import numpy as np
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import PolynomialFeatures

days = np.array([1, 2, 3, 4, 5, 6, 7, 8, 9,
    10]).reshape(-1, 1)
growth = np.array([1, 2, 3, 5, 15, 34, 48, 70, 136,
    185])

poly = PolynomialFeatures(degree=3)
days_poly = poly.fit_transform(days)

model = LinearRegression()
model.fit(days_poly, growth)

predicted_day_11 =
    model.predict(poly.transform(np.array([[11]])))[0]
print(f"Predicted growth on day 11:
    {predicted_day_11:.2f} cm")
```

Question: Based on the cubic model, what is the predicted bamboo growth on day 11?

A) 185 cm

B) 200 cm

C) 265.23 cm

D) 300 cm

Answer: **C) 265.23 cm**
Explanation: The code fits a cubic polynomial ($y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$) and predicts 260.50 cm for day 11, an extrapolation beyond day 10 (185 cm), capturing the accelerating growth trend.

7. **Sales Prediction with Python**
Scenario: A shop owner runs this Python code to predict sales based on ad budget:

```python
import numpy as np
from sklearn.linear_model import LinearRegression

X = np.array([[10], [20], [30], [40], [50]])  # Budget
Y = np.array([150, 300, 450, 600, 750])       # Sales

model = LinearRegression()
model.fit(X, Y)

slope = model.coef_[0]
intercept = model.intercept_

print(f"Slope: {slope}, Intercept: {intercept}")
```

Question: Predicted sales for a \$25 budget, best approximation?

A) 300
B) 375
C) 450
D) 525

Answer: **B) 375**
Explanation: Output: `Slope: 15.0, Intercept: 0.0`
$Y = 15 \times 25 + 0 = 375$. Each dollar increases sales by \$15.

8. **Error in Plant Height Prediction (Difficult)**
Scenario: A researcher runs this code to predict plant height:

```python
import numpy as np
from sklearn.linear_model import LinearRegression

X = np.array([[2], [4], [6], [8]])  # Sunlight hours
Y = np.array([10, 18, 26, 34])      # Height in cm

model = LinearRegression()
model.fit(X, Y)

slope = model.coef_[0]
intercept = model.intercept_

X_test = np.array([[5]])
predicted_height = model.predict(X_test)[0]
actual_height = 25
error = abs(actual_height - predicted_height)

print(f"Slope: {slope}, Intercept: {intercept}")
print(f"Predicted height for 5 hours:
   {predicted_height}")
print(f"Absolute error: {error}")
```

Output:
`Slope: 4.0, Intercept: 2.0`
`Predicted height for 5 hours: 22.0`
`Absolute error: 3.0`
Question: What does the error of 3.0 indicate?

A) Overestimates by 3 cm

B) Underestimates by 3 cm

C) Slope incorrect by 3

D) Intercept needs +3

Answer: **B)**
Explanation: Predicted: 22 cm, Actual: 25 cm. $|25 - 22| = 3$, and 22 ¡ 25, so it underestimates.

2 Numeric Type Questions

9. **Coding Speed Target**
Scenario: A student runs this code for coding speed:

```
import numpy as np
from sklearn.linear_model import LinearRegression

X = np.array([[1], [2], [3], [4], [5]])  # Hours
Y = np.array([10, 25, 40, 55, 70])       #
    Lines/hour

model = LinearRegression()
model.fit(X, Y)

slope = model.coef_[0]
intercept = model.intercept_

print(f"Slope: {slope}, Intercept: {intercept}")
```

Output: `Slope: 15.0, Intercept: -5.0`
Question: Hours needed for 100 lines/hour?
Answer: **7**
Explanation: $100 = 15X - 5$. $105 = 15X$, $X = 7$.
Hint: Solve for X with $Y = 100$.

10. **Quadratic Model Fit**
Scenario: A researcher fits a quadratic polynomial regression model (degree=2) to the bamboo growth data with this code:

```
import numpy as np
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import PolynomialFeatures
from sklearn.metrics import r2_score

days = np.array([1, 2, 3, 4, 5, 6, 7, 8, 9,
    10]).reshape(-1, 1)
growth = np.array([1, 2, 3, 5, 15, 34, 48, 70, 136,
    185])

poly = PolynomialFeatures(degree=2)
days_poly = poly.fit_transform(days)
```

```
11
12 model = LinearRegression()
13 model.fit(days_poly, growth)
14
15 predicted = model.predict(days_poly)
16 r2 = r2_score(growth, predicted)
17 print(f"R2 Score for quadratic model: {r2:.4f}")
```

Question: What is the R^2 score for the quadratic model, rounded to two decimal places?
Answer: **0.98**
Explanation: The R^2 score of 0.9804, rounded to 0.98, indicates a good fit, though a cubic model might better capture the rapid growth curve.

11. **Impressions for Clicks (Difficult)**
Scenario: A marketer runs this code for clicks:

```
1 import numpy as np
2 from sklearn.linear_model import LinearRegression
3
4 X = np.array([[100], [200], [300], [400]])  #
     Impressions
5 Y = np.array([5, 9, 13, 17])                # Clicks
6
7 model = LinearRegression()
8 model.fit(X, Y)
9
10 slope = model.coef_[0]
11 intercept = model.intercept_
12
13 print(f"Slope: {slope}, Intercept: {intercept}")
```

Output: `Slope: 0.04, Intercept: 1.0`
Question: Impressions for 25 clicks?
Answer: **600**
Explanation: $25 = 0.04X + 1$. $24 = 0.04X$, $X = 600$.

3 Multiple Select Questions (MSQs)

12. **Flower Growth Analysis**
Scenario: A gardener's linear regression has a slope of 0.2 and an intercept of 5. Which are true for 10 units of water?

A) Predicted growth is 7 units

B) Slope means 0.2 units growth per water unit

C) Intercept means growth starts at 5 with no water

D) Growth doubles if water doubles to 20 units

Answers: **A, B, C**
Explanation:
- A) $0.2 \times 10 + 5 = 7$, true.
- B) Slope $= 0.2$, true.
- C) Intercept $= 5$, true.
- D) $0.2 \times 20 + 5 = 9$, not 14, false.

13. **Employee Productivity**
Scenario: A manager's linear regression has a slope of 3 and an intercept of 40. Which are correct for 5 training hours?

A) Productivity is 55

B) Each hour boosts productivity by 3

C) No training means productivity is 0

D) Assumes a straight-line relationship

Answers: **A, B, D**
Explanation:
- A) $3 \times 5 + 40 = 55$, true.
- B) Slope $= 3$, true.
- C) Intercept $= 40$, false.
- D) Linear regression assumption, true.

14. **Calories Burned Analysis**
Scenario: A coach runs this code for calories burned:

```python
import numpy as np
from sklearn.linear_model import LinearRegression

X = np.array([[10], [20], [30], [40]])  # Minutes
Y = np.array([100, 180, 260, 340])      # Calories

model = LinearRegression()
model.fit(X, Y)

slope = model.coef_[0]
intercept = model.intercept_

X_test = np.array([[25]])
predicted_calories = model.predict(X_test)[0]

print(f"Slope: {slope}, Intercept: {intercept}")
print(f"Predicted calories for 25 minutes:
   {predicted_calories}")
```

Output:
`Slope: 8.0, Intercept: 20.0`
`Predicted calories for 25 minutes: 220.0`
Question: Which are true?

A) Predicted calories for 25 minutes are 220
B) Each minute burns 8 calories
C) Base burn is 20
D) Prediction for 0 minutes is 0

Answers: **A, C**
Explanation:
- A) Output shows 220, true.
- B) Slope = 8 is increase, not total, false.
- C) Intercept = 20, true.
- D) $8 \times 0 + 20 = 20$, false.

15. **Cubic Model Analysis**
Scenario: A student fits a cubic polynomial regression model to the bamboo growth data:

```python
import numpy as np
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import PolynomialFeatures

days = np.array([1, 2, 3, 4, 5, 6, 7, 8, 9,
    10]).reshape(-1, 1)
growth = np.array([1, 2, 3, 5, 15, 34, 48, 70, 136,
    185])

poly = PolynomialFeatures(degree=3)
days_poly = poly.fit_transform(days)

model = LinearRegression()
model.fit(days_poly, growth)

print(f"Coefficients: {model.coef_}")
print(f"Intercept: {model.intercept_:.2f}")
```

Output:
`Coefficients: [ 0. 7.248446 -2.58566434 0.37801088]`
`Intercept: -4.77`
Question: Which statements are true based on the model and document?

A) The model equation(approximate) is $y = -4.77 + 7.25x + 2.59x^2 + 0.38x^3$

B) The cubic term is positive, indicating accelerating growth.

C) A linear model (degree=1) would perfectly fit this data

D) The R^2 score is likely close to 1 due to the cubic fit

Answers: **B, D**
Explanation:
- A) False: The quadratic term sign is incorrect.
- B) $\beta_3 = 0.38 > 0$, indicating acceleration, true.
- C) A linear model would underfit this data.
- D) Cubic fit is optimal per document, likely $R^2 \approx 1$, true.