Evaluation of Linear Regression Model for CreditScore Prediction

Tran Quoc Thai
ID student: 2370759

April 9, 2025

1 Introduction

To assess the performance of our multiple linear regression model from the previous analysis, we compute two key evaluation metrics:

- **Mean Squared Error (MSE)**: Measures the average squared difference between actual and predicted values.
- **Coefficient of Determination ($\mathbf{R^2}$)**: Measures how well the independent variables explain the variance in the dependent variable.

2 Mean Squared Error (MSE)

The Mean Squared Error is given by:

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 \tag{1}$$

Using the estimated regression equation:

$$\hat{y} = 400 + 4.8x_1 + 3.2x_2 \tag{2}$$

We compute the predicted CreditScore values:
The squared errors are:

Age (x_1)	**Education** (x_2)	**Actual** y	**Predicted** $\hat{y}$
35	16	720	$400 + 4.8(35) + 3.2(16) = 713.6$
28	14	650	$400 + 4.8(28) + 3.2(14) = 662.4$
45	14.57	750	$400 + 4.8(45) + 3.2(14.57) = 724.2$
31	12	600	$400 + 4.8(31) + 3.2(12) = 641.6$
52	18	780	$400 + 4.8(52) + 3.2(18) = 784.8$
29	14	630	$400 + 4.8(29) + 3.2(14) = 667.2$
42	16	710	$400 + 4.8(42) + 3.2(16) = 725.6$
33	12	640	$400 + 4.8(33) + 3.2(12) = 650.4$

Table 1: Actual vs. Predicted CreditScore Values

$$
\begin{aligned}
(720 - 713.6)^2 &= 40.96 \\
(650 - 662.4)^2 &= 153.76 \\
(750 - 724.2)^2 &= 660.84 \\
(600 - 641.6)^2 &= 1722.56 \\
(780 - 784.8)^2 &= 23.04 \\
(630 - 667.2)^2 &= 1382.24 \\
(710 - 725.6)^2 &= 243.36 \\
(640 - 650.4)^2 &= 108.16
\end{aligned}
$$

Summing these,

$$\sum (y_i - \hat{y}_i)^2 = 40.96 + 153.76 + 660.84 + 1722.56 + 23.04 + 1382.24 + 243.36 + 108.16 = 4334.92 \tag{3}$$

Thus, the MSE is:

$$MSE = \frac{4334.92}{8} = 541.865 \tag{4}$$

3 Coefficient of Determination (R^2)

The R^2 value is given by:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \tag{5}$$

First, we calculate the mean of y:

$$\bar{y} = \frac{720 + 650 + 750 + 600 + 780 + 630 + 710 + 640}{8} = 685 \tag{6}$$

Next, we compute $\sum(y_i - \bar{y})^2$:

$$\begin{aligned}
(720 - 685)^2 &= 1225 \\
(650 - 685)^2 &= 1225 \\
(750 - 685)^2 &= 4225 \\
(600 - 685)^2 &= 7225 \\
(780 - 685)^2 &= 9025 \\
(630 - 685)^2 &= 3025 \\
(710 - 685)^2 &= 625 \\
(640 - 685)^2 &= 2025
\end{aligned}$$

Summing these,

$$\sum(y_i - \bar{y})^2 = 1225 + 1225 + 4225 + 7225 + 9025 + 3025 + 625 + 2025 = 28500 \quad (7)$$

Thus, R^2 is:

$$R^2 = 1 - \frac{4334.92}{28500} = 1 - 0.1521 = 0.8479 \quad (8)$$

4 Interpretation of Results

- **MSE**: The MSE of 541.865 suggests that the average squared error in predicting `CreditScore` is relatively small, indicating a good model fit.
- **$\mathbf{R^2}$ value**: An R^2 value of 0.8479 implies that approximately 84.79% of the variance in `CreditScore` is explained by `Age` and `Education`. This indicates a strong linear relationship between the independent variables and the target variable.

5 Conclusion

Based on the evaluation metrics:

- The relatively low MSE confirms that the model predictions are reasonably accurate.
- The high R^2 value suggests that the linear model captures most of the variability in `CreditScore`.
- Therefore, multiple linear regression is an appropriate method for this problem, though further improvements (e.g., incorporating additional features or non-linear transformations) could enhance the model's predictive power.