Logistic Regression for Predicting Risk Level

Tran Quoc Thai
ID student: 2370759

April 9, 2025

1 Logistic Regression Model

Logistic regression models the probability of a binary outcome using the **sigmoid function**:

$$h_{\mathbf{w}}(x) = \frac{1}{1 + e^{-(w_0 + w_1 x_1 + w_2 x_2)}}, \tag{1}$$

where:

- x_1 represents **Age**,
- x_2 represents **CreditScore**,
- w_0, w_1, w_2 are the model parameters.

1.1 Initial Prediction for T_1

The given parameters are:

$$w_0 = 0.5, \quad w_1 = -0.02, \quad w_2 = 0.01.$$

For **test record** T_1 **(Age = 37, CreditScore = 705)**, we compute the linear combination:

$$\begin{aligned} z &= w_0 + w_1 x_1 + w_2 x_2 \\ &= 0.5 + (-0.02 \times 37) + (0.01 \times 705) \\ &= 0.5 - 0.74 + 7.05 \\ &= 6.81. \end{aligned}$$

Applying the **sigmoid function**:

$$h_{\mathbf{w}}(x) = \frac{1}{1+e^{-6.81}}. \tag{2}$$

Since $e^{-6.81} \approx 0.0011$, we obtain:

$$h_{\mathbf{w}}(x) \approx \frac{1}{1+0.0011} = 0.9989. \tag{3}$$

Thus, the model predicts a **high probability (0.9989)** of T_1 being **High Risk**.

2 Computing the Cost Function

The logistic regression **binary cross-entropy loss** is:

$$J(w) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log h_{\mathbf{w}}(x^{(i)}) + (1-y^{(i)})\log(1-h_{\mathbf{w}}(x^{(i)}))\right]. \tag{4}$$

For T_1, assuming the actual risk level is **Low** $(y_1 = 0)$, the cost function simplifies to:

$$\begin{aligned} J(w) &= -\left[(1-0)\log(1-0.9989)\right] \\ &= -\log(0.0011) \\ &= -(-6.81) \\ &= 6.81. \end{aligned}$$

3 Interpretation

- The **high probability** of 0.9989 suggests that the model strongly predicts T_1 as **High Risk**.
- The **cost function value** of **6.81** is relatively high, indicating a poor fit if T_1 is actually Low Risk.
- **Further improvements**, such as feature scaling, regularization, or more training data, could enhance prediction accuracy.

4 Conclusion

The logistic regression model predicts a high risk for T_1, but the high cost function value suggests potential misclassification. This result highlights the importance of model refinement, including improved feature selection, parameter tuning, and potential use of regularization techniques. Addressing these aspects can lead to more reliable predictions and better generalization for unseen data.