Understanding Logistic Regression: A Mathematical Explanation

Mustafa AbdulRazek

October 8, 2024

1 Introduction to Logistic Regression

Logistic regression is a statistical model used for binary classification tasks, where the output variable is categorical, typically 0 or 1. The goal of logistic regression is to predict the probability that the output variable $y = 1$, given a set of input features $x_1, x_2, \ldots, x_n$. Unlike linear regression, logistic regression ensures that the predicted probabilities are between 0 and 1 by using the *sigmoid function*.

2 Odds and Log-Odds

Before we dive into logistic regression, let's review the concepts of *odds* and *log-odds*:

2.1 Odds

The *odds* of an event occurring (e.g., $y = 1$) are defined as the ratio of the probability that the event happens to the probability that it does not happen:

$$\text{odds} = \frac{P(y=1)}{1 - P(y=1)}$$

2.2 Log-Odds (Logit)

The *log-odds* (or *logit*) is the natural logarithm of the odds:

$$\text{log-odds} = \log\left(\frac{P(y=1)}{1 - P(y=1)}\right)$$

Log-odds can take any real value between $-\infty$ and $+\infty$.

3 Logistic Regression Model

In logistic regression, we model the *log-odds* as a *linear combination* of the input features:

$$\log\left(\frac{P(y=1)}{1-P(y=1)}\right) = w_1x_1 + w_2x_2 + \cdots + w_nx_n + b$$

where:

- $w_1, w_2, \ldots, w_n$ are the weights (or coefficients) of the model.
- $x_1, x_2, \ldots, x_n$ are the input features.
- b is the bias (or intercept) term.

This equation shows that the log-odds are modeled as a linear combination of the input features. To obtain the probability, we need to transform the log-odds back to the probability scale.

4 Solving for the Probability

Starting from the log-odds equation:

$$\log\left(\frac{P(y=1)}{1-P(y=1)}\right) = w_1x_1 + w_2x_2 + \cdots + w_nx_n + b$$

4.1 Exponentiation

To remove the logarithm, we exponentiate both sides:

$$e^{\log\left(\frac{P(y=1)}{1-P(y=1)}\right)} = e^{w_1x_1 + w_2x_2 + \cdots + w_nx_n + b}$$

which simplifies to:

$$\frac{P(y=1)}{1-P(y=1)} = e^{w_1x_1 + w_2x_2 + \cdots + w_nx_n + b}$$

4.2 Isolating $P(y=1)$

Next, we solve for $P(y=1)$. Multiplying both sides by $(1-P(y=1))$ gives:

$$P(y=1) = e^{w_1x_1 + w_2x_2 + \cdots + w_nx_n + b} \cdot (1-P(y=1))$$

Distributing e^z, we get:

$$P(y=1) = e^z - P(y=1)e^z$$

where $z = w_1x_1 + w_2x_2 + \cdots + w_nx_n + b$.

Now, move all terms involving $P(y=1)$ to one side:

$$P(y=1)(1+e^z) = e^z$$

Finally, solving for $P(y=1)$ gives:

$$P(y=1) = \frac{e^z}{1+e^z}$$

Substituting $z = w_1x_1 + w_2x_2 + \cdots + w_nx_n + b$, the equation becomes:

$$P(y=1) = \frac{e^{w_1x_1+w_2x_2+\cdots+w_nx_n+b}}{1+e^{w_1x_1+w_2x_2+\cdots+w_nx_n+b}}$$

4.3 Sigmoid Function

This equation can be simplified using the sigmoid function, defined as:

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

Thus, the probability can be expressed as:

$$P(y=1) = \sigma(w_1x_1 + w_2x_2 + \cdots + w_nx_n + b)$$

5 Visualization of the Sigmoid Function

The sigmoid function is crucial in logistic regression because it transforms the linear combination of the input features into a probability between 0 and 1. Below is a plot of the sigmoid function:

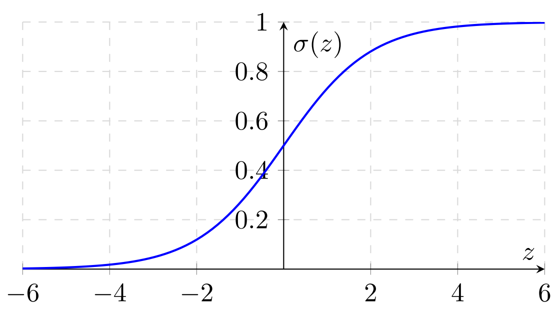

Figure 1: Sigmoid function $\sigma(z) = \frac{1}{1+e^{-z}}$

6 Interpreting the Coefficients

In logistic regression, the coefficients $w_1, w_2, \ldots, w_n$ represent how much the log-odds change with a one-unit change in the corresponding input feature:

$$\log\left(\frac{P(y=1)}{1-P(y=1)}\right) = w_1x_1 + w_2x_2 + \cdots + w_nx_n + b$$

Exponentiating the coefficients gives the odds ratio:

$$e^{w_j}$$

This tells us how the odds change for a one-unit increase in x_j.

For example, if $w_j = 0.5$, then $e^{0.5} \approx 1.65$, meaning the odds are 65% higher for a one-unit increase in x_j.

7 Conclusion

In logistic regression, we model the log-odds of the probability of an event as a linear combination of the input features. We then use the sigmoid function to map the log-odds back to a probability between 0 and 1. This ensures that the model provides valid probability estimates for classification tasks.