

Logistic Regression for Binary Classification

Problem Statement

In binary classification tasks, the goal is to develop a model capable of effectively distinguishing between two classes while providing probabilistic outputs. We aim to propose an optimization problem to minimize the logistic loss function, thereby enabling the modeling of probabilities and the classification of data points into two distinct categories.

Mathematical Formulation

Consider \mathbf{x}_i as the input vector and y_i as the associated label, taking values of either zero or one. Let's choose the sigmoid function $\sigma(z) = \frac{1}{1+e^{-z}}$ as the probability distribution function for the input vector \mathbf{x} , where $z = \mathbf{w}^T \mathbf{x} + b$ for instances with label 1.

Let \hat{y}_i represent the probability of \mathbf{x}^i having label 1 or belonging to group 1. Hence, we can express:

Probability of y_i being 1: \hat{y}_i

Probability of y_i being 0: $1 - \hat{y}_i$

Thus, the probability of y_i being true given the predicted probability \hat{y}_i can be formulated as:

$$P(y_i|\hat{y}_i) = (\hat{y}_i)^{y_i} \times (1 - \hat{y}_i)^{1-y_i}$$

The probability of the model being accurate can be represented by the product of all probabilities over the provided input vectors:

$$P(Model) = \prod_i P(y_i|\hat{y}_i)$$

According to the Maximum Likelihood Principle, we seek to maximize the probability. Therefore, we maximize $\log(P(Model))$ (Maximum Log Likelihood), or equivalently, minimize $-\log(P(Model))$, which can be referred to as the loss function.

Let $L = -\frac{\log(P(Model))}{n}$, where n is the number of input vectors provided. Our main objective is to solve the following optimization problem:

$$\min L = -\frac{1}{n} \sum_i [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

where $\hat{y}_i = \frac{1}{1+e^{-z}}$ and $z = \mathbf{w}^T \mathbf{x} + b$.

If we substitute $\hat{y}_i = \frac{1}{1+e^{-z}}$ into the objective function, we obtain a convex objective function in z , thereby converting the problem into a convex optimization problem.

Summary

Logistic regression offers a powerful framework for binary classification tasks, providing not only a means to classify data points but also probabilistic outputs. By minimizing the logistic loss function, we optimize the model's ability to accurately distinguish between two classes. Through the mathematical formulation described above, we establish the foundation for logistic regression's optimization problem, ultimately enabling the development of effective binary classification models.