Logistic Regression

1 Introduction

Logistic regression is a statistical method for modeling the probability of a binary outcome based on one or more predictor variables. It is widely used in various fields such as medicine, economics, and machine learning for classification tasks.

2 Mathematical Formulation

The logistic regression model relates the probability of the outcome Y to a linear combination of predictor variables through the logistic function.

2.1 Odds and Logit Function

The *odds* of the event $Y = 1$ occurring given predictors X is defined as:

$$\text{Odds}(Y = 1|X) = \frac{P(Y = 1|X)}{1 - P(Y = 1|X)}$$

The *logit* function, which is the logarithm of the odds, is given by:

$$\text{logit}(P(Y = 1|X)) = \ln\left(\frac{P(Y = 1|X)}{1 - P(Y = 1|X)}\right)$$

2.2 Linear Relationship

Logistic regression assumes a linear relationship between the log-odds of the outcome and the predictor variables:

$$\ln\left(\frac{P(Y = 1|X)}{1 - P(Y = 1|X)}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n$$

2.3 Logistic Function

Solving for $P(Y = 1|X)$, we obtain the logistic function:

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n)}}$$

2.4 Interpretation of Coefficients

Each coefficient β_i represents the change in the log-odds of the outcome per unit change in the predictor X_i, holding all other predictors constant.

2.5 Maximum Likelihood Estimation

The parameters $\beta_0, \beta_1, \ldots, \beta_n$ are estimated using the method of maximum likelihood, which maximizes the likelihood function:

$$L(\beta) = \prod_{i=1}^{N} [P(Y_i|X_i)]^{Y_i} [1 - P(Y_i|X_i)]^{1-Y_i}$$

where N is the number of observations, Y_i is the observed outcome, and X_i is the vector of predictor variables for observation i.

2.6 Decision Rule

For classification tasks, a threshold (commonly 0.5) is applied to the predicted probability to determine the class label:

$$\hat{Y} = \begin{cases} 1 & \text{if } P(Y = 1|X) \geq 0.5 \\ 0 & \text{if } P(Y = 1|X) < 0.5 \end{cases}$$

3 Algorithm

The following steps outline the algorithm for fitting a logistic regression model:

1. **Initialize** the parameters $\beta_0, \beta_1, \ldots, \beta_n$ to some initial values (e.g., zeros).
2. **Compute** the predicted probabilities $P(Y = 1|X_i)$ for each observation i using the logistic function:

$$P(Y = 1|X_i) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_n X_{in})}}$$

3. **Calculate** the gradient of the log-likelihood function with respect to each parameter β_j:

$$\frac{\partial \ell(\beta)}{\partial \beta_j} = \sum_{i=1}^{N} (Y_i - P(Y = 1|X_i)) X_{ij}$$

4. **Update** the parameters using gradient ascent (or a similar optimization method):

$$\beta_j \leftarrow \beta_j + \alpha \frac{\partial \ell(\beta)}{\partial \beta_j}$$

   where α is the learning rate.

5. **Repeat** steps 2-4 until convergence (i.e., until the change in the log-likelihood function is below a predefined threshold).
6. **Output** the final parameter estimates $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_n$.