Predicted Probability

The logistic regression model predicts the probability $\hat{p}_i$ that the target $y_i = 1$ given the features $\mathbf{x}_i$.

$$\hat{p}_i = \frac{1}{1 + e^{-(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta})}}$$

- β_0 is the intercept (bias term). - $\mathbf{x}_i$ is the feature vector for the i-th observation. - $\boldsymbol{\beta}$ is the vector of coefficients.

Weight Matrix W

The weight matrix W is a diagonal matrix where each diagonal element W_{ii} represents the variance of the predicted probability for the i-th observation:

$$W_{ii} = \hat{p}_i(1 - \hat{p}_i)$$

Where: - $\hat{p}_i$ is the predicted probability for the i-th observation.

Design Matrix X

The design matrix X includes the features and the intercept. If X originally contains only the features, we add a column of ones to account for the intercept:

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$$

- x_{ij} is the value of the j-th feature for the i-th observation. - p is the number of features. - n is the number of observations.

Hessian Approximation (Covariance Matrix Σ)

The covariance matrix Σ is calculated using the Hessian approximation. It is the inverse of the product of the transposed design matrix, the weight matrix, and the design matrix:

$$\Sigma = (X^\top W X)^{-1}$$

Where: - $X^\top$ is the transpose of the design matrix X. - W is the weight matrix.

Standard Errors

The standard error of each coefficient $\hat{\beta}_j$ is the square root of the corresponding diagonal element of the covariance matrix Σ:

$$\text{SE}(\hat{\beta}_j) = \sqrt{\Sigma_{jj}}$$

Where: - Σ_{jj} is the j-th diagonal element of the covariance matrix Σ.

z-scores

The z-score for each coefficient $\hat{\beta}_j$ is calculated by dividing the coefficient by its standard error:

$$z_j = \frac{\hat{\beta}_j}{\text{SE}(\hat{\beta}_j)}$$

Where: - $\hat{\beta}_j$ is the estimated coefficient. - $\text{SE}(\hat{\beta}_j)$ is the standard error of $\hat{\beta}_j$.

p-values

The p-value for each coefficient is calculated from the z-score using the cumulative distribution function (CDF) of the standard normal distribution:

$$p_j = 2 \cdot \left(1 - \Phi\left(|z_j|\right)\right)$$

Where: - $\Phi(|z_j|)$ is the CDF of the standard normal distribution evaluated at $|z_j|$.

Summary of the Equations

- Predicted Probability: $\hat{p}_i = \frac{1}{1+e^{-(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta})}}$
- Weight Matrix: $W_{ii} = \hat{p}_i(1 - \hat{p}_i)$
- Covariance Matrix: $\Sigma = (X^\top W X)^{-1}$
- Standard Error: $\text{SE}(\hat{\beta}_j) = \sqrt{\Sigma_{jj}}$
- z-score: $z_j = \frac{\hat{\beta}_j}{\text{SE}(\hat{\beta}_j)}$
- p-value: $p_j = 2 \cdot \left(1 - \Phi\left(|z_j|\right)\right)$