Gradient Descent

Jacob Robertson

November 2024

Logistic Regression and Log-Likelihood Function

The sigmoid function is defined as:

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

where

$$z = X \cdot W + b$$

with W being the weights and b the bias.

Log-Likelihood Function

The log-likelihood function $L(W, b)$ measures how well the logistic regression model predicts the actual outcomes. It is defined as:

$$L(W, b) = \sum_{i=1}^{N} \left(y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right)$$

where:

- N is the number of samples,
- y_i is the actual class label (0 or 1),
- $\hat{y}_i$ is the predicted probability of the positive class for the i-th sample, calculated using the sigmoid function.

Gradients of the Log-Likelihood

To optimize the logistic regression model, we compute the gradients of the log-likelihood function with respect to the weights W and bias b:

Gradient with respect to weights W:

The gradient is given by:

$$\frac{\partial L}{\partial W} = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i) X_i$$

This represents how much the log-likelihood changes with a small change in the weights, averaged over all samples.

Gradient with respect to bias b:

The gradient for the bias is:

$$\frac{\partial L}{\partial b} = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)$$

Using Gradients for Optimization

In gradient descent, we use these gradients to update the weights and bias in the opposite direction of the gradient, which is aimed at maximizing the log-likelihood:

Weight Update:

$$W \leftarrow W - \eta \frac{\partial L}{\partial W}$$

Bias Update:

$$b \leftarrow b - \eta \frac{\partial L}{\partial b}$$