Logistic Regression

Introduction

Logistic Regression is a statistical method used for **binary classification**. It models the probability that a given input belongs to a particular category.

Mathematical Formulation

Logistic Function

The core of logistic regression is the logistic function, which is defined as:

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

Here, $\sigma(z)$ is the logistic function, e is the base of the natural logarithm, and z is the input to the function, usually the linear combination of features and weights.

Model Function

In logistic regression, we predict the probability that a given input $\mathbf{x}$ belongs to the default class (class 1), which can be represented as:

$$P(y = 1|\mathbf{x}) = \sigma(\mathbf{w}^T\mathbf{x})$$

- $P(y = 1|\mathbf{x})$ is the probability that input $\mathbf{x}$ belongs to class 1.
- $\mathbf{w}$ is the weight vector.
- $\mathbf{w}^T\mathbf{x}$ represents the dot product of vectors $\mathbf{w}$ and $\mathbf{x}$.

Decision Boundary

The decision boundary for logistic regression is set at 0.5 probability. If $\sigma(\mathbf{w}^T\mathbf{x})$ is greater than or equal to 0.5, the output is classified as class 1; otherwise, it's classified as class 0.

Loss Function

The loss function in logistic regression, also known as the **cross-entropy loss**, measures the performance of the model. It's formulated as:

$$L(\mathbf{w}) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log(\sigma(\mathbf{w}^T\mathbf{x}^{(i)})) + (1-y^{(i)})\log(1-\sigma(\mathbf{w}^T\mathbf{x}^{(i)}))\right]$$

- m is the number of training examples.
- $y^{(i)}$ is the actual label of i-th example.
- $\mathbf{x}^{(i)}$ is the feature vector of the i-th example.
- $L(\mathbf{w})$ is the loss function.

Question: How to understand the cross-entropy loss?

Model Training

Training a logistic regression model involves finding the best $\mathbf{w}$ that minimizes the loss function $L(\mathbf{w})$. This is typically done using optimization algorithms like Gradient Descent.

Gradient Descent

Gradient Descent is used to find the minimum of the loss function. The weights are updated as follows:

$$\mathbf{w} \leftarrow \mathbf{w} - \eta\frac{\partial L(\mathbf{w})}{\partial \mathbf{w}}$$

- η is the learning rate.
- $\frac{\partial L(\mathbf{w})}{\partial \mathbf{w}}$ are the partial derivatives of the loss function with respect to $\mathbf{w}$, respectively.

Conclusion

Logistic Regression is a fundamental machine learning algorithm for binary classification. It's based on the logistic function and uses a log loss loss function. The model's parameters are optimized using algorithms like Gradient Descent. Despite its simplicity, logistic regression can be very effective for linearly separable classes.