Logistic Regression

Abigail Naa Amankwaa Abeo

African Masters in Machine Intelligence (AMMI)

June 12, 2023

Overview

What is Regression and Types?

A regression is a statistical technique that relates a dependent variable to one or more independent (explanatory) variables.

Types:

- Linear
- Polynomial
- Non-Linear
- Ridge
- Lasso
- Elastic net
- Logistic

What is Logistic Regression?

- Logistic Regression is a popular statistical model for binary classification.
- It predicts the probability of an event occurring based on input features.
- It is widely used in various fields, including machine learning, healthcare, finance, and social sciences.

Model Formulation

- Logistic Regression models the relationship between the input features and the probability of the binary outcome.
- The probability of the classes is modeled using the sigmoid function, where,

$$g(z) = g(w^T x) = \frac{1}{1 + e^{-w^T x}}$$

.

Model Formulation

- The decision boundary is defined by a threshold probability.

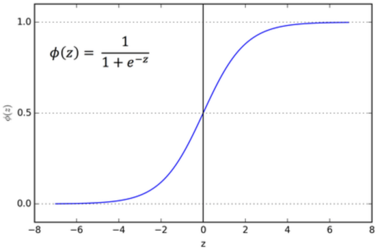

Figure: Graph of Sigmoid Function

Model Formulation

- The model parameters are estimated using maximum likelihood estimation.
  Let us assume that

$$P(y = 1|x; w) = h_w(x)$$

$$P(y = 0|x; w) = 1 - h_w(x)$$

where,

$$h_w(x) = g(w^T x) = \frac{1}{1 + e^{-w^T x}} \tag{3.1}$$

Note that this can be written more compactly as

$$p(y|x; w) = (h_w(x))^y (1 - h_w(x))^{1-y} \tag{3.2}$$

Cross Entropy Loss or Negative log likelihood

Assuming that the n training examples were generated independently, we can then write down the negative log-likelihood of the parameters as

$$L(w) = -\log\left(\prod_{i=1}^{n} p(y^{(i)}|x^{(i)}; w)\right) \tag{3.3}$$

To simplify the derivation, we can rewrite the negative log-likelihood given by (3.3) using the logarithmic properties:

$$L(w) = -\sum_{i=1}^{n} \log\left(p(y^{(i)}|x^{(i)}; w)\right)$$

$$L(w) = -\sum_{i=1}^{n} \log\left(h_w(x^{(i)})^{y^{(i)}}(1 - h_w(x^{(i)}))^{1-y^{(i)}}\right)$$

Cross Entropy Loss or Negative log likelihood

So,

$$L(w) = -\sum_{i=1}^{n} \log \left(y^{(i)} \log h_w(x^{(i)}) + (1 - y^{(i)}) \log(1 - h_w(x^{(i)})) \right) \tag{3.4}$$

Model Training

- The logistic regression model is trained using optimization algorithms like gradient descent.

$$w = w - \alpha \nabla_w l \tag{4.1}$$

- The objective is to minimize the log loss or cross-entropy loss function.
- The model's performance can be evaluated using various metrics such as accuracy, precision, recall.

Example Application

- Let's consider a spam email classification task.
- We can use logistic regression to predict whether an email is spam or not based on its content and other features.
- By training the model on labeled data, it can learn to distinguish between spam and non-spam emails.
- The trained model can then be used to classify new, unseen emails.

References

[1] Andrew Ng, CS229 Lecture Notes.

[2] Shai Shalev-Shwartz & Shai Ben-David, UNDERSTANDING MACHINE LEARNING From Theory to Algorithms.