Open in app

This account is under investigation or was found in violation of the Medium Rules.

uday sai

4 Followers About

Logistic Regression and Softmax Regression

uday sai May 24, 2020 · 4 min read

> Logistic Regression (also called Logit Regression) is commonly used to estimate the probability that an instance belongs to a particular class (e.g., what is the probability that this email is spam?). If the estimated probability is greater than 50%, then the model predicts that the instance belongs to that class (called the positive class, labeled "1"), or else it predicts that it does not (i.e., it belongs to the negative class, labeled "0"). This makes it a binary classifier. Some regression algorithms can be used for classification as well (and vice versa) and Logistic Regression is one amongst these algorithms.

Estimating Probabilities

Open in app

This account is under investigation or was found in violation of the Medium Rules.

Linear Regression m... he logistic is the sigmoid function which results in between 0 and 1. The logistic function equation and S-shaped function is defined as follows:

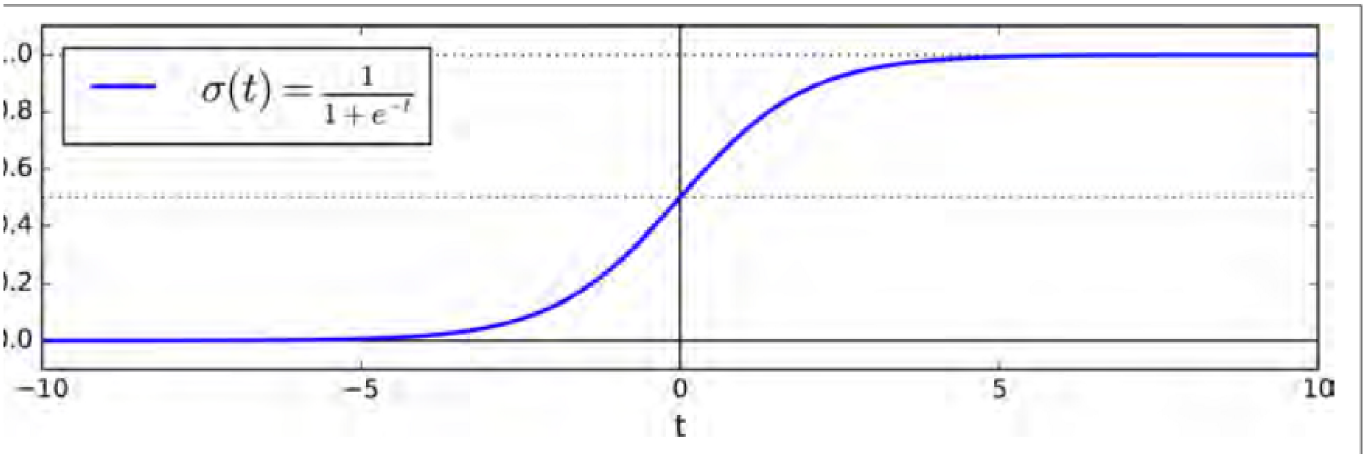

Image : Logistic Regression Equation

Once the Logistic Regression model has estimated the probability p = hθ(x) that an instance x belongs to the positive class, it can make its prediction ŷ easily.

Logistic Regression model prediction

y =0 if p < 0 . 5, 1 if p ≥ 0 . 5 .

Notice that σ(t) < 0.5 when t < 0, and σ(t) ≥ 0.5 when t ≥ 0, so a Logistic Regression model predicts 1 if θ^T· x is positive, and 0 if it is negative.

Training and Cost Function

The objective of training is to set the parameter vector θ so that the model estimates high probabilities for positive instances (y =1) and low probabilities for negative instances (y = 0). This idea is captured by the cost function shown in the below for a single training instance x.

Equation , Cost function of a single training instance

c θ =− log p if y = 1,− log 1 − p if y = 0 .

Open in app

This account is under investigation or was found in violation of the Medium Rules.

instance, and it will also be very large if the model estimates a probability close to 1 for a negative instance. The bad news is that there is no known closed-form equation to compute the value of θ that minimizes this cost function (there is no equivalent of the Normal Equation). But the good news is that this cost function is convex, so Gradient Descent (or any other optimization algorithm) is guaranteed to find the global minimum (if the learning rate is not too large and you wait long enough).

Softmax Regression

The Logistic Regression model can be generalized to support multiple classes directly, without having to train and combine multiple binary classifiers (as discussed in Chapter 3). This is called Softmax Regression, or Multinomial Logistic Regression. The idea is quite simple: when given an instance x, the Softmax Regression model first computes a score sk(x) for each class k, then estimates the probability of each class by applying the softmax function (also called the normalized exponential) to the scores. The equation to compute sk(x) should look familiar, as it is just like the equation for Linear Regression prediction.

Once you have computed the score of every class for the instance x, you can estimate the probability pk that the instance belongs to class k by running the scores through the softmax function it computes the exponential of every score, then normalizes them (dividing by the sum of all the exponentials). The Softmax Regression classifier predicts only one class at a time (i.e., it is multiclass, not multioutput) so it should be used only with mutually exclusive classes such as different types of plants. You cannot use it to recognize multiple people in one picture. Now that you know how the model estimates probabilities and makes predictions, let's take a look at training. The objective is to have a model that estimates a high probability for the target class (and consequently a low probability for the other classes). Minimizing the cost function, called the cross entropy, should lead to this objective because it penalizes the model when it estimates a low probability for a target class. Cross entropy is frequently used to measure how well a set of estimated class probabilities match the target classes (we will use it again several times in the following chapters).

Open in app

This account is under investigation or was found in violation of the Medium Rules.

to switch it to Softm... solver that supports Softmax Regression, such as the "lbfgs" solver (see Scikit-Learn's documentation for more details). It also applies $\ell 2$ regularization by default, which you can control using the hyperparameter C.

References:

udaysai50/Full-Stack-Data-Science

Permalink Dismiss GitHub is home to over 50 million developers working together to host and review code, manage...

github.com

Hands-On Machine Learning with Scikit-Learn and TensorFlow by Aurélien Géron

Data Science Machine Learning Logistic Regression Softmax Python

About Help Legal

Get the Medium app

Download on the App Store

GET IT ON Google Play