Maximum likelihood in logistic regression

Created by Donglai Ma, October 2020

Goal

This document is help you understand how people came up the cost function of logistic regression:

$J(\theta)=-\frac{1}{m} \sum_{i=1}^{m}\left[y^{(i)} \log \left(h_{\theta}\left(x^{(i)}\right)\right)+\left(1-y^{(i)}\right) \log \left(1-\left(h_{\theta}\left(x^{(i)}\right)\right)\right]\right.$

Function set

Imagine a 2-class problem. C1 and C2

The purpose of logistic regression is that we want to find the probability of which class it should be for a given x.

So for a given x, we want to find$P_{w,b}\left(C_1 \mid x\right)$, if $P_{w,b}\left(C_1 \mid x\right)>0.5$, the result is C1, otherwise, output C2 .

A good way to change output to 0~1 range is sigmoid function.

$P_{w,b}\left(C_1 \mid x\right)=\sigma(z)$
$\sigma(z)=\frac{1}{1+\exp(-z)}$

$z=w \cdot x+b=\sum_{i} w_{i} x_{i}+b$

Now you got your function $f_{w,b}(x)=\sigma\left(\sum_{i} w_{i} x_{i}+b\right)$ whose output is between 0 and 1

Likelihood of the function

From the function set we could see that $f_{w,b}(x)=P_{w,b}\left(C_1 \mid x\right)$

And we got a lot of training data

$$\begin{array}{lllll} x^1 & x^2 & x^3 & \cdots & x^N \\ C_1 & C_1 & C_2 & \cdots & C_1 \end{array}$$

Assume there is a real probability in this world, and our training data is generated from this probability $P_{w,b}\left(C_1 \mid x\right)$
So what is the probability of generating this whole training dataset?

$L(w,b)=P_{w,b}\left(C_1 \mid x^1\right) P_{w,b}\left(C_1 \mid x^2\right)\left(1-P_{w,b}\left(C_1 \mid x^3\right)\right) \cdots\left(P_{w,b}\left(C_1 \mid x^N\right)\right)$
$=f_{w,b}\left(x^1\right) f_{w,b}\left(x^2\right)\left(1-f_{w,b}\left(x^3\right)\right) \cdots f_{w,b}\left(x^N\right)$

So the most likely w* and b* is the one with the largest L(w,b)

$w^*, b^*=\arg\max_{w,b} L(w,b)$

Multiply operation is easy to cause overflow so we write this into log liklihood

$w^*, b^*=\arg\min_{w,b}-\ln L(w,b)$

so

$$
\begin{aligned}
& -\ln L(w, b)=\ln f_{w, b}\left(x^1\right)+\ln f_{w, b}\left(x^2\right)+\ln \left(1-f_{w, b}\left(x^3\right)\right) \cdots \\
& \hat{y}^n: 1 \text { for class } 1,0 \text { for class } 2 \\
& =\sum_n-\left[\hat{y}^n \ln f_{w, b}\left(x^n\right)+\left(1-\hat{y}^n\right) \ln \left(1-f_{w, b}\left(x^n\right)\right)\right]
\end{aligned}
$$