predictive model project

January 21, 2020

1 Introduction

In this predictive model, gradient descent optimization algorithm is used to minimize the sum of residuals $R(\alpha, \beta) := \sum_{i=1}^{N} L(f(x_i|\alpha, \beta), y_i)$.

1.1 Derivative of Loss function with respect to parameters

Assuming we have train samples $(x_1, y_1), \ldots, (x_N, y_N)$ and loss function $L(y, y') := |y - y'|^a$, where $\hat{y} = f(x|\alpha, \beta)$. And $f(x|\alpha, \beta) = \beta X + \alpha$, Then the prediction of $\hat{y}$ can be written as

$$\hat{y} = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

letting $\alpha = \beta_0 x_0$, then we can rewrite $\hat{y}$ as below

$$\begin{aligned} \hat{y} &= \beta_0 x_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \qquad (1) \\ x_0 &= 1 \end{aligned}$$

.

equation (1) can be written in matrix form as

$$\hat{y} = \beta^T X \qquad (2)$$

$$L(y, \beta^T X) := |y - \beta^T X|^a \qquad (3)$$

We need to minimize the error in equation (3) w.r.t β.

$$\frac{\partial L(y, \beta^T X)}{\partial \beta} = a|y - \beta^T X|^{a-1} \ \frac{\partial |y - \beta^T X|}{\partial \beta} \qquad (4)$$

we know that given $y = |x|$, $\frac{dy}{dx} = \frac{x}{|x|}$ then equation (4) becomes

$$\begin{aligned} \frac{\partial L(y, \beta^T X)}{\partial \beta} &= a|y - \beta^T X|^{a-1} \ \frac{\partial (|y - \beta^T X|)}{\partial \beta} \qquad (5) \\ &= a|y - \beta^T X|^{a-1} \ . \frac{(y - \beta^T X)}{|y - \beta^T X|}(-X) \\ &= -aX \frac{|y - \beta^T X|^a}{(y - \beta^T X)^2}(y - \beta^T X) \text{ —— This simplifies to} \\ &= -aX \frac{|y - \beta^T X|^a}{(y - \beta^T X)} \end{aligned}$$

when $a > 2$ and $y - \beta^T X$ is large, the derivative of the cost function goes to infinity, which means there is no learning. So, from my analysis, I highly recommend using a value of a between [1,3)