

Stage - 1 - Linear Model

Author: hdup

Email: huangdan@youhujia.com

Remember Following Core Concepts

1. Model (Task)
2. Cost (Performance)
3. Optimize (Experience)

Linear Model

(A). Linear Regression

1), Model

For a single input data X

$$h(x) = W^T X + b \quad (1)$$

2), Cost

Say, MSE (Mean Square Error) cost function

$$Cost|_{w,b}(X, y) = \frac{1}{2m} \sum_{i=1}^m (h_i - y_i)^2 \quad (2)$$

Here, we have m samples

3), Optimize

To leverage Gradient Descent algorithm, we need find out the derivative form of Cost function.

First, we are minimizing function of parameters. That means for given data $\{X, y\}$, we need to find out $\{W, b\}$, so that $Cost$ reaches a value as small as possible.

$$Cost|_{x,y}(W, b) = \frac{1}{2m} \sum_{i=1}^m (h_i - y_i)^2 \quad (3)$$

Let's see what the partial derivatives are.

For w

$$\frac{\partial Cost}{\partial w} = \frac{\partial Cost}{\partial h} \cdot \frac{\partial h}{\partial w} \quad (4)$$

For b

$$\frac{\partial Cost}{\partial b} = \frac{\partial Cost}{\partial h} \cdot \frac{\partial h}{\partial b} \quad (5)$$

First, w.r.t. $\frac{\partial Cost}{\partial h}$

$$\frac{\partial Cost}{\partial h} = 2 \cdot \frac{1}{2m} \cdot \sum_{i=1}^m (h_i - y_i) = \frac{1}{m} \cdot \sum_{i=1}^m (h_i - y_i) \quad (6)$$

Next, w.r.t. $\frac{\partial h}{\partial w}$ and $\frac{\partial h}{\partial b}$

$$\frac{\partial h}{\partial w} = \frac{\partial (wx + b)}{\partial w} = x \quad (7)$$

$$\frac{\partial h}{\partial b} = \frac{\partial (wx + b)}{\partial b} = 1 \quad (8)$$

Combine them together

$$\frac{\partial Cost}{\partial w} = \frac{1}{m} \cdot \sum_{i=1}^m (h_i - y_i) x_i \quad (9)$$

$$\frac{\partial Cost}{\partial b} = \frac{1}{m} \cdot \sum_{i=1}^m (h_i - y_i) \quad (10)$$

Finally

$$\Delta w = \alpha \cdot \frac{\partial Cost}{\partial w} \quad (11)$$

$$\Delta b = \alpha \cdot \frac{\partial Cost}{\partial b} \quad (12)$$

Update $\{w, b\}$ by

$$w = w - \Delta w \quad (13)$$

$$b = b - \Delta b \quad (14)$$

Where α is learning rate

(A). Logistic Regression

Remember, logistic regression is just a generalized linear model

1), Model

To handle binary classification problems (True or False), a simple linear model is not a good choice, we need to generalize the $-\infty$ to $+\infty$ linear space into a better space like 0 to 1

$$h_{\text{generalized}}|_{W,b}(X) = g(W^T X + b) \quad (15)$$

$$\text{Sigmoid}(a) = \frac{1}{1 + e^{-a}} \quad (16)$$

$$\text{Let: } g(x) = \text{Sigmoid}(x) \quad (17)$$

$$\text{We have: } h_g(X) = \frac{1}{1 + e^{-(W^T X + b)}} \quad (18)$$

2), Cost

Say, true values are 0 and 1

$$\text{Cost}|_{y=1} = -\ln(h) \quad (19)$$

$$\text{Cost}|_{y=0} = -\ln(1 - h) \quad (20)$$

Combine them together:

$$\text{Cost}|_{\log} = -\frac{1}{m} \cdot \sum_{i=1}^m y_i \ln(h_i) + (1 - y_i) \ln(1 - h_i) \quad (21)$$

3), Optimize

Similar to linear regression, we want to leverage gradient descent.

Note:

$$\frac{\partial Cost}{\partial h} = -\frac{1}{m} \cdot \sum_{i=1}^m \left(\frac{y_i}{h_i} - \frac{(1-y_i)}{1-h_i} \right) \quad (22)$$

Let's mark:

$$h(a) = Sigmoid(a) \quad (23)$$

$$a(x) = wx + b \quad (24)$$

$$\frac{\partial h}{\partial a} = \frac{\partial Sigmoid}{\partial a} \quad (25)$$

$$\frac{\partial h}{\partial a} = -1 \cdot (1 + e^{-a})^{-2} \cdot e^{-a} \cdot (-1) = \frac{1 + e^{-a} - 1}{(1 + e^{-a})^{-2}} = h(a)(1 - h(a)) \quad (26)$$

So that:

$$\frac{\partial h}{\partial w} = \frac{\partial h}{\partial a} \cdot \frac{\partial a}{\partial w} = h(1 - h)x \quad (27)$$

$$\frac{\partial h}{\partial b} = \frac{\partial h}{\partial a} \cdot \frac{\partial a}{\partial b} = h(1 - h) \quad (28)$$

$$\frac{\partial Cost}{\partial w} = \frac{\partial Cost}{\partial h} \cdot \frac{\partial h}{\partial w} = -\frac{1}{m} \cdot \sum_{i=1}^m \left(\frac{y_i}{h_i} - \frac{(1-y_i)}{1-h_i} \right) \cdot h_i(1 - h_i)x_i \quad (29)$$

$$\frac{\partial Cost}{\partial w} = -\frac{1}{m} \cdot \sum_{i=1}^m (x_i y_i (1 - h_i) - x_i h_i (1 - y_i)) = \frac{1}{m} \cdot \sum_{i=1}^m (h_i - y_i)x_i \quad (30)$$

$$\frac{\partial Cost}{\partial b} = \frac{1}{m} \cdot \sum_{i=1}^m (h_i - y_i) \quad (31)$$