

# 1 Simple Regression Model

## 1.1 Linear

### 1.1.1 One Dimension

Each instance has one attribute  $x$  corresponds to one label  $y$ .

$$y = wx + b, \quad y, x, w, b \in \mathbb{R}$$

### 1.1.2 Multiple Dimensions

Each instance has multiple attributes  $x_1, \dots, x_n$  and multiple labels  $y_1, \dots, y_m$ .

$$\begin{aligned} \mathbf{y} &= \mathbf{w}\mathbf{x} + \mathbf{b}, \\ \mathbf{x} &\in \mathbb{R}^n, \\ \mathbf{w} &\in \mathbb{R}^{m \times n}, \\ \mathbf{y}, \mathbf{b} &\in \mathbb{R}^m, \\ \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} &= \begin{pmatrix} w_{11} & \dots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{m1} & \dots & w_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \end{aligned}$$

## 1.2 Nonlinear(Polynomial Model)

### 1.2.1 One Dimension

Each instance has one attribute  $x$  corresponds to one label  $y$ .

$$\begin{aligned} y &= w_2x^2 + w_1x + b, \\ &\vdots \\ y &= w_kx^k + w_{k-1}x^{k-1} + \dots + w_1x + b \end{aligned}$$

### 1.2.2 Multiple Dimensions

Each instance has multiple attributes  $x_1, \dots, x_n$  and multiple labels  $y_1, \dots, y_m$ .

$$\begin{aligned} \mathbf{y} &= \mathbf{w}_2\mathbf{x}^2 + \mathbf{w}_1\mathbf{x} + \mathbf{b}, \\ &\vdots \\ \mathbf{y} &= \mathbf{w}_k\mathbf{x}^k + \mathbf{w}_{k-1}\mathbf{x}^{k-1} + \dots + \mathbf{w}_1\mathbf{x} + \mathbf{b} \end{aligned}$$

where

$$\mathbf{x}^p = \begin{pmatrix} x_1^p \\ \vdots \\ x_n^p \end{pmatrix}, \quad \mathbf{w}_p \in \mathbb{R}^{m \times n}, \quad \mathbf{y}, \mathbf{b} \in \mathbb{R}^m$$

## 2 Logistic Regression

Put the function of linear regression into sigmoid function, the output value will lie in  $(0, 1)$ .

$$f_{w,b}(x) = \sigma(w \cdot x + b) = \sigma \left( \sum_i w_i x_i + b \right)$$

In the training set  $\{(x_k, \hat{y}_k)\}_k$ ,  $\hat{y}_k \in \{0, 1\}$ . 1 for class  $C_1$ , 0 for class  $C_2$ . If  $(x_1, x_2, x_3, \dots)$  corresponds to  $(1, 1, 0, \dots)$ . The loss function

$$L(w, b) = f_{w,b}(x_1)f_{w,b}(x_2)(1 - f_{w,b}(x_3)) \dots, \quad w^*, b^* = \arg \max_{w, b} L(w, b)$$

Note that

$$w^*, b^* = \arg \min -\ln L(w, b)$$

And

$$\begin{aligned} -\ln L(w, b) &= -\ln f_{w,b}(x_1) - \ln f_{w,b}(x_2) - \ln(1 - f_{w,b}(x_3)) \dots \\ &= \sum_k -[\hat{y}_k \ln f_{w,b}(x_k) + (1 - \hat{y}_k) \ln (1 - f_{w,b}(x_k))] \end{aligned}$$

The relation in the brackets [ ] is called the cross entropy between two Bernoulli distribution.

### 2.1 Comparison with Linear Regression

Simple computation shows that

$$\frac{\partial -\ln L(w, b)}{\partial w_i} = \sum_k -(\hat{y}_k - f_{w,b}(x_k))x_{k,i}$$

So  $w_i$  will update in the same way with linear regression. The only difference between them is the range of output. Logistic regression lies in  $(0, 1)$  while linear regression can be any real number.