Simple Regression Model

Linear

One Dimension

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

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

Multiple Dimensions

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

$$\begin{aligned} &\boldsymbol{y} = \boldsymbol{w}\boldsymbol{x} + \boldsymbol{b}, \\ &\boldsymbol{x} \in \mathbb{R}^n, \\ &\boldsymbol{w} \in \mathbb{R}^{m \times n}, \\ &\boldsymbol{y}, \boldsymbol{b} \in \mathbb{R}^m, \end{aligned}$$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} w_{11} & \dots & w_{1_n} \\ \vdots & \ddots & \vdots \\ w_{m1} & \dots & w_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$$

Nonlinear(Polynomial Model)

One Dimension

Each instance has one attribute x correponds to one label y.

$$\begin{aligned} y &= w_2 x^2 + w_1 x + b, \\ &\vdots \\ y &= w_k x^k + w_{k-1} x^{k-1} + \dots + w_1 x + b \end{aligned}$$

Multiple Dimensions

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

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

where

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

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 $(x_1, x_2, x_3, ...)$ corresponds to $(1, 1, 0, ...)$. The loss function

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

Note that

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

And

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

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

Comparison with Linear Regression

Simple computation shows that

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

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.