logistic regression

logistic distribution X

$$F(x) = P(X \leq x) = \frac{1}{1+e^{-(x-\mu)/\gamma}}$$ 分布函数

$$f(x) = F'(x) = \frac{e^{-(x-\mu)/\gamma}}{\gamma(1+e^{-(x-\mu)/\gamma})^2}$$ 密度函数

1、二项 logistic Regression Model

$X \in R^n \rightarrow Y \in \{0,1\}$ 二分类

$$P(Y=1|x) = \frac{\exp(w \cdot x + b)}{1+\exp(w \cdot x + b)}$$

$$P(Y=0|x) = \frac{1}{1+\exp(w \cdot x + b)}$$

$w \in R^n$ weight

$b \in R$ bias

$x \in R^n$

or. $\Rightarrow$ $w \cdot x + b = w \cdot x$, $w = (w, b)$

rewrite $x = (x, 1)$

$$\Rightarrow \begin{cases} P(Y=1|x) = \dfrac{\exp(w \cdot x)}{1+\exp(w \cdot x)} \\ P(Y=0|x) = \dfrac{1}{1+\exp(w \cdot x)} \end{cases}$$

假如某事件发生的概率为 p，则该事件的 odds 为：

$$\text{odds}(p) = \frac{p}{1-p}$$

该事件的 log odds / logit 为

$$\text{logit}(p) = \log \frac{p}{1-p}$$

而对 $P(Y=1|x)$ 而言 $\boxed{\text{logit}(p(Y=1|x)) = w \cdot x}$

极大似然估计：$\prod_{i=1}^{N} [\pi(x_i)]^{y_i} [1-\pi(x_i)]^{1-y_i}$

设 $P(Y=1|x) = \pi(x)$, $P(Y=0|x) = 1-\pi(x)$

则：

$$L(w) = \sum_{i=1}^{N} [y_i \log \pi(x_i) + (1-y_i) \log(1-\pi(x_i))]$$

$$= \sum_{i=1}^{N} \left[y_i \log \frac{\pi(x_i)}{1-\pi(x_i)} + \log(1-\pi(x_i)) \right]$$

$$= \sum_{i=1}^{N} [y_i (w \cdot x_i) - \log(1+\exp(w \cdot x_i))]$$

$$\frac{\partial L(w)}{\partial w} = \sum_{i=1}^{N} \left[(y_i \cdot x_i) - \frac{x_i \exp(w \cdot x_i)}{1+\exp(w \cdot x_i)} \right]$$

update $w \to w' = w + \alpha \frac{\partial L(w)}{\partial w}$ (极大值)

randomly choose one direction x^i

2. 多项 Logistic Regression Model , Softmax

$$X \in \mathbb{R}^n \longrightarrow Y \in \{1, 2, 3, \cdots K\}$$

$$P(Y=j|x) = \frac{\exp(w_j \cdot x)}{\sum_{i=1}^{K} \exp(w_i \cdot x)} \quad j=1,2,\cdots K$$

引入指示性函数 $1\{\cdot\}$ $\begin{cases} 1\{\text{值为真}\}=1 \\ 1\{\text{值为假}\}=0 \end{cases}$

则 $$L(w) = -\sum_{i=1}^{N} \sum_{j=1}^{K} 1\{y^i=j\} \log \frac{\exp(w_j \cdot x_i)}{\sum_{l=1}^{K} \exp(w_l \cdot x_i)}$$

极小值

所以引入权重衰减 (weight decay)

$$L'(w) = L(w) + \frac{\lambda}{2} \sum_{i=1}^{K} \sum_{j=0}^{n} w_{ij}^2 \qquad w_i \in R^{n+1}$$

$\Rightarrow$ 极小值 $$\nabla_{w_j} L'(w) = -\sum_{i=1}^{N} \left[x_i \left(1\{y^i=j\} - \frac{\exp(w_j \cdot x_i)}{\sum_{l=1}^{K} \exp(w_l \cdot x_i)} \right) \right] + \lambda w_j$$

$\lambda > 0$

$$w_j' = w_j - \alpha \nabla_{w_j} L'(w)$$

randomly pick $\bar{i}$