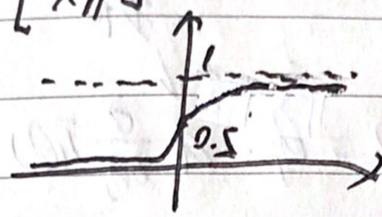
逻辑回归

$h_\theta(x) = g(\theta^T x)$

z

$z = [\theta_0 \ \theta_1 \cdots \theta_n] \cdot \begin{bmatrix} x_0 \\ \vdots \\ x_n \end{bmatrix} = \theta^T x$

$g(z) = \dfrac{1}{1+e^{-z}}$



$h_\theta(x) = g(\theta^T x) = \dfrac{1}{1+e^{-\theta^T x}}$

→ 输出为1的概率

$\begin{cases} h(x) > 0.5 & \text{正样本} \\ h(x) < 0.5 & \text{负样本} \end{cases}$

代价函数:

$Cost(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & y=1 \\ -\log(1-h_\theta(x)) & y=0 \end{cases}$

$J(\theta) = \dfrac{1}{m} \sum_{i=1}^{m} Cost(h_\theta(x^{(i)}), y^{(i)})$

$Cost(h_\theta(x), y) = -y\log(h_\theta(x)) - (1-y)\log(1-h_\theta(x))$

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1-y^{(i)}) \log(1-h_\theta(x^{(i)}))\right]$$

求解 $\min_{\theta} J(\theta)$

梯度下降:

$$\theta_j = \theta_j - \overset{\text{步长}}{\alpha} \cdot \frac{\partial}{\partial \theta_j} J(\theta)$$

$$\frac{\partial}{\partial \theta_j} J(\theta) = \frac{\partial J(\theta)}{\partial g(\theta^T x)} \cdot \frac{\partial g(\theta^T x)}{\partial \theta^T x} \cdot \frac{\partial \theta^T x}{\partial \theta_j}$$

$$\frac{\partial J(\theta)}{\partial g(\theta^T x)} = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)} \cdot \frac{1}{g(\theta^T x)} + (y^{(i)}-1) \cdot \frac{1}{1-g(\theta^T x)} \right]$$

$$\frac{\partial g(\theta^T x)}{\partial \theta^T x} = g'(z) = \left(\frac{1}{1+e^{-z}}\right)' = \frac{e^{-z}}{(1+e^{-z})^2} = \frac{1}{1+e^{-z}} \cdot \left(1 - \frac{1}{1+e^{-z}}\right)$$

$$= g(z) \cdot (1-g(z))$$

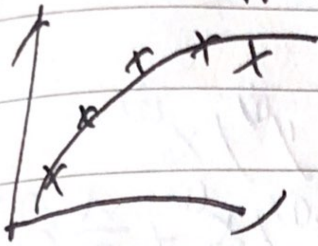
$$= g(\theta^T x) \cdot (1-g(\theta^T x))$$

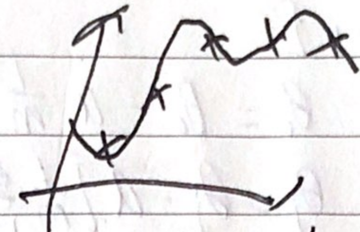
$$\frac{\partial \theta^T x}{\partial \theta_j} = \frac{\partial(\theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n)}{\partial \theta_j} = x_j$$

$$\frac{\partial}{\partial \theta_j} J(\theta) = \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

$$\theta_j = \theta_j - \alpha \cdot \frac{1}{m}\sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) \cdot x_j^{(i)}$$

过拟合解决方案





$\theta_0+\theta_1x+\theta_2x^2$

$\theta_0+\theta_1x+\theta_2x^2+\theta_3x^3+\theta_4x^4$

penalize and make θ_3, θ_4 really small

$$\min \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)})-y^{(i)}\right)^2+1000\theta_3^2+1000\theta_4^2$$

$\theta_3\approx 0$ $\theta_4\approx 0$

Regularization：减轻过拟合

small value for $\theta_0, \theta_1 \cdots \theta_n$.

$$J(\theta)=\frac{1}{2m}\left[\sum_{i=1}^{m}\left(h_\theta(x^{(i)})-y^{(i)}\right)^2+\lambda\sum_{i=1}^{n}\theta_i^2\right]$$

θ_0不惩罚

regularization parameter

λ如果很大，$\rightarrow$ $\theta_1 \cdots \theta_n \approx 0$

$h_\theta(x)=\theta_0$

造成欠拟合

梯度下降

repeat {

$$\theta_0 = \theta_0 - a \cdot \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_0^{(i)}$$

$$\theta_j = \theta_j - a \cdot \left[\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)} + \frac{\lambda}{m}\theta_j\right]$$

}

$$\downarrow$$

$$\theta_j(1 - a\frac{\lambda}{m}) - a \cdot \frac{1}{m} \cdots$$

$1 - a\frac{\lambda}{m} < 1$ 只小一点 e.g. 0.99

normal equation

$$\theta = (X^T X + \lambda L)^{-1} X^T y$$

$$L = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}_{(n+1)\times(n+1)}$$

$X^T X + \lambda L$ 保证可逆