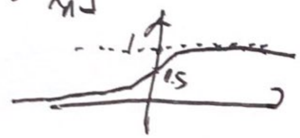
$$h(x) = g(\theta^T x)$$

$$\Downarrow$$

$$z = [\theta_0\ \theta_1\ \dots\ \theta_n] \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} = \theta^T x$$

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



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

$\downarrow$

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：$Cost(h(x), y) = (h(x))^y \cdot [1 - h(x)]^{1-y}$

$$\begin{cases} y=1, & h(x) \\ y=0, & 1-h(x) \end{cases}$$

为简化问题，对表达式求对数

$$Cost(h(x), y) = y \log h(x) + (1-y) \log(1-h(x))$$

给定一个样本，通过代价函数可以求出，样本所属类别的概率，这个概率越大越好.

对整个样本集：

$$J(\theta) = \sum_{i=1}^{m} [y_i \cdot \log(h(x_i)) + (1-y_i) \log(1-h(x_i))]$$

求解 $J(\theta)$ 的最大值

梯度上升法：

$$x_{i+1} = x_i + \alpha \cdot \frac{\partial f(x_i)}{x_i}$$

$\uparrow$

步长，控制更新的幅度

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

$$\frac{\partial J(\theta)}{\partial g(\theta^T x)} = y \cdot \frac{1}{g(\theta^T x)} + (y-1) \cdot \frac{1}{1-g(\theta^T x)}$$

$$g'(z) = \left(\frac{1}{1+e^{-z}}\right)' = \frac{e^{-z}}{(1+e^{-z})^2} = \frac{1}{1+e^{-z}} \cdot \left(1 - \frac{1}{1+e^{-z}}\right) = g(z)[1-g(z)]$$

$$\frac{\partial g(\theta^T x)}{\partial \theta^T x} = g(\theta^T x) \cdot (1 - g(\theta^T x))$$

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

表示一个样本的所有特征

$\therefore$ 原式 $= [y - h(x)] \cdot x_j$

梯度上升迭代公式：

$$\theta_j = \theta_j + a \sum_{i=1}^{m} (y_i - h(x_i)) \cdot x_{ji}$$