Logistic Regression

2019年6月28日 16:40

非线性回归

模型

测试数据为 $X(x_0, x_1, \cdots x_n)$

未知参数为 $\theta(\theta_0, \theta_1, \cdots \theta_n)$ （要学习的参数）.

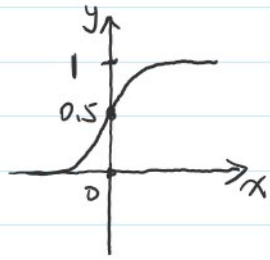
$$z = \theta_0 x_0 + \theta_1 x_1 + \cdots + \theta_n x_n$$

$\downarrow$ 向量表示.

$$z = \theta^T \cdot X.$$

引入 sigmoid function 使数据平滑化在 0,1 之间.

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



预测函数 $h_\theta(x) = g(\theta^T x) = \frac{1}{1+e^{-\theta^T x}}$

用概率来描述预测函数

正例 $(y=1)$: $h_\theta(x) = P(y=1|x;\theta)$

负例 $(y=0)$: $1-h_\theta(x) = P(y=0|x;\theta)$.

$y=1$ 或 $y=0$ 在自变量 x 和已定参数 θ 下的条件概率.

Cost 函数

Cost 函数

对于线性回归

$$Cost(h_\theta x, y) = \sum_{i=1}^{n} (\underbrace{h_\theta(x_i)}_{预测值} - \underbrace{y_i}_{真实值})^2$$

其中 $h_\theta(x_i) = \theta_0 + \theta_i x_i$

找到 θ_0, θ_i 使得 cost 函数值最小

对于非线性回归

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

此处取了负对数，之前需使 $h_\theta(x)$ 最大化，则这时让 $-\log(h_\theta(x))$ 最小化即可.

将 $y=1$ 与 $y=0$ 情况合并.

则目标函数

$$J(\theta) = \frac{1}{n} \sum_{i=1}^{n} cost(h_\theta(x_i), y_i)$$

$$= -\frac{1}{n}\left[\sum_{i=1}^{n}(y_i \cdot \log(h_\theta(x_i)) + (1-y_i)\cdot \log(1-h_\theta(x_i))\right].$$

最后通过学习训练得出的 θ 使 $J(\theta)$ 最小即可.

如何求解 θ：使用 gradient decent (梯度下降算法).

不断更新 θ，直到找到最优点.

更新法则

$$\theta_j = \theta_j - (l) \frac{\partial J(\theta)}{\partial \theta_j} \quad (j=0,1,\cdots n).$$

$$\theta_j = \theta_j - (\alpha) \frac{\partial}{\partial \theta_j} \cdots \quad (j \cdots)$$

$$\Downarrow$$

$$\theta_{j_{new}} = \theta_{j_{old}} - \underbrace{(l)}_{\text{learning rate}} \sum_{i=1}^{n} \left(h_\theta(x_i) - y_i \right) x_{j(i)} \quad (j = 0, 1, \cdots n).$$

对所有的θ同时更新，重复更新直到收敛．