多元线性回归

1 最小二乘法

多元线性回归

$$y_i=\beta_0+\beta_1 x_{i1}+\cdots+\beta_p x_{ip}+\varepsilon_i. \quad i\in[1,n]$$

$$\min_{\beta_0\cdots\beta_p}\sum_{i=1}^{n}(y_i-\hat\beta_0-\hat\beta_1 x_{i1}-\cdots-\hat\beta_p x_{ip})^2$$

最小二乘法：使 y_i 与 $\hat y_i$ 之间离差平方和最小来估计 β_j 的方法.

求解

$$\begin{cases}\dfrac{\partial Q}{\partial\beta_0}=-2\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_{i1}-\cdots-\beta_p x_{ip})=0\\[2ex] \dfrac{\partial Q}{\partial\beta_j}=-2\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_{i1}-\cdots-\beta_p x_{ip})x_{ij}=0\end{cases}$$

$$\therefore\begin{cases}n\beta_0+\sum x_{i1}\beta_1+\cdots+\sum x_{ip}\beta_p=\sum y_i\\ \sum x_{i1}\cdot\beta_0+\sum x_{i1}^2\cdot\beta_1+\cdots+\sum x_{i1}x_{ip}\beta_p=\sum x_{i1}y_i\\ \cdots\\ \sum x_{ip}\cdot\beta_0+\sum x_{ip}^2\cdot\beta_1+\cdots+\sum x_{ip}^2\beta_p=\sum x_{ip}y_i\end{cases}$$

$$X'X\beta=X'Y$$

$$\beta=(X'X)^{-1}X'Y$$

$y_1=\beta_0+\beta_1 x_{1}$

p个变量 x_p

n组数据

2 梯度下降法

多元线性回归.

梯度下降

$$J(\theta)=\frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)})-y^{(i)}\right)^2$$

m组数据

$$h_\theta(x)=\theta^T X=\theta_0+\theta_1x_1+\theta_2x_2+\cdots+\theta_n x_n$$

$i\in[0,n]$

求解:

$$\theta_j:=\theta_j-\alpha\frac{\partial}{\partial\theta_j}J(\theta)$$

j代表第j个变量

$$=\theta_j-\alpha\frac{\partial}{\partial\theta_j}\frac{1}{2m}\sum^{m}\left(h_\theta(x^{(i)})-y^{(i)}\right)^2$$

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

3 牛顿法

多元线性回归

牛顿法

$$\min.\ J(\theta)=\frac{1}{2}\sum_{i=1}^{m}\left(\theta^T x^{(i)}-y^{(i)}\right)^2$$

公式:

$$\theta := \theta - H^{-1}\nabla_\theta J(\theta)$$

$\nabla_\theta J(\theta)$：$J(\theta)$关于θ_j的偏导数向量.

$H_{ij}=\dfrac{\partial^2 J(\theta)}{\partial\theta_i\partial\theta_j}$　　$H: n\times n$.

①
$$\frac{\partial J(\theta)}{\partial\theta_j}=\frac{1}{2}\sum_{i=1}^{m}2\cdot\left(\theta^T x^{(i)}-y^{(i)}\right)\frac{\partial}{\partial\theta_j}\left(\theta^T x^{(i)}-y^{(i)}\right)$$
$\rightarrow$ $\nabla_\theta J(\theta)$的第j个元素

$$=\sum_{i=1}^{m}\left(\theta^T x^{(i)}-y^{(i)}\right)\cdot\frac{\partial}{\partial\theta_j}\left(\theta^T x^{(i)}-y^{(i)}\right)$$
$$=\sum_{i=1}^{m}\left(\theta^T x^{(i)}-y^{(i)}\right)\cdot x_j^{(i)}$$

$$\therefore \nabla_\theta J(\theta)=\begin{bmatrix}\sum_{i=1}^{m}\left(\theta^T x^{(i)}-y^{(i)}\right)\cdot x_1^{(i)}\\ \sum_{i=1}^{m}\left(\theta^T x^{(i)}-y^{(i)}\right)\cdot x_2^{(i)}\\ \cdots\\ \sum_{i=1}^{m}\left(\theta^T x^{(i)}-y^{(i)}\right)\cdot x_n^{(i)}\end{bmatrix}=X^T X\theta-X^T\vec{y}$$

$$X=\begin{bmatrix}x_1^{(1)} & x_2^{(1)} & \cdots & x_n^{(1)}\\ x_1^{(2)} & x_2^{(2)} & \cdots & x_n^{(2)}\\ & \vdots & & \\ x_1^{(m)} & x_2^{(m)} & \cdots & x_n^{(m)}\end{bmatrix}\quad \vec{y}=\begin{bmatrix}y^{(1)}\\ y^{(2)}\\ \vdots\\ y^{(m)}\end{bmatrix}$$

$$\begin{bmatrix}\sum_{i=1}^{m}x_1^{(i)}x_1^{(i)} & \sum_{i=1}^{m}x_1^{(i)}x_2^{(i)} & \cdots & \sum_{i=1}^{m}x_1^{(i)}x_n^{(i)}\\ \sum_{i=1}^{m}x_2^{(i)}x_1^{(i)} & & \cdots & \\ & & & \sum_{i=1}^{m}x_n^{(i)}x_n^{(i)}\end{bmatrix}$$

②
$$\frac{\partial^2}{\partial\theta_j\partial\theta_k}J(\theta)=\frac{\partial}{\partial\theta_k}\sum_{i=1}^{m}\left(\theta^T x^{(i)}-y^{(i)}\right)\cdot x_j^{(i)}=\sum_{i=1}^{m}x_k^{(i)}x_j^{(i)}\qquad \therefore H=X^T X$$

$$\therefore \theta_1=\theta_0-H^{-1}\nabla_\theta J(\theta)$$
$$=\theta_0-(X^T X)^{-1}(X^T X\theta_0-X^T\vec{y})=\theta_0-\theta_0+(X^T X)^{-1}X^T\vec{y}=(X^T X)^{-1}X^T\vec{y}$$