批量梯度下降

2020年4月27日 星期一 下午8:15

1. 梯度下降法

有：$X=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ \vdots & & & \\ x_{N1} & \cdots & \cdots & x_{Nn} \end{bmatrix}$；$\theta=\begin{bmatrix} \theta_0 \\ \vdots \\ \theta_{n-1} \end{bmatrix}$；$y=\begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}$；其中 $\underbrace{\begin{pmatrix} x_{11} \\ \vdots \\ x_{N1} \end{pmatrix}=\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}}$

$y=\theta_0+x\theta_1$

得：$h(x)=\sum_{i=1}^{N} x_{i1}\theta_0+x_{i2}\theta_1\cdots x_{in}\theta_{n-1}$；

$$\text{error}=X\theta-y=\begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & & \\ x_{N1} & \cdots & x_{Nn} \end{pmatrix}\begin{pmatrix} \theta_0 \\ \vdots \\ \theta_{n-1} \end{pmatrix}-\begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}=\begin{pmatrix} x_{11}\theta_0+x_{12}\theta_1\cdots x_{1n}\theta_{n-1}-y_1 \\ \vdots \\ x_{N1}\theta_0+x_{N2}\theta_1\cdots x_{Nn}\theta_{n-1}-y_N \end{pmatrix}$$

关于θ的函数，寻找使得error最小的θ。

$$J=\frac{1}{2N}\sum_{i=1}^{N}\left(x_{i1}\theta_0+x_{i2}\theta_1\cdots x_{in}\theta_{n-1}-y_i\right)^2$$

取平均值，2是为了好看 ==

为了计算误差时得到的正负误差不会相互抵消

$$\frac{\partial J}{\partial \theta_j}=\frac{1}{N}\sum_{i=1}^{N}\left(x_{i1}\theta_0+x_{i2}\theta_1\cdots x_{in}\theta_{n-1}-y_i\right)\cdot x_{i\,j+1}$$

$$☆\ \theta_j^{t+1}=\theta_j^{t}-\alpha\frac{\partial J}{\partial \theta_j^{t}}=\theta_j^{t}-\alpha\cdot\frac{1}{N}\sum_{i=1}^{N}\left(h(x_i)-y_i\right)\cdot x_{i\,j+1}$$

梯度下降法的迭代方法。

2. 正规方程

$$\text{error}=X\theta-y$$

$$J=\frac{1}{2N}\|X\theta-y\|^2=\underbrace{J(\theta)}_{min}$$

$$=\frac{1}{2N}(X\theta-y)^T(X\theta-y)$$

$$=\frac{1}{2N}(\theta^TX^T-y^T)(X\theta-y)$$

$$=\frac{1}{2N}(\theta^TX^TX\theta-\theta^TX^Ty-y^TX\theta+y^Ty)$$

$$\theta^TX^Ty=(X\theta)^Ty=y^T(X\theta)$$

$$=\frac{1}{2N}(\theta^TX^TX\theta-2y^TX\theta+y^Ty)$$

假设 $y^Tx=(a_1\cdots a_n)$

$$y^Tx\theta=(a_1\cdots a_n)\begin{pmatrix} \theta_0 \\ \vdots \\ \theta_{n-1} \end{pmatrix}=a_1\theta_0+\cdots a_n\theta_{n-1}$$

$$\frac{\partial}{\partial\theta}(y^Tx\theta)=\begin{pmatrix} \frac{\partial(a_1\theta_0+\cdots a_n\theta_{n-1})}{\partial\theta_0} \\ \vdots \\ \frac{\partial(a_1\theta_0+\cdots a_n\theta_{n-1})}{\partial\theta} \end{pmatrix}\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}=(y^Tx)^T=x^Ty$$

$$\boxed{\begin{aligned} f(\theta)&=\theta^TA\theta \\ f'(\theta)&=(A+A^T)\theta \end{aligned}\quad \text{矩阵求导公式}}$$

$$J=\frac{1}{2N}(\theta^Tx^Tx\theta-2x^Ty+y^Ty)$$

$$=\frac{1}{2N}\left((\underbrace{x^Tx+xx^T}_{\text{对称矩阵}})\theta-2x^Ty\right)$$

$$=\frac{1}{2N}(2x^Tx\theta-2x^Ty)$$

$$\frac{\partial J}{\partial\theta}=\frac{\partial}{\partial\theta}\frac{1}{N}(x^Tx\theta-x^Ty)=0$$

$$x^Tx\theta=x^Ty$$

$$\theta=(x^Tx)^{-1}x^Ty$$