先考虑一维:

目标: $J(\theta)=\sum_{i=1}^{m}(h_\theta(x^{(i)})-y^{(i)})^2$; 其中 θ 为参数, 目标为求出 θ.

$h_\theta(x^{(i)})$ 为预测值 $(y^{(i)})$ 满足 $\sim N(y^{(i)},\sigma^2)$

满足正态分布可得出:

$$L(\theta|y^{i})=P(h_\theta(x^{(i)}))=\frac{1}{\sigma\sqrt{2\pi}}\cdot e^{[-\frac{1}{2\sigma^2}(h_\theta(x^{(i)})-y^{(i)})^2]}$$

联合似然函数: $\prod_{i=1}^{m}\frac{1}{\sigma\sqrt{2\pi}}\exp[-\frac{1}{2\sigma^2}(h_\theta(x^{(i)})-y^{(i)})^2]=\frac{1}{\sigma\sqrt{2\pi}}\exp(\sum_{i=1}^{m}-\frac{1}{2\sigma^2}\dots)$

可知: 最大似然: 当: $\sum_{i=1}(h_\theta(x^{(i)})-y^{(i)})^2$ 最小时取到

(此时 x 为一维) 则 θ 只有一个

梯度下降

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

$$\begin{cases}\frac{\partial J}{\partial\theta_0}=\frac{1}{m}\sum_{i=1}^{m}(\theta_0+\theta_1x_i-y_i)\\ \frac{\partial J}{\partial\theta_1}=\frac{1}{m}\sum_{i=1}^{m}(\theta_0+\theta_1x_i-y_i)x_i\end{cases}$$

$$\theta_i=\theta_i-\eta\frac{\partial J}{\partial\theta_i}$$

多元:

$$\partial J(\theta)=\begin{bmatrix}\frac{\partial J}{\partial\theta_0}\\ \frac{\partial J}{\partial\theta_1}\\ \frac{\partial J}{\partial\theta_2}\end{bmatrix}=\frac{1}{M}\begin{bmatrix}\frac{\partial J}{\partial\theta_0}=\frac{2}{M}\sum_{i=1}^{M}(\hat{\theta}x-\hat{y})x_0\\ \vdots\\ \frac{\partial J}{\partial\theta_m}=\frac{2}{m}\sum_{i=1}^{m}(\theta x-\hat{y})x_m\end{bmatrix}$$