下面是拟合的函数

$$h_\theta(x)=\theta_0+\theta_1x_1+\theta_2x_2$$

$$h_\theta(x)=\sum_{i=0}^{n}\theta_ix_i=\theta^Tx$$

其中ϵ是残差

$$y^{(i)}=\theta^Tx^{(i)}+\varepsilon^{(i)}$$

假设残差服从正态分布

误差$\varepsilon^{(i)}(1\le i\le m)$是独立同分布的，服从均值为0，方差为某定值$\sigma^2$的高斯分布。

- 原因：中心极限定理

实际问题中，很多随机现象可以看做众多因素的独立影响的综合反应，往往近似服从正态分布。

似然函数 $y^{(i)}=\theta^Tx^{(i)}+\varepsilon^{(i)}$

可以得到残差的密度函数即为均值为 0（均值的其他部分可以放到截距项θ_1里面），方差为σ^2的正态分布

$$p(\epsilon^{(i)})=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(\epsilon^{(i)})^2}{2\sigma^2}\right)$$

其中残差等于真实值减去预测值，代换得到$y^{(i)}-\theta^Tx^{(i)}$的密度函数为

$$p(y^{(i)}|x^{(i)};\theta)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y^{(i)}-\theta^Tx^{(i)})^2}{2\sigma^2}\right)$$

那么似然函数即为所有样本出现的概率乘积，也就是同时看到 y1 y2 y3…… ym 的概率

$$\begin{aligned}L(\theta) &= \prod_{i=1}^{m}p(y^{(i)}\mid x^{(i)};\theta)\\ &= \prod_{i=1}^{m}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y^{(i)}-\theta^Tx^{(i)})^2}{2\sigma^2}\right)\end{aligned}$$

我们将乘积改成加和，于是取对数

$$\begin{aligned}\ell(\theta) &= \log L(\theta)\\ &= \log\prod_{i=1}^{m}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y^{(i)}-\theta^Tx^{(i)})^2}{2\sigma^2}\right)\end{aligned}$$

其中 log 里面的乘积可以展开为加和的形式

$$= \sum_{i=1}^{m} \log \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2}\right)$$

再把其中的包含π 和 σ的常数项展开

$$= m \log \frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{\sigma^2} \cdot \frac{1}{2} \sum_{i=1}^{m} (y^{(i)} - \theta^T x^{(i)})^2$$

然后去掉其中的常数项，是不是看到了熟悉的公式

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

求目标函数最优解

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

$$J(\theta) = (X\theta - y)^T \cdot (X\theta - y)$$

将转置展开

$$= (\theta^T X^T - y^T)(X\theta - y)$$

多项式展开

$$= \theta^T \boxed{X^T X} \theta - \theta^T X^T y - y^T X \theta + y^T y$$

对θ求偏导

$$\frac{\partial J(\theta)}{\partial \theta} = X^T X \theta - X^T y$$

另偏导为 0，解出θ

$$\theta = (X^T X)^{-1} \cdot X^T \cdot y$$

若X^TX不可逆或防止过拟合，增加λ扰动

$$\theta = \left(X^T X + \lambda I\right)^{-1} X^T y$$

"简便"方法记忆结论

$$X\theta = y \Rightarrow X^T X\theta = X^T y$$
$$\Rightarrow \theta = \left(X^T X\right)^{-1} X^T y$$

其中 I 是单位阵，出现过拟合的原因是求得的参数θ比较大，函数出现振荡，所以要解决过拟合就得使θ变小

将目标函数增加平方和损失：

$$J\left(\vec{\theta}\right) = \frac{1}{2}\sum_{i=1}^{m}\left(h_{\vec{\theta}}\left(x^{(i)}\right) - y^{(i)}\right)^2 + \lambda\sum_{j=1}^{n}\theta_j^2$$