Machine Learning

线性回归：在一定的区间内，找到它的值.

$Y = X_1\theta_1 + X_2\theta_2$　　求解最合适的 θ_1 和 θ_2.

没办法解　　所有人：

尽可能　　合当前的数据.

$$h_\theta = \theta_0 + \theta_1 X_1 + \theta_2 X_2$$

置项　　权重项

$$h_\theta(X) = \sum_{i=0}^{n} \theta_i X_i = \theta^T X$$

自己动手加入一行 X_0.
X_0 内所有的数据都是 1.

误差

$$y^{(i)} = \theta^T x^{(i)} + \epsilon^{(i)}$$

真实值　预测值　误差.

误差是独立的，因为样本是独立的
服从 $\mu = 0$　$\sigma^2 = \theta^2$ 的高斯分布
同样也是同分布的.

Gaussian Distribution:

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

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

用累乘是因为联合.
概率密度 = 边缘密度的乘积.

似然函数　$L(\theta) = \prod_{i=1}^{m} p(y^{(i)} \mid x^{(i)};\theta)$

找到一个参数跟数据组合恰好是真实值.

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

乘法转为加法，寻找极值点而不是极值.

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

B越小越好.

$$J(\theta) = \frac{1}{2} \sum_{i=1}^{m} (y^{(i)} - \theta^T x^{(i)})^2. \Rightarrow$$ 最小二乘法

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

求什么样的θ可以让总体越低越好.

求导: $\nabla_\theta J(\theta) = \nabla_\theta \left(\frac{1}{2}(X\theta - y)^T (X\theta - y)\right)$

$$= \nabla_\theta \left(\frac{1}{2}(\theta^T X^T X\theta - \theta^T X^T y - y^T X\theta + y^T y)\right)$$

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

乘逆转质

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

没有学习的过程?