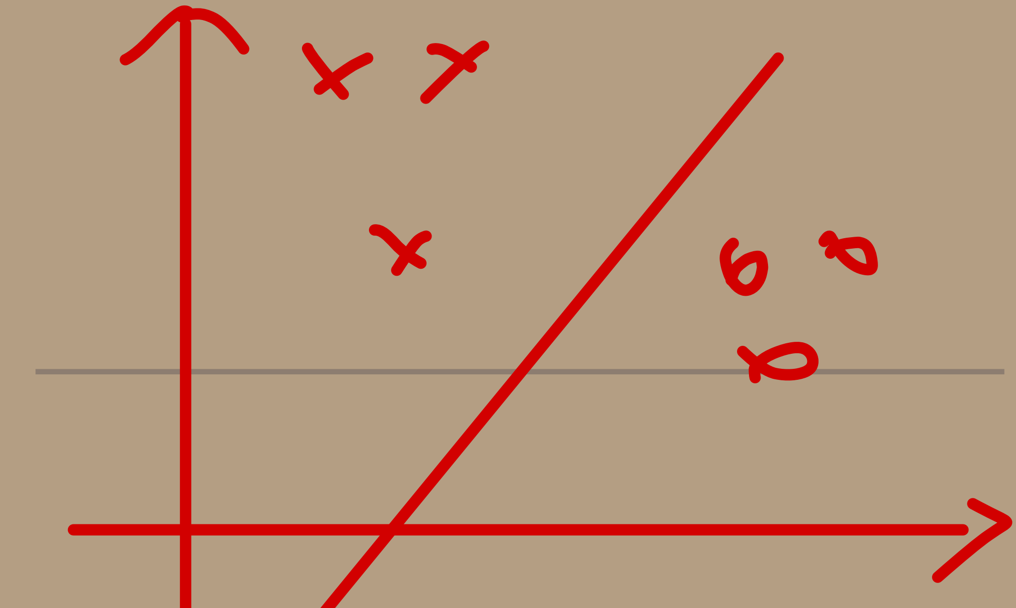

$f(x) = w^T x$

最小二乘估计

noise $\sim N(0, \sigma^2)$

MLE

Linear Regression:

设：样本集：$D=\{(x_1,y_1),(x_2,y_2),\cdots(x_N,y_N)\}$，$x_i\in\mathbb{R}^p$，$y_i\in\mathbb{R}$。

$$X=\begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{bmatrix}_{N\times p} \qquad Y=\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}_{N\times 1}$$

①最小二乘估计：hyposisss function: $f(x)=w^Tx$

Loss function: $L(w)=\sum_{i=1}^{n}\|w^Tx_i-y_i\|^2$，$w=\begin{bmatrix} w_1 \\ \vdots \\ w_p \end{bmatrix}_{p\times 1}$

goal: $\underset{w}{\operatorname{argmin}}\, L(w)$

$$L(w)=\underbrace{\sum_{i=1}^{n}(w^Tx_i-y_i)^2}_{\underbrace{(w^Tx_1-y_1,\ w^Tx_2-y_2\cdots w^Tx_N-y_N)}_{\underbrace{(w^Tx_1,w^Tx_2\cdots w^Tx_N)}_{w^T(x_1,x_2\cdots x_N)}-\underbrace{(y_1,y_2\cdots y_N)}_{Y^T}}}\begin{pmatrix} w^Tx_1-y_1 \\ \vdots \\ w^Tx_N-y_N \end{pmatrix}$$

$\underbrace{1\times p\quad p\times N\quad N\times 1}$

$$\therefore L(w)=(w^TX^T-Y^T)(w^TX^T-Y^T)^T=(w^TX^T-Y^T)(Xw-Y)$$

$$=w^TX^TXw\underline{-w^TX^TY-Y^TXw}+Y^TY$$

w^TX^TY为实数，Y^TXw为其转置。

$$=w^TX^TXw-2w^TX^TY+Y^TY$$

$$\frac{\partial L(w)}{\partial w}=2X^TXw-2X^TY=0$$

$$\Rightarrow w=(X^TX)^{-1}X^TY$$

②极大似然估计：数据本身具有随机性，存在噪声。

最小二乘估计：

$$L(w) = \sum_{i=1}^{N} \|w^T x_i - y_i\|^2$$

$$\hat{w} = \arg\min_{w} L(w)$$

$$\hat{w} = (X^T X)^{-1} X^T Y$$

噪声 $\varepsilon \sim N(0, \sigma^2)$

$$y = f(w) + \varepsilon$$

$$f(w) = w^T x$$

$$y = w^T x + \varepsilon$$

$$y|x;w \sim N(w^T x, \sigma^2)$$

$$P(y|x;w) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-w^T x)^2}{2\sigma^2}\right)$$

MLE：(极大似然估计)

$$\mathcal{L}(w) = \log P(Y|X;w) = \log \prod_{i=1}^{N} P(y_i|x_i;w)$$

$$= \sum_{i=1}^{N} \log P(y_i|x_i;w)$$

$$= \sum_{i=1}^{N} \log \frac{1}{\sqrt{2\pi}\sigma} + \log \exp\left\{-\frac{(y_i - w^T x_i)^2}{2\sigma^2}\right\}$$

$$= \sum_{i=1}^{N} \log \frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{2\sigma^2}(y_i - w^T x_i)^2$$

$$\hat{w} = \arg\max_{w} \mathcal{L}(w)$$

$$= \arg\max_{w} -\sum_{i=1}^{N} \frac{1}{2\sigma^2}(y_i - w^T x_i)^2$$

$$= \arg\min_{w} (y_i - w^T x_i)^2$$

最小二乘估计隐含了噪声服从高斯分布的假设。

$\therefore$ LSE $\Longleftrightarrow$ MLE (Noise is gaussian distribution)。