$$p(x,y) = p(x|y)p(y) = p(y|x)p(x)$$

1 heads, 0 tails

$\theta_{ML} = 1$

$$\text{posterior} \quad p(\theta|D) = \frac{\text{prior}\ p(\theta)\ \text{likelihood}\ p(D|\theta)}{p(D)}$$

θ: parameter, D: data, $p(D)$: evidence

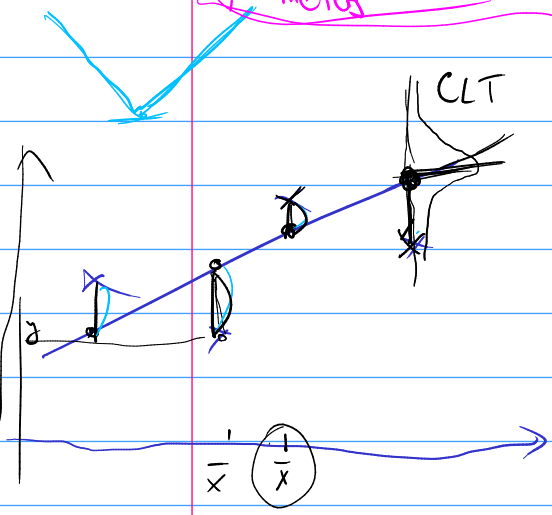
$$p(\theta|D) \propto p(D|\theta)p(\theta)$$

heads, tails

$$p(D|\theta) = \theta^k(1-\theta)^l \to \max$$

p(heads)

$$\theta_{ML} = \frac{k}{k+l}$$



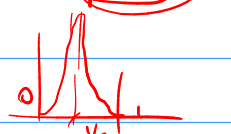
$$D = \{(\bar{x}_n, y_n)\}_{n=1}^N$$

$$y_n \approx \bar{w}^T \bar{x}_n$$

$$L = \sum_n (y_n - \bar{w}^T \bar{x}_n)^2$$

$$p(\theta|D) \propto \propto p(D|\theta) \cdot p(\theta)$$

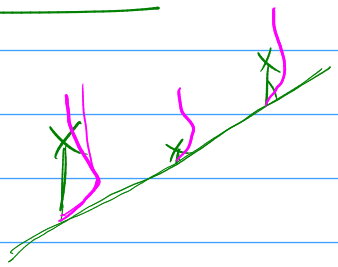
θ



$$p(y|\bar{w},\bar{x}) = N(y|\bar{w}^T\bar{x}, \sigma^2) \leftarrow \text{likelihood}$$

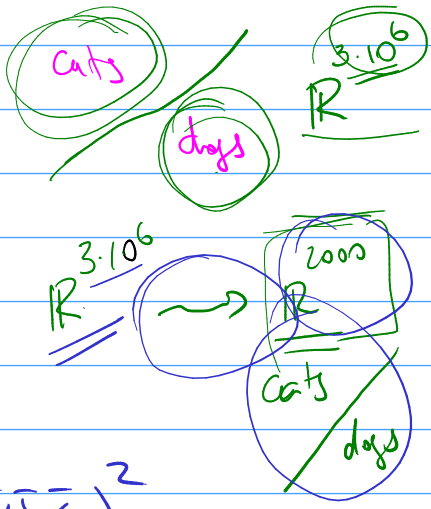
$$p(C_1|\bar{x}) \sim \sigma(\bar{w}^T\bar{x})$$

$$p(D|\bar{w}) = \prod_{n=1}^N p(y_n|\bar{w},\bar{x}_n) =$$



$$= \prod_{n=1}^N N(y_n|\bar{w}^T\bar{x}_n, \sigma^2) =$$

$$= \prod_{n=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2\sigma^2}(y_n - \bar{w}^T\bar{x}_n)^2}$$



$$\ln p(D|\bar{w}) = -\frac{N}{2}\ln(2\pi\sigma^2) - \sum_{n=1}^N \frac{1}{2\sigma^2}(y_n - \bar{w}^T\bar{x}_n)^2$$

$$= \text{const} - \text{const} \cdot \sum_{n=1}^N (y_n - \bar{w}^T\bar{x}_n)^2 + \ldots ?$$

max

min

$$p(\theta|D) \propto p(\theta)p(D|\theta)$$

$$\ln p(\theta|D) = \text{const} + \log p(\theta) + \log p(D|\theta)$$

$-c\sum(\cdot)^2$

$\to$ max

$$p(\bar{w}) = \mathcal{N}(\bar{w} \mid \bar{0}, \sigma_0^2 \cdot I) = \prod_{i=1}^{d} \mathcal{N}(w_i \mid 0, \sigma_0^2) = \prod_{i=1}^{d} \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{1}{2\sigma_0^2} w_i^2}$$

$$\log p(\bar{w}) = -\frac{1}{2}\log(2\pi\sigma_0^2) - \frac{1}{2\sigma_0^2} \sum_{i=1}^{d} w_i^2$$

$$\sum_{i=1}^{d} w_i^2 = \|\bar{w}\|^2$$

$$\ln p(\bar{w} \mid D) = \text{const} - C_1 \sum_n \left(y_n - \bar{w}^T \bar{x}_n\right)^2 - C_2 \|\bar{w}\|^2 \to \max$$

$$L(\bar{w}) = \sum_n \left(y_n - \bar{w}^T \bar{x}_n\right)^2 + \alpha \cdot \|\bar{w}\|^2$$

Backpropagation

regularizer

ridge regression

$\alpha \cdot \sum |w_i|$

L_1 - regularization

lasso regression