

ML

DL

Supervised

Unsupervised

General

CNNs

Tricks

Labeled data

Unlabeled

$$D = \{(x, y)\}$$

$$D = \{x\}$$

input

target

Clustering:

likelihood prior

parameters

Bayes theorem

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$$

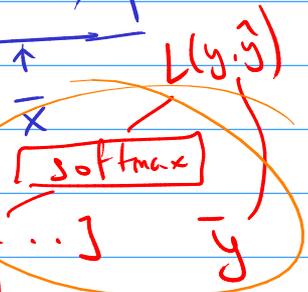
posterior evidence

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

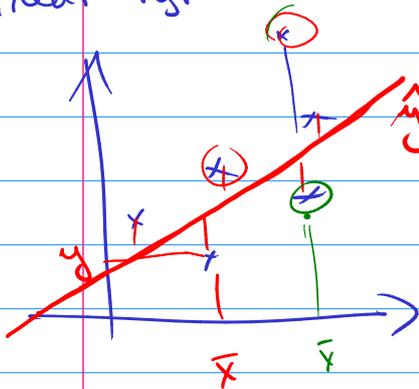
linear regression

$$L(\hat{y}, y) = (y - \hat{y})^2$$

logistic regression



Linear rgr



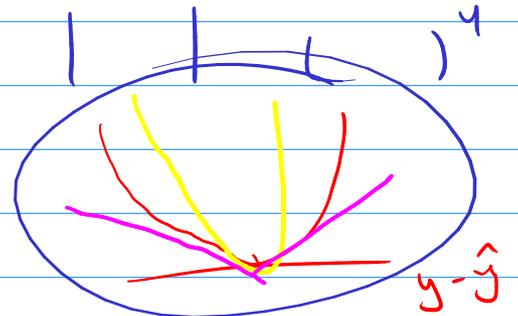
$$\hat{y} = w_1 x_1 + \dots + w_d x_d$$

$$\hat{y} = \bar{w}^T \bar{x}$$

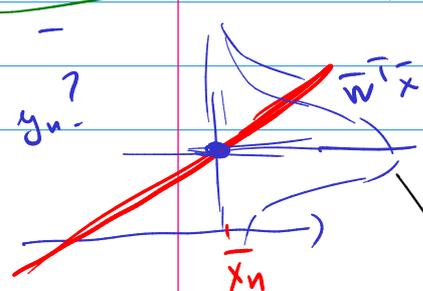
$$L(y, \hat{y}) = (y - \hat{y})^2 = (y - \bar{w}^T \bar{x})^2$$

$$p(D|\bar{w}) = \prod_{d \in D} p(d|\bar{w})$$

$$D = \{(x_n, y_n)\}_{n=1}^N$$

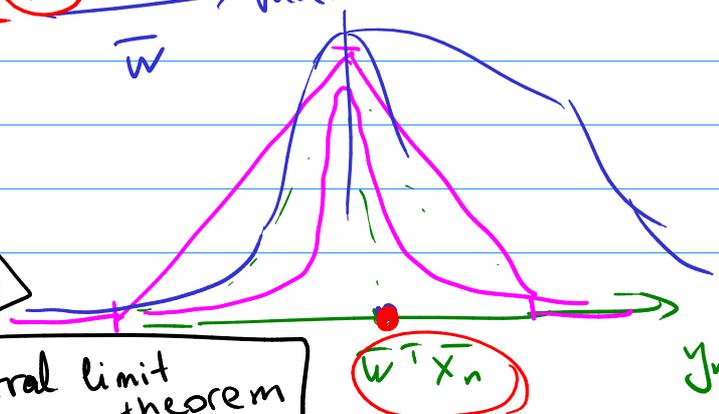


$$= \prod_{n=1}^N p(y_n | \bar{w}, x_n) \rightarrow \max_{\bar{w}}$$



$$y_n = \bar{w}^T \bar{x} + \epsilon$$
$$\epsilon \sim \mathcal{N}(0, \sigma^2)$$

central limit theorem



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

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

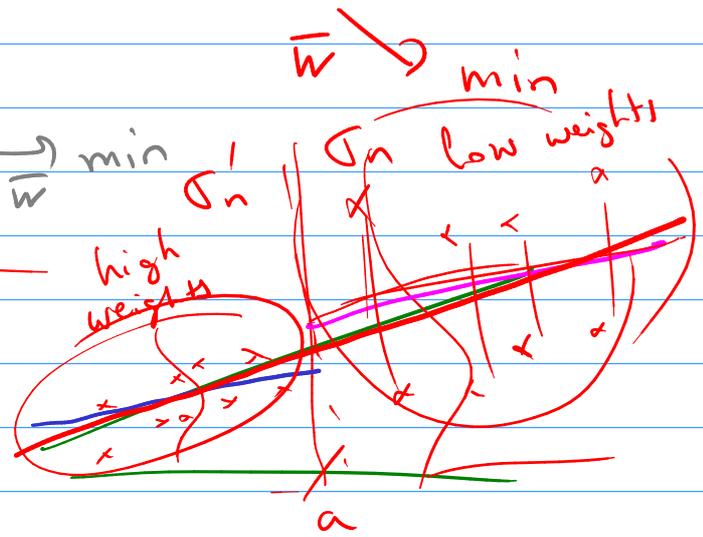
$\bar{w} \rightarrow \max$

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

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

$\bar{w} \rightarrow \max$

$$\sum_{n=1}^N \frac{1}{\sigma_n^2} (y_n - \bar{w}^T \bar{x}_n)^2 \xrightarrow{\bar{w}} \min$$



Bernoulli

$$p(\text{heads}) = \theta$$

$$p(\text{tails}) = 1 - \theta$$

$$p(D | \theta) = \prod_{d \in D} p(d | \theta) =$$

h h t h t t h

$$= \theta^4 (1 - \theta)^3 \xrightarrow{\theta} \max$$

$$\log p(D | \theta) = 4 \log \theta + 3 \log(1 - \theta) = \theta^4 (1 - \theta)^3 \xrightarrow{\theta} \max$$

$$\frac{\partial \log p}{\partial \theta} = \frac{4}{\theta} - \frac{3}{1 - \theta} = 0$$

$$4(1-\theta)_{MC} - 3\theta_{MC} = 0$$

$$\theta_{MC} = \frac{4}{7}$$

$$D = \{h\}$$

$$\theta_{MC} = \frac{1}{2}$$

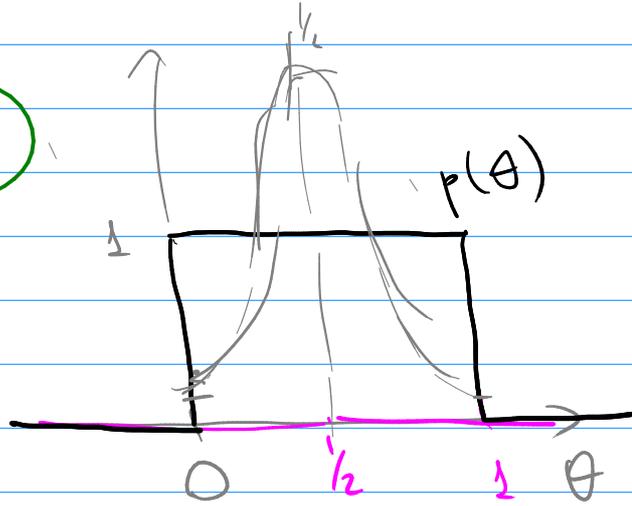
$$p(D|\theta) = \theta$$

$$\theta_{MC} \rightarrow \infty$$

ht

$$\theta_{MC} = \frac{1}{2}$$

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$$



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

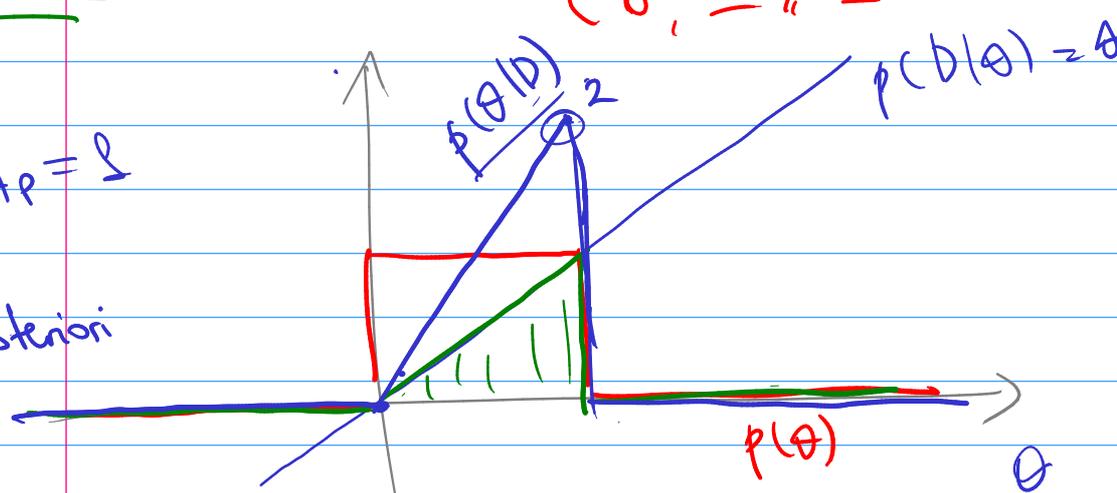
$\theta \in \mathbb{R}$

$$p(\theta) = \begin{cases} 1, & \theta \in [0, 1] \\ 0, & \text{---} \end{cases}$$

$$p(\theta|D) \propto \theta \cdot p(\theta) = \begin{cases} \theta, & \theta \in [0, 1] \\ 0, & \text{---} \end{cases}$$

$$\theta_{MAP} = \frac{1}{2}$$

max a posteriori

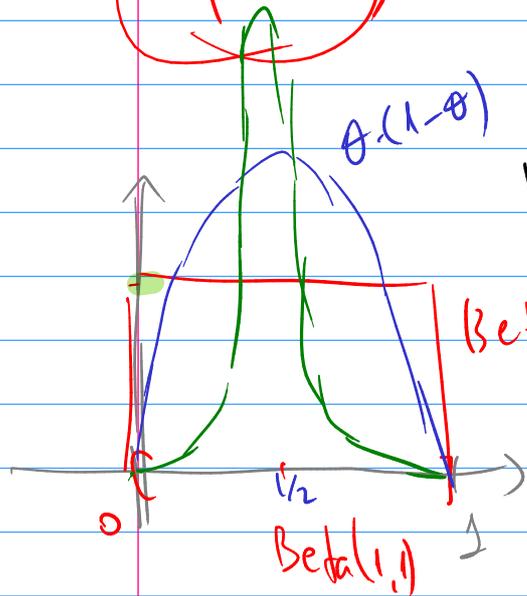


$$p(\theta) = \begin{cases} c \cdot e^{-\frac{1}{2\sigma^2}(\theta - \frac{1}{2})^2}, & \theta \in [0, 1] \\ 0, & \text{---} \end{cases}$$

$$p(\theta|D) \propto \theta^n (1-\theta)^m \cdot e^{-\frac{1}{2\sigma^2}(\theta - \frac{1}{2})^2}$$

$$\underbrace{\theta^n (1-\theta)^m}_{p(D|\theta)} \times \underbrace{\theta^{\alpha-1} (1-\theta)^{\beta-1}}_{p(\theta)}, \theta \in [0,1]$$

$$\underbrace{p(\theta|D)}_{\propto} \theta^{n+\alpha-1} (1-\theta)^{m+\beta-1}$$



$$\text{Beta}(x | \alpha, \beta) = \frac{1}{B(\alpha, \beta)} \cdot x^{\alpha-1} (1-x)^{\beta-1} \quad x \in [0,1]$$

$$0, x \notin [0,1]$$

Beta $\theta(1-\theta)$

Conjugate prior

$$p(\theta) = \text{Beta}(\theta | \alpha, \beta) \times p(D|\theta) = \theta^n (1-\theta)^m \propto$$

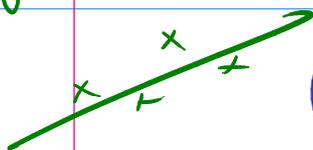
$$\propto p(\theta|D) = \text{Beta}(\theta | \alpha+n, \beta+m)$$

$$p(\bar{w} | D) \propto p(D|\bar{w}) \cdot p(\bar{w})$$

$$\bar{f}(x) \rightarrow \text{[Lin Keys]}$$

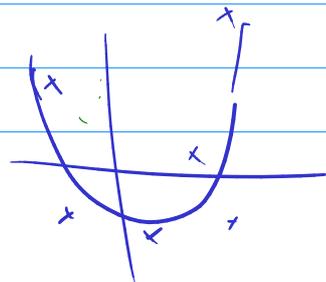
$$\frac{f}{x}$$

$$y \sim \bar{w}^T \bar{x}$$



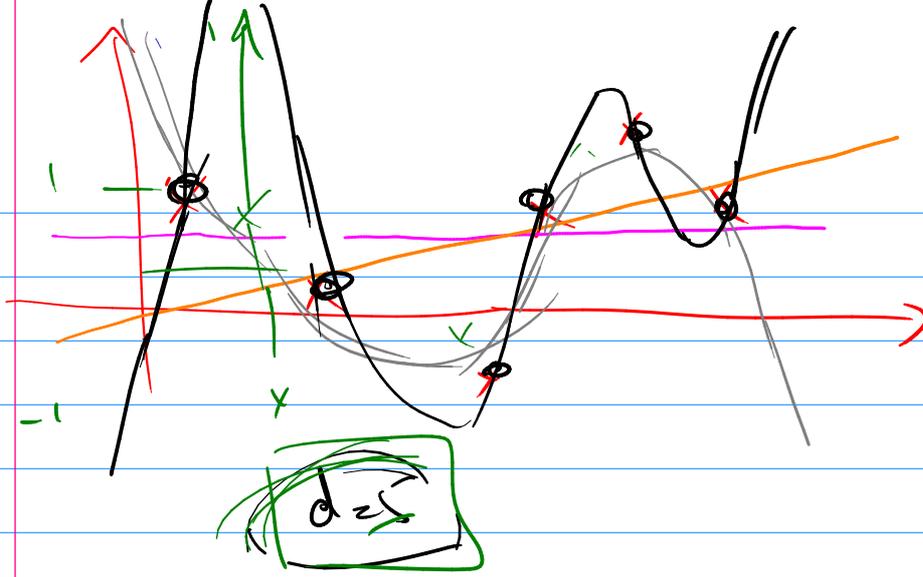
$$\begin{pmatrix} w_0 \\ w_1 \\ w_n \end{pmatrix}^T \begin{pmatrix} 1 \\ x \end{pmatrix}$$

$$y \sim$$



$$y \sim w_0 + w_1 x + w_2 x^2$$

$$= \begin{pmatrix} w_0 \\ w_1 \\ w_2 \end{pmatrix}^T \begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix}$$



- 1) $d=0$
 - 2) $d=1$ $\begin{pmatrix} w_0 \\ w_1 \end{pmatrix}^T \begin{pmatrix} 1 \\ x \end{pmatrix}$
 - 3) $\begin{pmatrix} w_0 \\ w_1 \\ w_2 \end{pmatrix}^T \dots$
- $d=3$

$$p(\bar{w}) = \mathcal{N}(\bar{w} | \bar{0}, \begin{pmatrix} \sigma^2 & 0 \\ 0 & -\sigma^2 \end{pmatrix} = \sigma_0^2 \cdot \mathbf{I}_2) =$$

$$= \frac{1}{(2\pi\sigma_0^2)^{d/2}} \cdot e^{-\frac{1}{2\sigma_0^2} \bar{w}^T \bar{w}}$$

WMAP \rightarrow max

$$\log p(\bar{w} | D) = \log p(D | \bar{w}) + \log p(\bar{w}) + \text{const} =$$

$$= \dots - \frac{1}{2\sigma^2} \sum_{n=1}^N (y_n - \bar{w}^T \bar{x}_n)^2 + \dots - \frac{1}{2\sigma_0^2} \bar{w}^T \bar{w} + \text{const}$$

$$L(\bar{w}) = \sum_{n=1}^N (y_n - \bar{w}^T \bar{x}_n)^2 + \frac{\sigma^2}{\sigma_0^2} \cdot \bar{w}^T \bar{w}$$

L_2 -regularization

L_1 -regularization
 $\lambda \cdot \sum |w_i|$
 Laplace distribution