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

loss function

$$-\left(\log p(D | \bar{\theta}) + \log p(\bar{\theta})\right) \xrightarrow{\bar{\theta}} \min$$

$$\bar{\theta}_{MAP} = \arg\min = \arg\max\, p(\bar{\theta} | D)$$

point estimate

distribution

	#1	#2
1/4	M	M
1/4	M	F
1/4	F	M
~~1/4~~	~~F~~	~~F~~

2/3

$$p(\exists F | \exists M) = \frac{p(MF \text{ or } FM)}{p(MM, MF, FM)} = \frac{2/4}{3/4}$$

$$p(\text{guilty} | \text{evidence})$$

$$\underbrace{p(\text{blood} | \text{innocent})}_{= 0.01} \neq 1 - p(\text{guilty} | \text{blood})$$

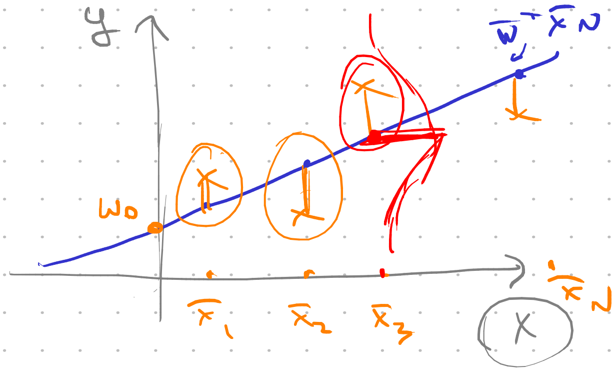
$$p(A|B) \neq p(B|A)$$

$$p(\text{innoc} | \text{blood}) = \frac{p(\text{blood} | \text{inn}) \cdot p(\text{inn})}{p(\text{blood})}$$

$$p(\text{murdered} | \text{violence}) = \frac{1}{2500}$$

$$p\left(\text{murdered by husband} \,\middle|\, \text{murdered, violence}\right) = 7/8$$

② Linear regression



$$y \approx w_0 + w_1 x_1 + \dots + w_d x_d$$

$$y \overset{2}{\approx} \bar w^T \bar x = \begin{pmatrix} w_1 \\ \vdots \\ w_d \end{pmatrix}^T \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix}$$

$x_0 = 1$

$$\bar x \mapsto \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_d \end{pmatrix} \qquad \begin{pmatrix} w_0 \\ \vdots \\ w_d \end{pmatrix}^T$$

$$L = \sum_{n=1}^{N} \left(y_n - \bar w^T \bar x_n\right)^2 \xrightarrow[\bar w]{} \min$$

$$\sum |y_n - \bar w^T \bar x_n| \to \min \qquad \sum (\dots)^4 \to \min \qquad \sum e^{|\dots|} \to \min$$

$$\bar\theta = \bar w \qquad D = \{(\bar x_1, y_1) \dots (\bar x_N, y_N)\}$$

$$p(D|\bar w) \overset{1}{=} \prod_{n=1}^{N} p(y_n | \bar w, \bar x_n) = \prod_{n=1}^{N} \overset{3}{\mathcal{N}}(y_n | \overset{2}{\bar w^T \bar x_n}, \overset{4}{\sigma^2}) =$$

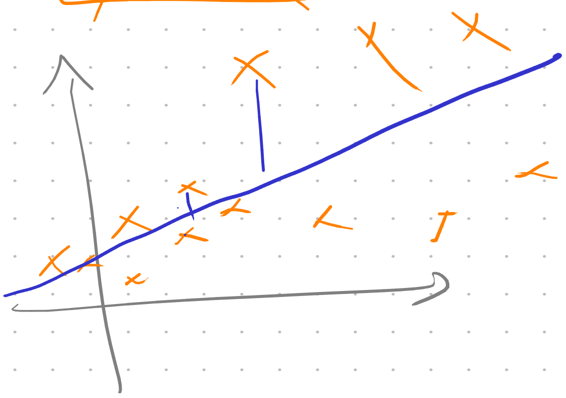
homoscedasticity

$$= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2\sigma^2}\left(y_n - \bar w^T \bar x_n\right)^2} \xrightarrow[\bar w]{} \max$$

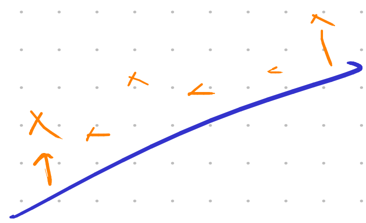
$\bar w_{ML}$

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

$$= \cancel{-\frac{N}{2} \log 2\pi\sigma^2} - \frac{1}{2\sigma^2} \boxed{\sum_{n=1}^{N} \left(y_n - \bar w^T \bar x_n\right)^2} \xrightarrow[\bar w]{} \max$$



$\xrightarrow[\bar w]{} \min$

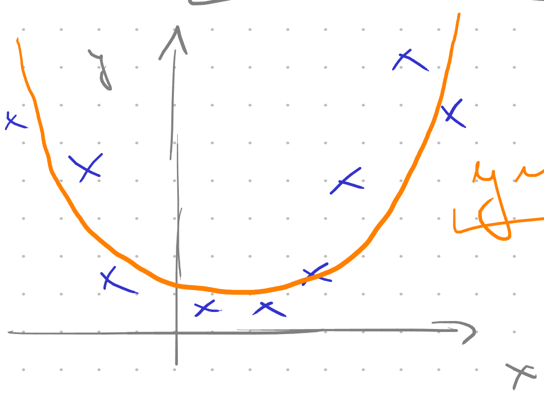


$$\max_{\bar w} \; p(\bar w | D) \propto p(D|\bar w)\, p(\bar w)$$ regularizer

$$\log p(\bar w|D) = \text{Const} + \log p(D|\bar w) + \log p(\bar w) \quad \xrightarrow[\bar w]{} \max$$

$\bar x$ — LR — $\bar w^T \bar x$

$\bar x$ — $\boxed{\text{Feature extraction}}$ — $\bar\varphi(\bar x)$ — $\boxed{\text{LR}}$ — $\bar w^T \varphi(\bar x)$



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

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

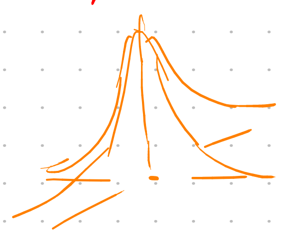
$$L = \sum_n \left(y_n - (w_0 + w_1 x + w_2 x^2)\right)^2$$

D $\quad \bar x \in \mathbb{R}^d \quad \underset{100}{(x_1 \dots x_d)} \sim \; \mapsto 5000$ overfitting

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

$$\bar x = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} \mapsto \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_d \\ x_1^2 \\ x_1 x_2 \\ x_1 x_3 \\ \vdots \\ x_d^2 \end{pmatrix} \qquad \frac{d(d+1)}{2}$$

$p(\bar w)$



$$p(\bar w) = \mathcal{N}(\bar w \,|\, \bar 0, \sigma_0^2 I) \qquad \sum w_i^2$$

$$\log p(\bar w) = \text{Const} - \frac{1}{2\sigma_0^2} \cdot \|\bar w\|^2$$

ridge regression

$$\cancel{\tfrac{1}{2}} \sum_n \left(y_n - \bar w^T \bar x_n\right)^2 + \frac{\sigma^2}{\cancel{2}\sigma_0^2} \sum w_i^2 \longrightarrow \min$$

L_2-regularizer

Sparsity-inducing

$$\underbrace{\sum_n (y_n - \bar{w}^T \bar{x}_n)^2}_{} + \underbrace{\alpha \cdot \sum |w_i|}_{L_1\text{-regularizer}} \longrightarrow \min$$

Lasso regression

$p(\bar{w})$

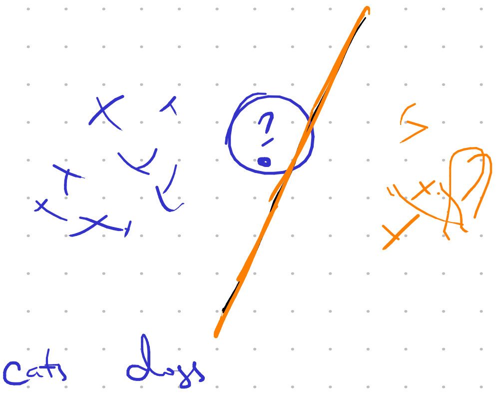
neural network

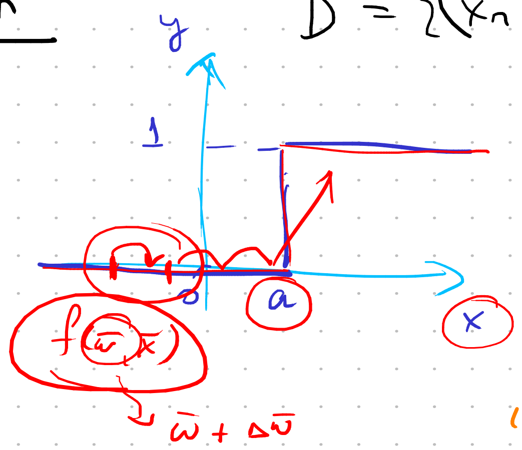
$$\underbrace{\sum_n \big(y_n - f(\bar{w}, \bar{x}_n)\big)^2 + \alpha \cdot \sum w_i^2}_{} \longrightarrow \min$$

③ Logistic regression

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

$\in \{1, 2, \dots, K\}$





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

C_1, C_2

$\in [0,1]$

$p(C_1 | \bar{x})$

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

$$p(C_1|\bar{x}) = \frac{p(\bar{x}|C_1)\,p(C_1)}{p(\bar{x})} = \frac{p(\bar{x}|C_1)\,p(C_1)}{p(\bar{x}|C_1)\,p(C_1) + p(\bar{x}|C_2)\,p(C_2)} =$$

$$\frac{1}{1 + \frac{p(\bar{x}|C_2)\,p(C_2)}{p(\bar{x}|C_1)\,p(C_1)}} = \frac{1}{1 + e^{-\boxed{\log \frac{p(\bar{x}|C_1)\,p(C_1)}{p(\bar{x}|C_2)\,p(C_2)}}}}$$

log-odds

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

odds $= \dfrac{p(C_2|\bar{x})}{p(C_1|\bar{x})}$ $\cancel{\sim \bar{w}^T \bar{x}}$

$\in (0, \infty)$

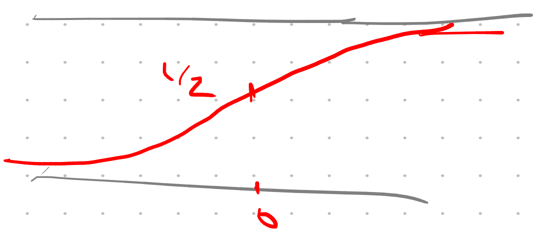
log. reg.:

linear model

$$p(C_1|\bar{x}) = \frac{1}{1+e^{-\bar{w}^T\bar{x}}} = \sigma(\bar{w}^T\bar{x})$$

$$\sigma(a) = \frac{1}{1+e^{-a}}$$

$$p(C_2|\bar{x}) = 1 - \sigma(\bar{w}^T\bar{x})$$



$$p(D|\bar{w}) = \prod_n p(y_n|\bar{w}, \bar{x}_n) =$$

$$= \prod_n \underbrace{\sigma}\ \underbrace{(1-\sigma)}\ \underbrace{\sigma}\ \underbrace{\sigma(1-\sigma)}$$

$C_1, C_2, \dots, C_K$

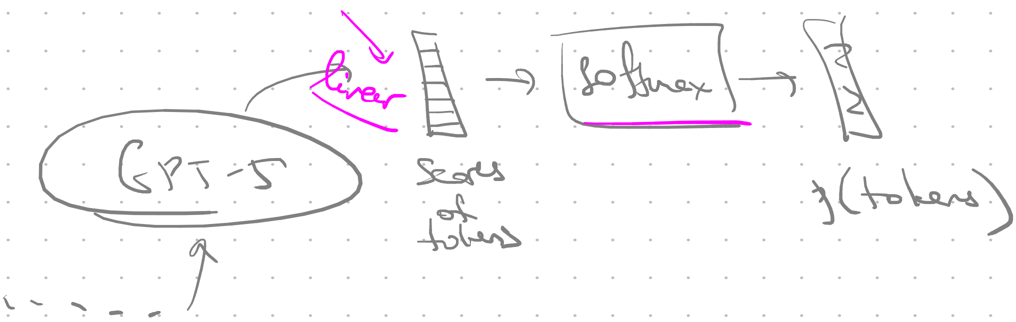
$$p(C_k|\bar{x}) = \frac{p(C_k)p(\bar{x}|C_k)}{\sum_{s=1}^{K} p(C_s)\cdot p(\bar{x}|C_s)} = \frac{e^{p(C_k)p(\bar{x}|C_k)}}{\sum_s e^{- s -\ -}}$$

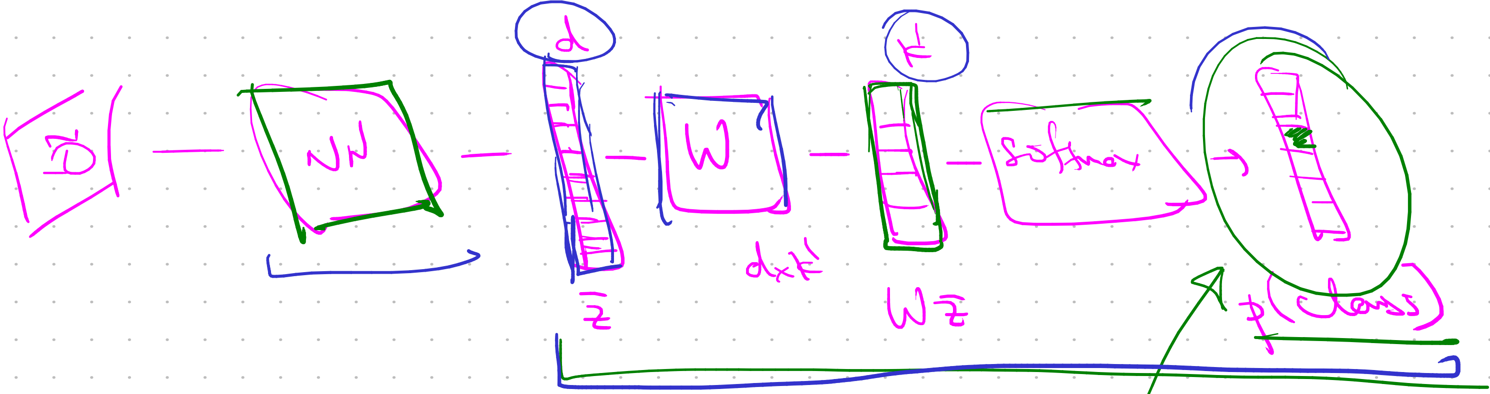
score affinity $\bar{w}_k^T\bar{x} \approx p(C_k)\,p(\bar{x}|C_k)$

$$\frac{e^{-2}}{e^{1}+e^{-1}+e^{-2}} = \frac{e^{0}}{e^{3}+e^{1}+e^{0}}$$

$$\bar{a} \mapsto \left(\dots,\ e^{a_k} \Big/ \sum_{s=1}^{K} e^{a_s},\ \dots \right)$$

$$\text{softmax}(a_1, a_2, \dots, a_K) = \left(\dots,\ \frac{e^{a_k}}{\sum_s e^{a_s}},\ \dots \right)$$

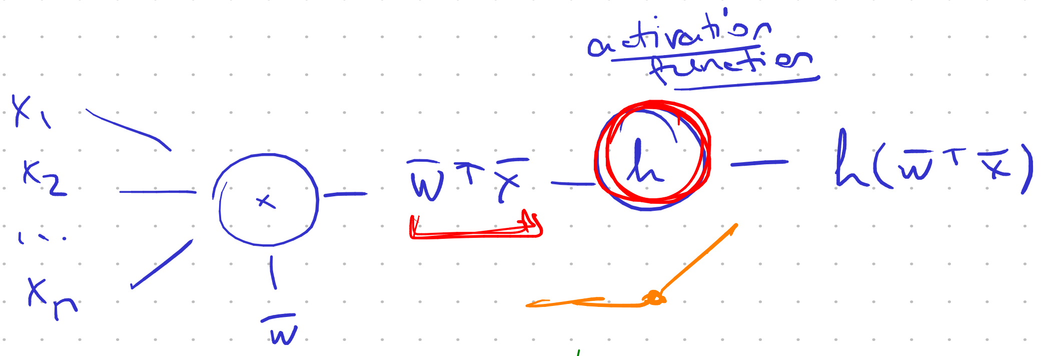


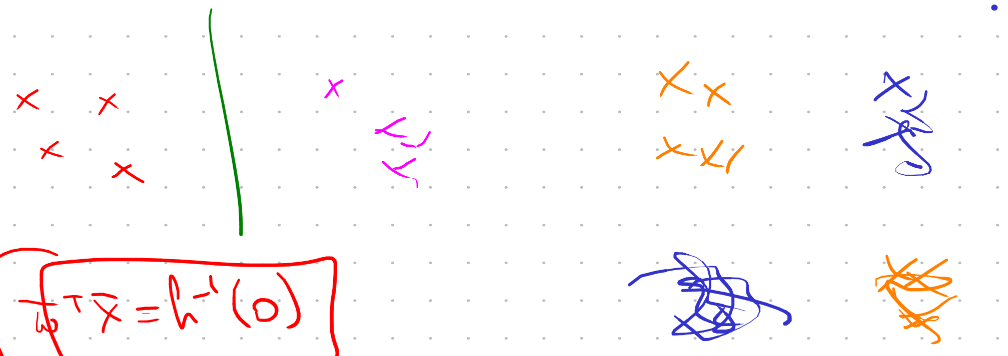


logistic regression

cat $\longrightarrow$ $[0, 0, 1, 0, 0]$

loss $= -\log p(\text{cat})$





$h(\bar{w}^T\bar{x}) > 0 \Rightarrow C_1$

$< 0 \Rightarrow C_2$

$\boxed{h(\bar{w}^T\bar{x}) = 0}$ $\quad$ $\boxed{\bar{w}^T\bar{x} = h^{-1}(0)}$

$\boxed{\text{AIC}, \ \text{BIC}}$ -- model selection

quadr. regr. $\quad$ lin. reg.

$D \rightarrow M_1, M_2, M_3, \ldots, M_M$

$\bar{\theta}_{MAP}$

$$p(M_k \mid D) \propto \cancel{p(M_k)}\, p(D \mid M_k)$$

$$p(\bar{\theta} \mid D, M) = \frac{p(\bar{\theta} \mid M)\, p(D \mid \bar{\theta}, M)}{p(D \mid M)}$$