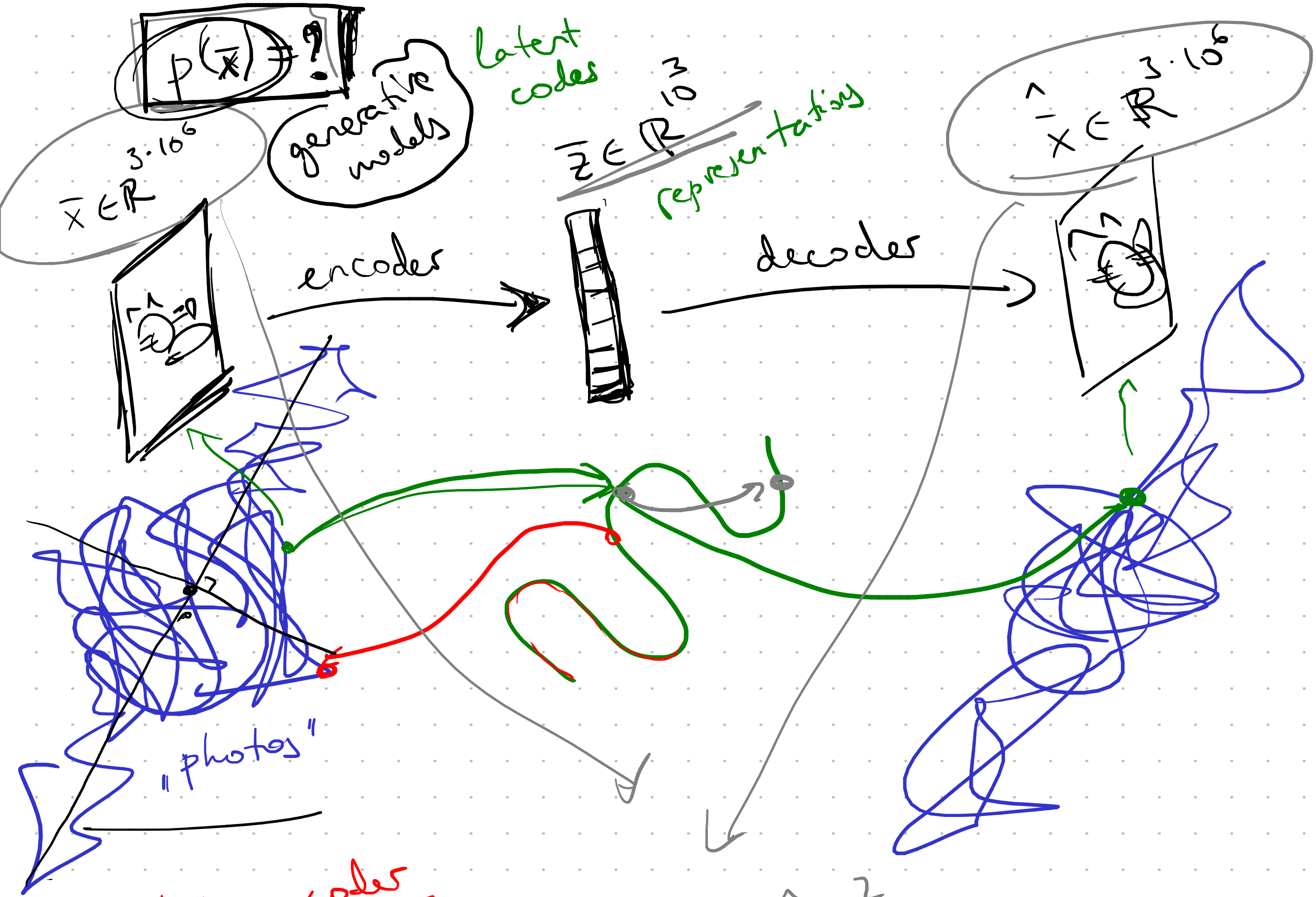
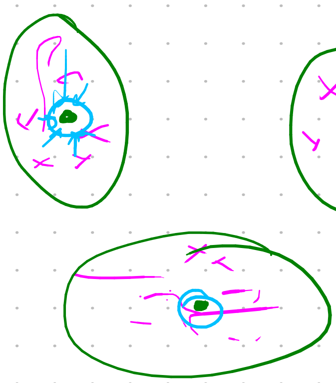
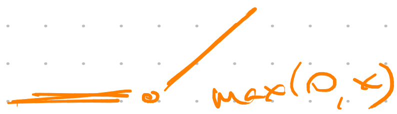


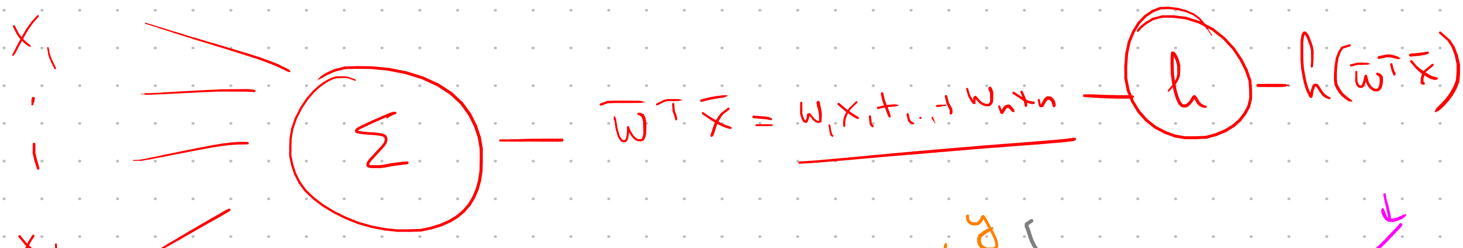
Autoencoder

$$L = \|\bar{x} - \hat{x}\|^2$$

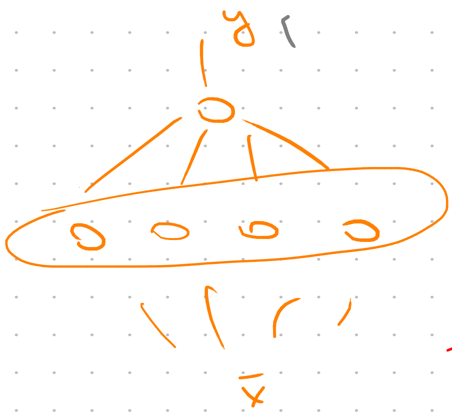


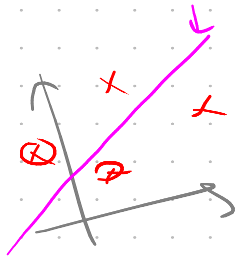


activation function



$\bar{w}$





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

D - data

$\bar{\theta}$ - model params

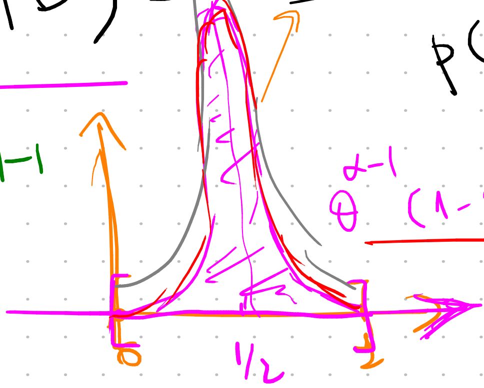
posterior prior likelihood

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

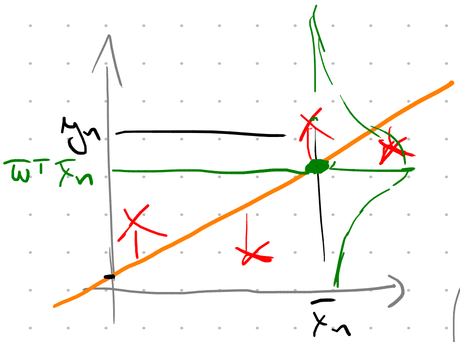
$p(D|\theta) = \theta^h(1-\theta)^t$

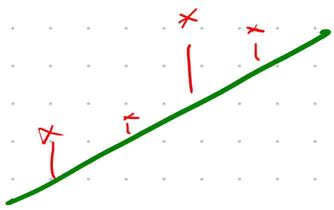
$p(\text{heads}|\theta)$

$\theta^{(\alpha+h)-1}(1-\theta)^{(\beta+t)-1}$



$\theta^{\alpha-1}(1-\theta)^{\beta-1}$





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

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

$$y = \underbrace{\bar w^T \bar x}_{1} + \varepsilon$$

homoscedesticity

$$\underbrace{\varepsilon \sim \mathcal N(0, \overset{1}{\sigma^2})}_{2}$$

$$p(D|\bar w) = p(\bar y|\bar w, X) \overset{0}{=} \prod_{n=1}^N p(y_n|\bar w, \bar x_n) \overset{1,2}{=}$$

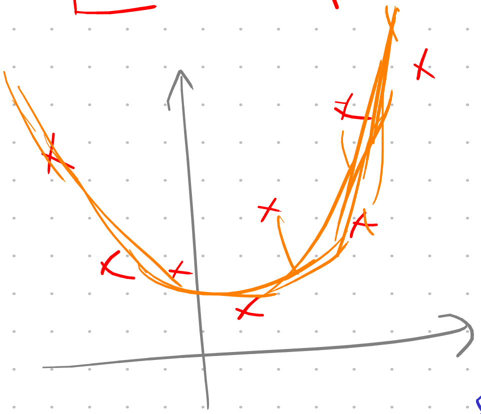
$$= \prod_{n=1}^N \mathcal N(y_n|\bar w^T \bar x_n, \sigma^2) = \prod_{n=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_n - \bar w^T \bar x_n)^2}$$

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

$$= \overbrace{-\frac{N}{2}\log(2\pi\sigma^2)}^{\text{const}} - \frac{1}{2\sigma^2}\boxed{\sum_{n=1}^N (y_n - \bar w^T \bar x_n)^2}$$

$$\bar w_{ML}: \quad \log p(D|\bar w) \xrightarrow{\bar w} \max \iff \xrightarrow[\bar w]{} \min$$

prior: $p(\bar{w})$



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

$x \xrightarrow{\bar{\varphi}} \begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix}$

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

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

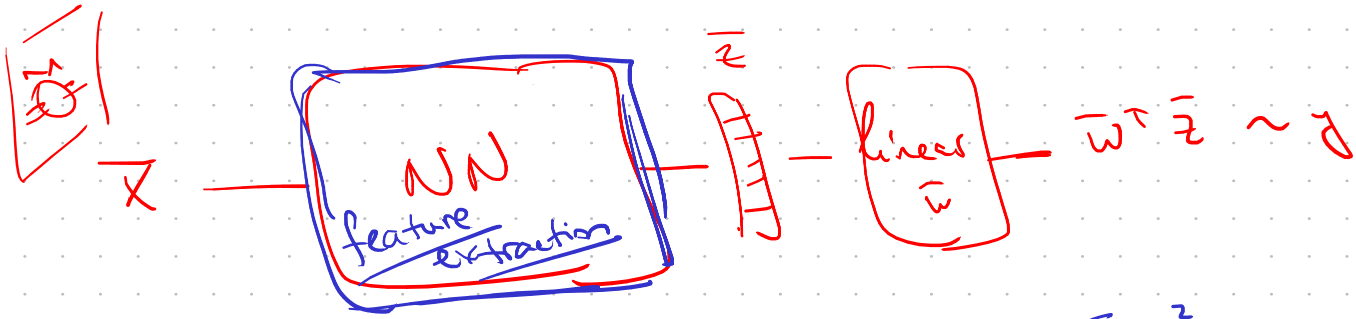
$\bar{x} \in \mathbb{R}^d$

$\bar{w} \in \mathbb{R}^d$

$\bar{x} \mapsto \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_d \\ x_1^2 \\ \cancel{x_1 x_2} \\ x_1 x_3 \\ \vdots \\ x_d^2 \end{pmatrix}$

$\frac{d(d+1)}{2}$

$\bar{w} \in \mathbb{R}$



$$p(\bar{w}) = N(\bar{w} \mid \bar{0}, \alpha \cdot I) = Const \cdot e^{-\frac{\alpha}{2} \cdot \sum w_i^2}$$

$$\log p(\bar{w} \mid D) = \log p(\bar{w}) + \log p(D \mid \bar{w}) - \underbrace{\cancel{\log p(D)}}_{Const} =$$

$$= Const + \log p(D \mid \bar{w}) + \log p(\bar{w}) =$$

$$= Const - Const \cdot \sum_n (y_n - \bar{w}^T \bar{x}_n)^2 - Const \cdot \sum w_i^2$$

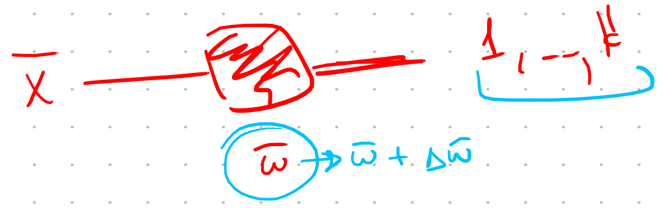
L_2-regularizer

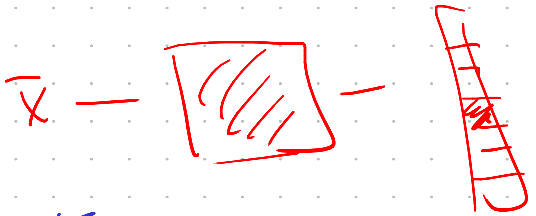
$\alpha \cdot \sum |w_i|$

L_1-reg.

Logistic regression

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





$p(C_k|\bar{x}) \in [0,1]$

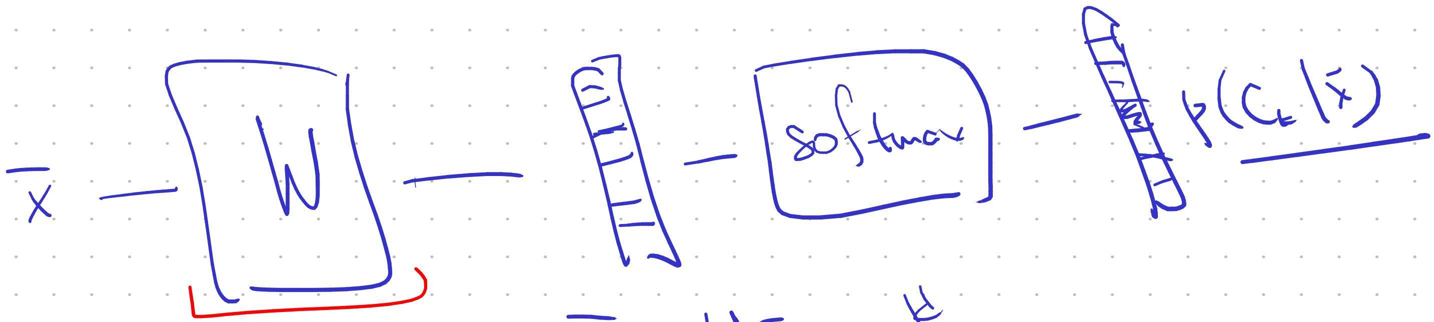
$\sum p(C_k|\bar{x}) = 1$

$$p(C_k|\bar{x}) = \frac{p(\bar{x}|C_k)\,p(C_k)}{p(\bar{x}|C_1)\,p(C_1) + \dots + p(\bar{x}|C_K)\,p(C_K)} =$$

$$= \frac{e^{\log p(C_k)\,p(\bar{x}|C_k)}}{\sum_s e^{\log p(C_s)\,p(\bar{x}|C_s)}}$$

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

$\log p(C_s)\,p(\bar{x}|C_s) \approx \bar{w}_s^T \bar{x}$



$\bar{a} = W\bar{x} \in \mathbb{R}^K$

$$\text{softmax}(\bar{a})_i = \frac{e^{a_i}}{\sum e^{a_j}}$$

Cross-entropy

$y = [0, 0, \dots, 1, \dots 0]$

$\boxed{\log p(C_k|\bar{x}) \to \max}$