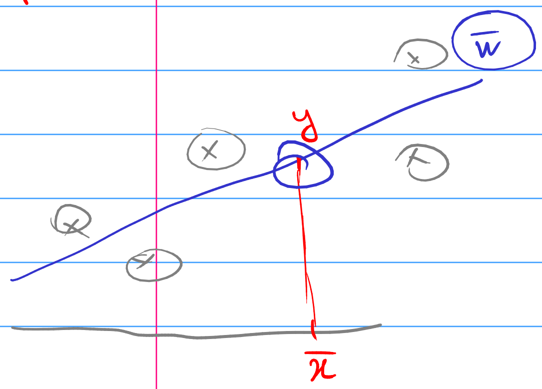
$$\underbrace{p(\theta|D)}_{\text{posterior}} = \frac{\overbrace{p(D|\theta)}^{\text{likelihood}}\overbrace{p(\theta)}^{\text{prior}}}{p(D)}$$

$$\theta_{ML} = \operatorname{argmax}\ p(D|\theta)$$

$$\theta_{MAP} = \operatorname{argmax}\ p(\theta|D) =$$

$$= \operatorname{argmax}\left(\log p(D|\theta) + \log p(\theta)\right)$$

Predictive distribution

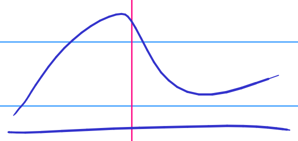


$$p(y|\bar{x}, D) =$$

$$= \int p(y, \bar{\theta}|\bar{x}, D)\, d\bar{\theta} =$$

$$= \int p(y|\theta, \bar{x}, \cancel{D})\, p(\theta|\cancel{\bar{x}}, D)\, d\theta$$

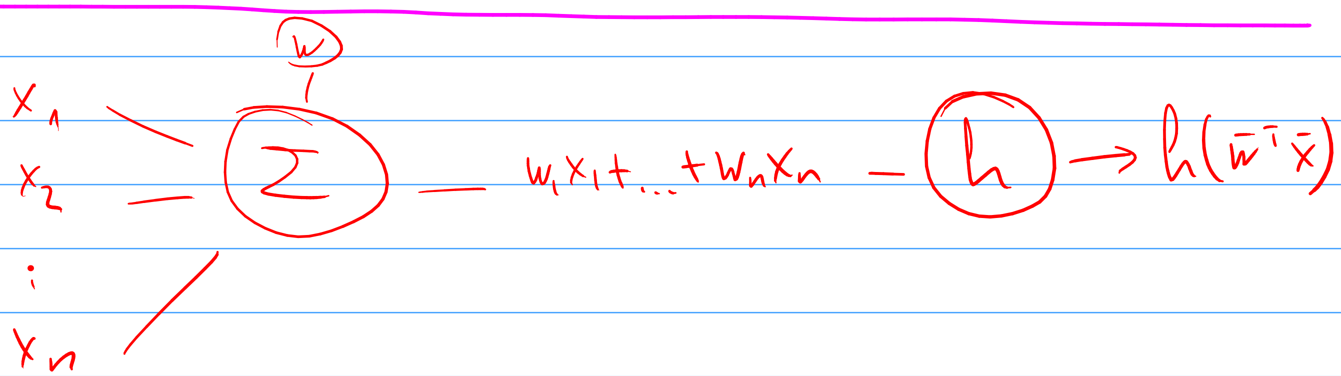
$$p(y|\bar{x}, D) = \int \underbrace{p(y|\theta, \bar{x})}_{\text{likelihood}}\, \underbrace{p(\theta|D)}_{\text{posterior}}\, d\theta =$$



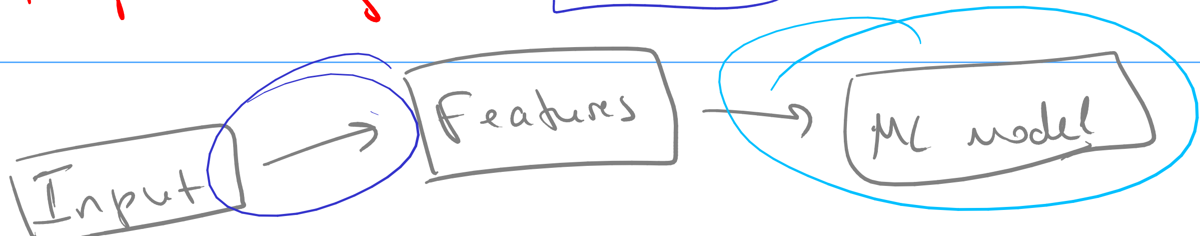
$$= \mathbb{E}_{p(\theta|D)}\left[p(y|\theta, \bar{x})\right] \approx \frac{1}{k}\sum_{\theta_k \sim p(\theta|D)} p(y|\theta_k, \bar{x})$$

Variational appr. $q(\theta) \approx p(\theta|D)$

Sampling MCMC $\theta_1, \theta_2, \ldots \theta_k \sim p(\theta|D)$



Deep Learning 2005/6

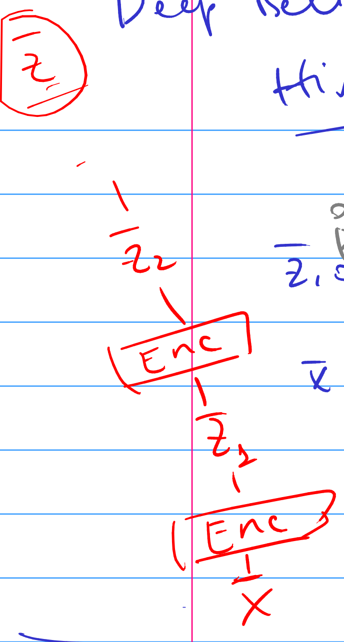


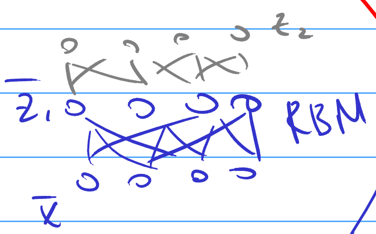
Deep Belief Networks

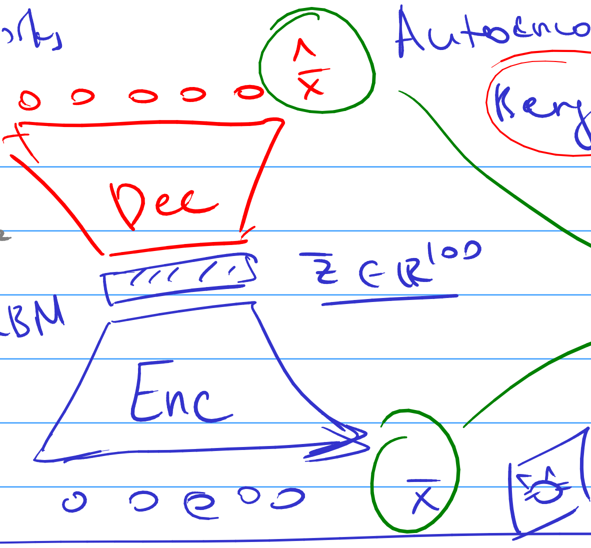
Hinton

Autoencoder

Bengio







$\bar{z} \in \mathbb{R}^{100}$

$L(\bar{x}, \hat{\bar{x}}) = \|\bar{x} - \hat{\bar{x}}\|_2^2$

Speech recognition

2009-10

DBN – GMM

Siri

HMM, RNN, Transf.

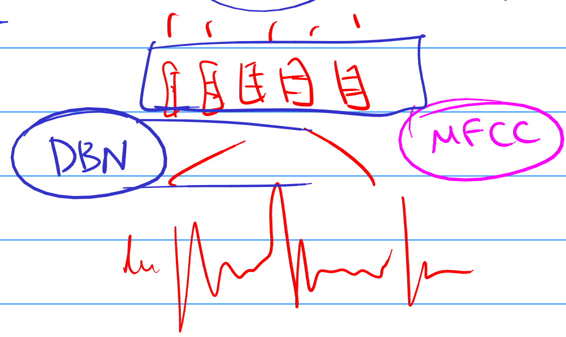


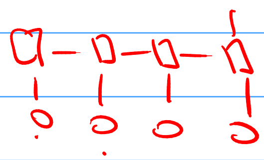
Image processing

NLP

CRF

RNN

LSTM



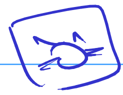
2017 Transformers

self-attention

attention mechanisms

Classification

$f: \bar{x} \longrightarrow \{C_1, \dots C_K\}$



$C_1\,,\; C_2$

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

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

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

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

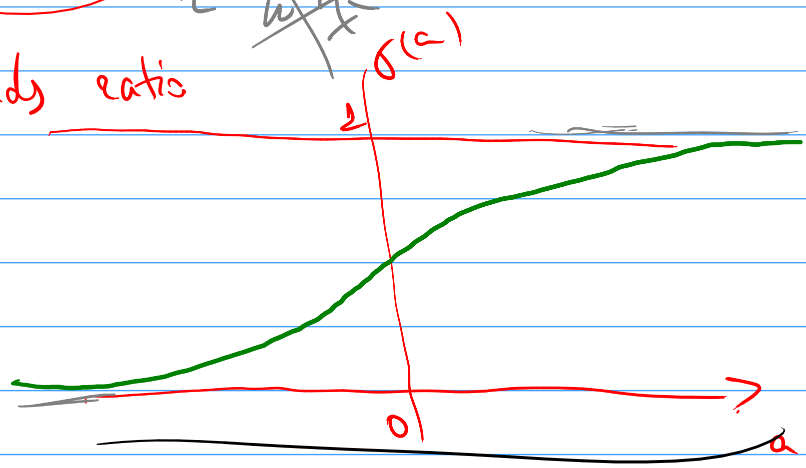
odds ratio

$(0, \infty)$

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

$(-\infty, \infty)$

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



logistic regression

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

$$p(D|\bar{w}) = \prod_n \sigma(\bar{w}^T\bar{x}_n)^{[y=1]}\,\left(1 - \sigma(\bar{w}^T\bar{x}_n)\right)^{[y=0]}$$

$y \in \{0, 1\}$
$\quad\;\; C_2 \;\; C_1$

$C_1, C_2, \ldots C_K$

$$\mathrm{softmax}(a_1, a_2, \ldots a_K) = \left(\ldots, \frac{e^{a_j}}{e^{a_1} + \ldots + e^{a_K}}, \ldots \right)$$

$$e^{a_j} = p(\bar{x} \mid C_j)\, p(C_j), \qquad a_j = \bar{w}_j^T \bar{x}$$

$$W = \underbrace{\begin{pmatrix} | \\ \bar{w}_j \\ | \end{pmatrix}}_{K} D$$