Discriminative

$\bar{x} \rightsquigarrow y$

Скотный прогноз

Generative

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

logistic regression

$$p(y_k|\bar{x}) = \frac{p(\bar{x}|y_k)p(y_k)}{\sum_l p(\bar{x}|y_l)p(y_l)}$$

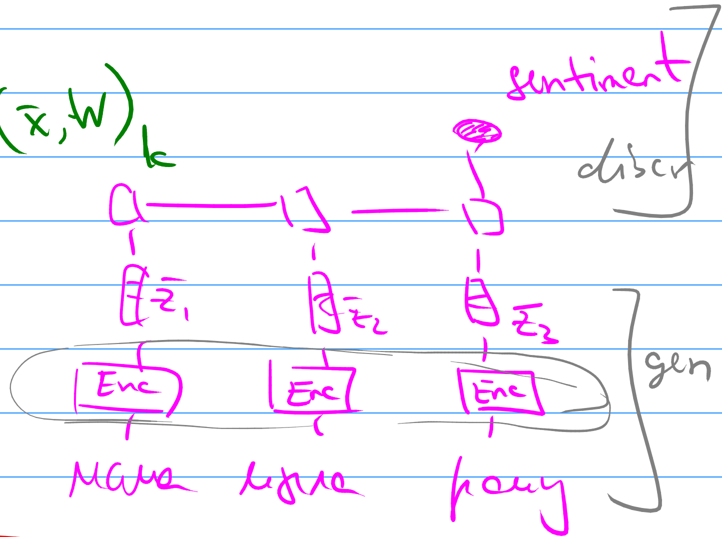
$$p(y|\bar{x}) = \frac{p(\bar{x}, y)}{p(\bar{x})}$$

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

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

$$p(y_k|\bar{x}) = \frac{e^{\bar{w}_k^T \bar{x}}}{\sum_l e^{\bar{w}_l^T \bar{x}}} = \mathrm{softmax}(\bar{x}, W)_k$$

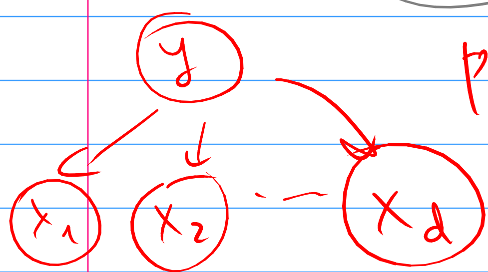
$$p(\bar{x}, y) = p(y|\bar{x})\, p(\bar{x})$$



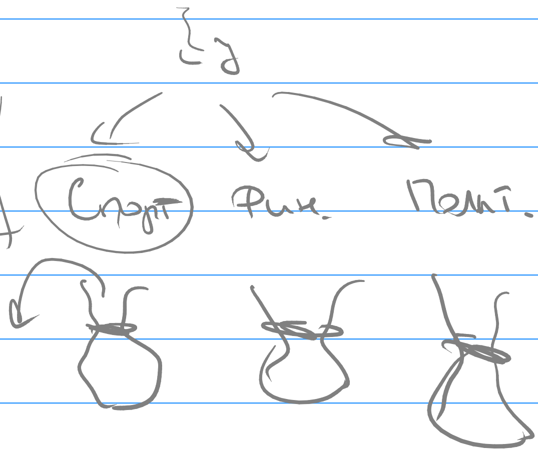
---

① Naive Bayes / log. regr.

$$p(\bar{x}, y) = \boxed{p(y)} \cdot \boxed{\prod_{i=1}^{d} p(x_i|y)}$$



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



$$p(y_k) = \frac{\#\{d \in C_k\} + 1}{N + K}\ ; \quad p(x_i|y_k) = \frac{\#\{i \in d,\ d \in C_k\} + 1}{\#\{\text{words in } C_k\} + |V|}$$

$$p(D|\theta) = \prod_{(\bar{x}, y) \in D} p(y) \prod_{j=1}^{N_d} p(x_j|y) =$$

$$= \prod_{n=1}^{N} \prod_{k=1}^{K} p(\bar{x}_n, y_k)^{[y_n=k]} =$$

$$= \prod_n \prod_k p(y_k)^{[y_n=k]} \cdot \prod_{j=1}^{N_d} \prod_{i=1}^{V} p(x_i|y_k)^{[y_n=k]\cdot[x_j=i]} =$$

$$= e^{\sum_n \sum_k \left([y_n=k] \underbrace{\log p(y_k)}_{=\theta_k} + \sum_j \sum_i [y_n=k, x_j=i] \cdot \underbrace{\log p(x_i|y_k)}_{=\theta_{ik}} \right)}$$

$\bar{\theta} = \{\theta_k, \theta_{ik}\}$

$$= e^{\sum_{n,k} \boxed{[y_n=k]} \cdot \theta_k + \sum_{n,k,i,j} \boxed{[y_n=k, x_j=i]} \cdot \theta_{ik}}$$

$$\boxed{p(y|\bar{x},\theta) \propto e^{\bar{\theta}^T \bar{f}}}$$

$$p(D|\theta) = e^{\sum_{(\bar{x},y)\in D} \sum_{s=1}^{S} f_s(\bar{x},y)\cdot \theta_s}$$

$$p(\bar{x},y|\theta) = e^{\sum_s f_s(\bar{x},y)\theta_s} = e^{\bar{\theta}^T \bar{f}}$$

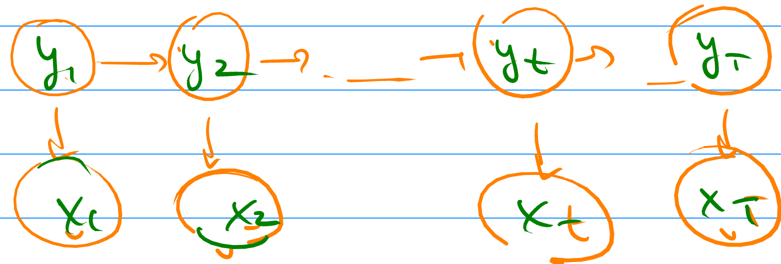
generative - discriminative pair

Log. regr.:

$$p(y|\bar{x},\theta) \propto e^{\bar{\theta}^T \bar{x}}$$

$$p(\bar{x},y|\theta) = \frac{1}{Z(\bar{x})} e^{\bar{\theta}^T \bar{x}} \cdot p(\bar{x})$$

② HMM / CRF



$$p(\bar{x},\bar{y}) = p(y_1)\,p(x_1|y_1)\,p(y_2|y_1)\dots$$

$$\dots p(y_t|y_{t-1})\,p(x_t|y_t)\dots p(y_T|y_{T-1})\,p(x_T|y_T)$$

$\pi_i = p(y_1=i)$

π, A, B $\quad p(y_t=j|y_{t-1}=i) = a_{ij}$ $\quad p(x_t=k|y_t=i) = b_{i}(k)$

$$p(D|\theta) = \prod_{(\bar{x},\bar{y})} p(\bar{x},\bar{y}|\theta) = \prod_{(\bar{x},\bar{y})} p(y_1)\,p(x_1|y_1) \dots p(x_T|y_T) =$$

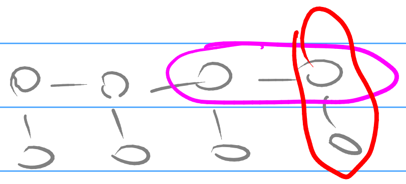
$$p(y_1) = \prod_{i=1}^{n} p(y_1=i)^{[y_1=i]} \qquad p(x_1|y_1) = \prod_i \prod_k p(x_t=k|y_t=i)^{[x_1=k,y_1=i]}$$

$$\theta_i \qquad \mu_{ik}$$

$$= e^{\sum_{(\bar{x},\bar{y})} \left[\sum_{i=1}^{n} [y_1=i] \log p(y_1=i) + \sum_{t=1}^{T} \sum_{i=1}^{n} \sum_{k=1}^{m} [x_t=k,\, y_t=i] \log p(x_t=k|y_t=i) + \sum_{t=2}^{T} \sum_{i,j=1}^{n} [y_t=j,\, y_{t-1}=i] \log p(x_t=j|x_{t-1}=i) \right]}$$

$$\theta_{ij}$$

$$p(D|\theta) = e^{\sum_{(\bar{x},\bar{y})} \left(\sum_i f_i(y_1)\theta_i + \sum_{ijt} f(y_t,y_{t-1})\theta_{ij} + \sum_{i,t,k} f(y_t,x_t)\mu_{ik} \right)}$$



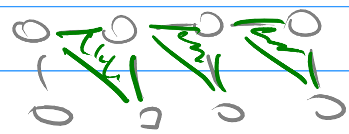
Conditional random fields

CRF

$$p(\bar{y}|\bar{x}) \propto e^{\sum_{ijt} f(y_t,y_{t-1})\theta_{ij} + \sum_{itk} f(y_t,x_t)\mu_{ik}}$$

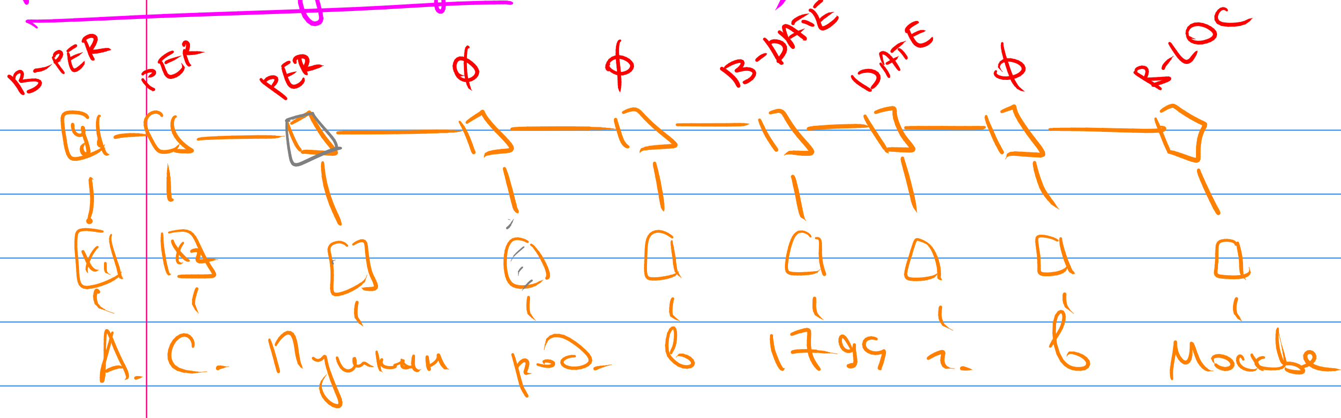
linear chain CRF

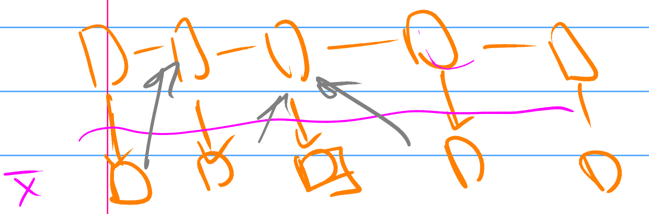
$$p(\bar{y}|\bar{x}) = \frac{1}{Z(\bar{x})} e^{\sum_t \sum_{s=1}^{S} \theta_s f_s(y_t,y_{t-1},x_t)}$$

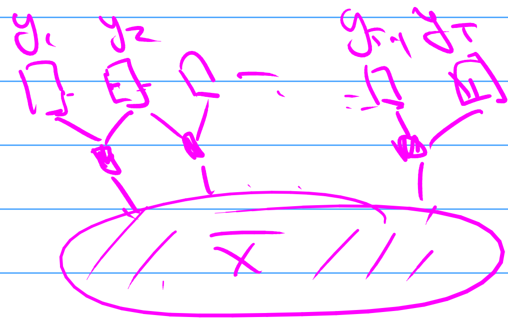


$$Z(\bar{x}) = \sum_{\bar{y}=y_1-y_T} p(\bar{x},\bar{y})$$

Named entity recognition (NER)

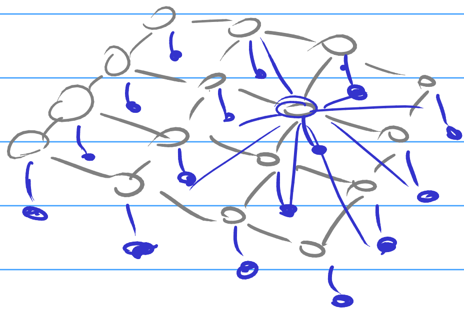






$Z(\bar{x})$

$$p(\bar{y}|\bar{x}) = \frac{1}{Z(\bar{x})} e^{\sum_t \sum_s \theta_s \cdot f(y_t, y_{t-1}, \bar{x})}$$



MRF - Markov random field

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

Generative models

Explicit density

Implicit density

$p(\bar{x}) = \dots \bar{\theta}$

$\bar{z} \longrightarrow \boxed{} \longrightarrow \bar{x}$

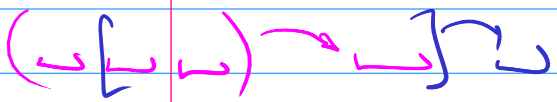
$$p(\bar{x}|\bar{\theta}) = p(x_1, x_2, \dots, x_d|\bar{\theta}) = \approx q(\bar{x}|\bar{\theta})$$

$$= p(x_1|\bar{\theta})\, p(x_2|x_1, \bar{\theta}) \dots p(x_i|\bar{x}_{1..i-1}, \bar{\theta}) \dots$$

$$\dots p(x_d|\bar{x}_{1..d-1}, \theta)$$

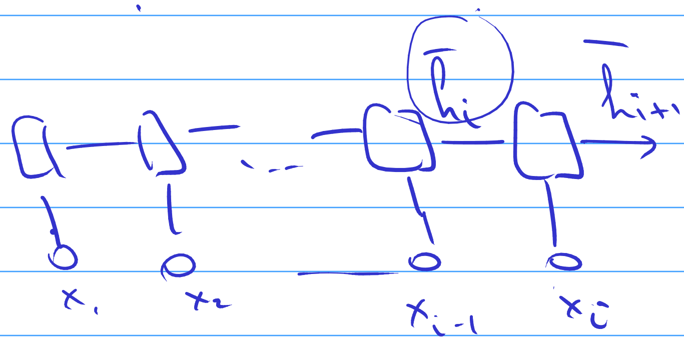
$p(x_i|\bar{x}_{1..i-1}, \bar{\theta})$ – autoregressive model

$p(x_i|\bar{x}_{i-k, i-1}, \bar{\theta})$



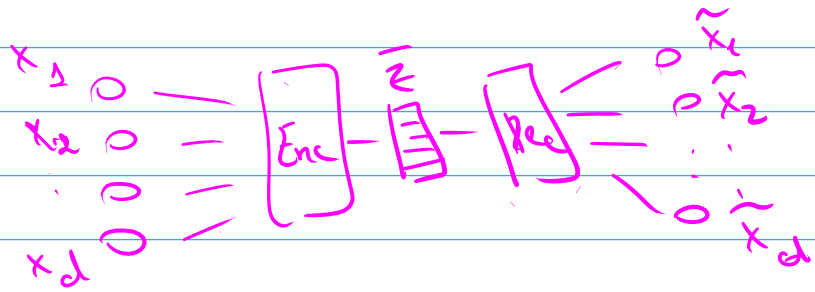
$p(x_i|\bar{h}_i, \bar{\theta})$, где

$\bar{h}_i = f(x_{i-1}, \bar{h}_{i-1}, \bar{\theta})$

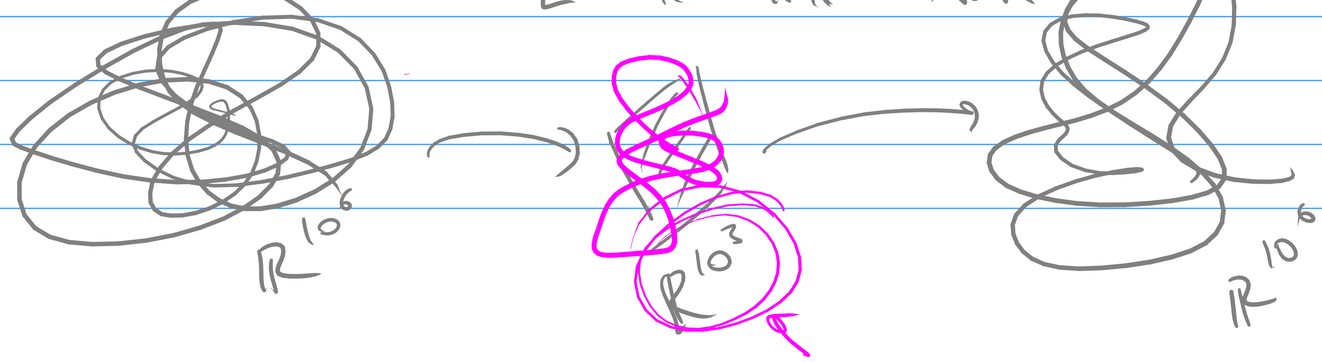


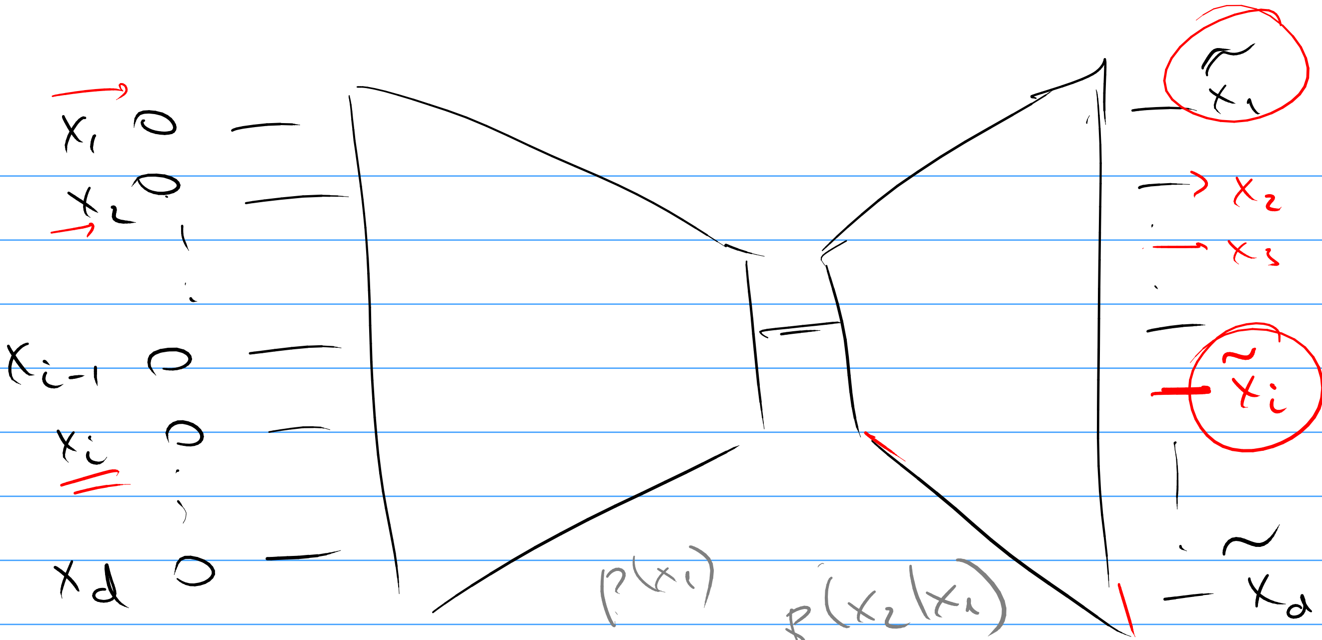
MADE – Masked Autoenc. for Distr. Estim.

$p(\bar{x})$ Autoencoder

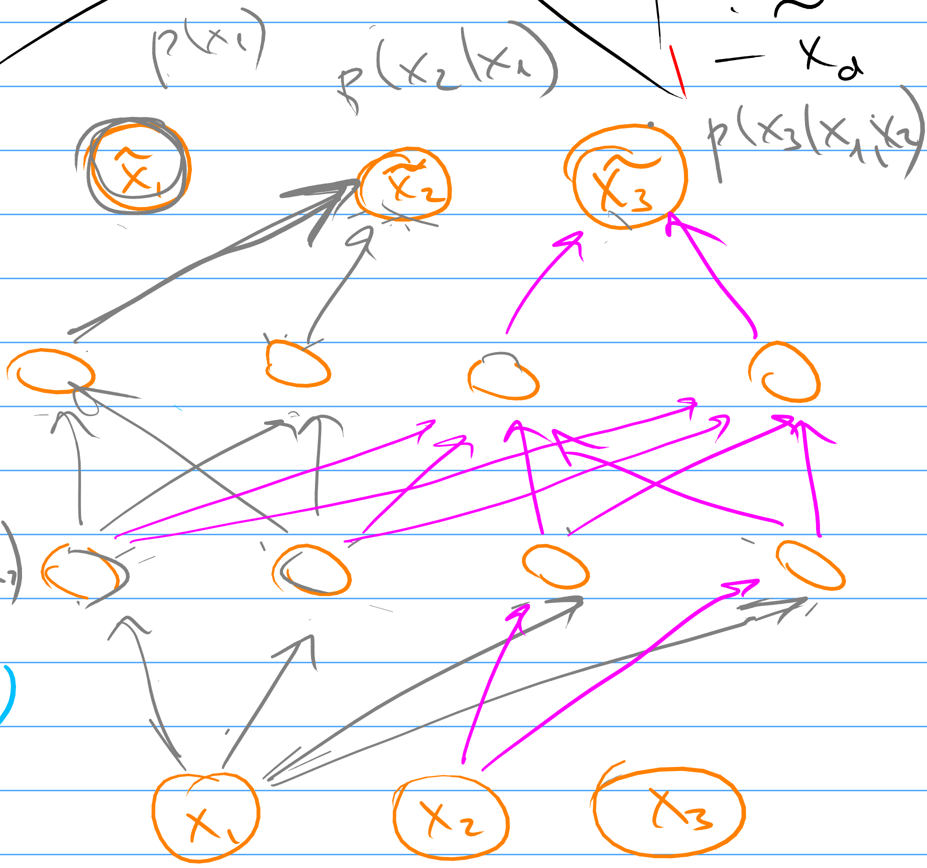


$L = \|\tilde{x} - \bar{x}\| \to \min$





$p(x_i \mid \bar{x}_{1,\dots,i-1})$



$$p(x_1, x_2, x_3) =$$

$$= p(x_1)\, p(x_2 \mid x_1)\, p(x_3 \mid x_1, x_2)$$

$$= p(x_2)\, p(x_3 \mid x_2)\, p(x_1 \mid x_2, x_3)$$

Pixel RNN   Pixel CNN