conformer

Output

Decoder

Language models

Pronounciation

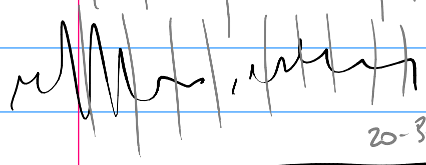
Acoustic models — HMM

Recurrent Transformers

Transducers

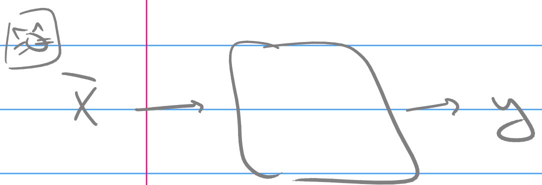
Acoustic features

MFCC $\mathbb{R}^{40}$



20-30ms

2009-2010 DBN → Siri, Cortana



$\bar{x} \to \; \to y$ "cat"

$p(y|\bar{x})$ discr.

$p(\bar{x}, y)$ generative

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

Naive Bayes

$\bar{x} = (x_1, \dots, x_k)$

y

bag of words

$$\boxed{p(\bar{x}|y) = \prod_{k=1}^{K} p(x_k|y)}$$

generative

$$p(y, \bar{x}) = p(y) \cdot \prod_{k=1}^{K} p(x_k|y)$$

Новости

спорт   полит   рын.

Logistic regression

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

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

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

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

$$p(\bar{y}|\bar{x}) \propto e^{0}$$

discriminative

$$p(y|\bar{x}) = \frac{1}{Z(\bar{x})} e^{\bar{w}_y^T\bar{x}} = \frac{1}{Z(\bar{x})} e^{w_1 x_1} e^{w_2 x_2} \ldots$$

$$p(y_k|\bar{x}) = \frac{e^{\bar{w}_k^T\bar{x}}}{\Sigma \ldots}$$

$$p(y_k|\bar{x}) \propto e^{\bar{w}_k^T\bar{x}}$$

Generative-discr. pair:

NB

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

LR

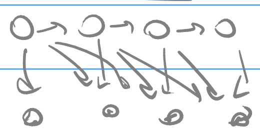
$$\boxed{p(y|\bar{x})} = \frac{1}{Z(\bar{x})} \cdot \prod_{i=1}^{n} f(y, x_i)$$

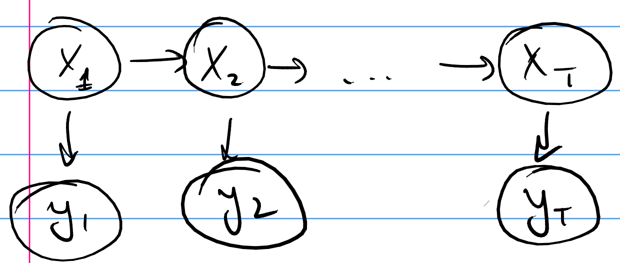
$$+\ p(\bar{x})$$

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

$p(A)$

Hidden Markov Models — generative





$A = (a_{ij})$ $\quad a_{ij} = p(x_{t+1}=j \mid x_t = i)$

$B = (b_j(k))$ $\quad b_j(k) = p(y_t = k \mid x_t = j)$

$$p(\bar{x},\bar{y}) = \underbrace{p(x_1)}_{\pi}\, p(y_1|x_1)\, p(x_2|x_1) \cdot p(y_2|x_2) \ldots p(x_T|x_{T-1})\, p(y_T|x_T)$$

A

B

$$p(\bar{x},\bar{y}) = \frac{1}{Z} \cdot e^{\sum_{i,j} \boxed{\ln a_{ij}} \cdot \left(\sum_t [x_{t-1}=i, x_t=j]\right) + \boxed{\ln b_j(k)} \cdot \left(\sum_t [x_t=j, y_t=k]\right)}$$

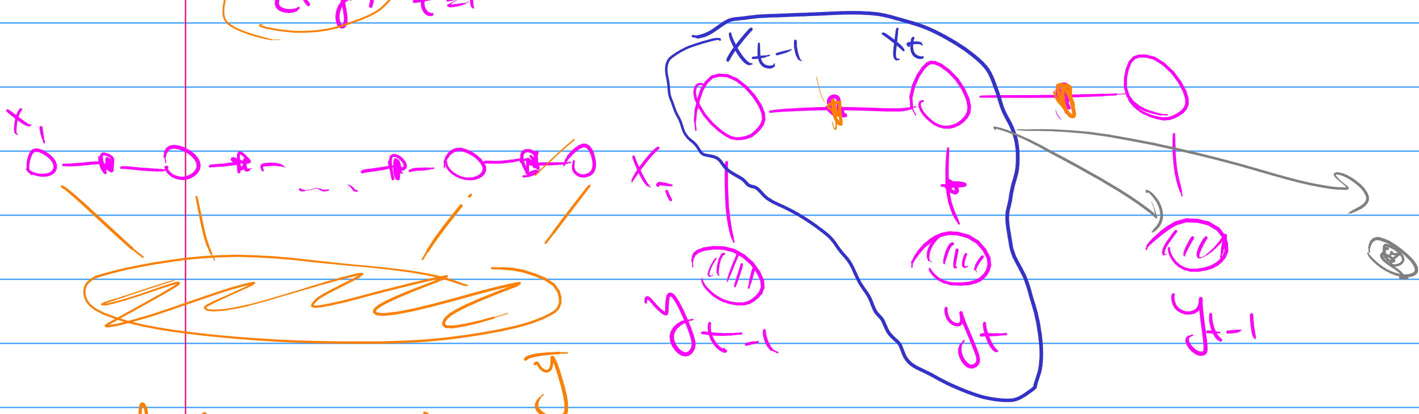
θ_s $\quad f_s(x_{t-1}, x_t, y_t)$

Conditional random field, CRF

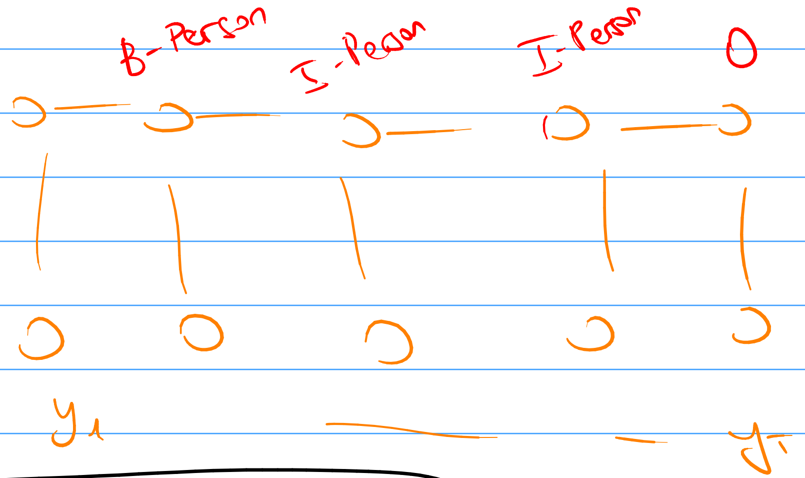
linear chain CRF

$$p(\bar{y},\bar{x}) = \frac{1}{Z} \prod_{t=1}^{T} e^{\sum_{s=1}^{S} \theta_s f_s(x_{t-1}, x_t, y_t)} \qquad f_s = \begin{cases} [\,\ldots\,] \\ [\,\ldots\,] \end{cases} \quad \text{HMM}$$

$$p(\bar{x}|\bar{y}) = \frac{1}{Z(\bar{y})} \prod_{t=1}^{T} e^{\sum_{s=1}^{S} \theta_s f_s(x_{t-1}, x_t, y_t, y_{t+1}, y_{t+2})}$$



$$f_s(x_{t-1}, x_t, \bar{y})$$



---

discr.

$$p(y|\bar{x})_{\bar{\theta}}$$

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

gen.

$$\underset{\in \mathbb{R}^{3 \cdot 10^6}}{p(y,\bar{x})} = p(y|\bar{x})\, p(\bar{x})$$

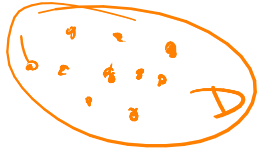
$$\in [0,1]$$

$$p(D|\bar{\theta}) = \prod p(\bar{x}|\theta) \xrightarrow[\bar{\theta}]{} \max$$

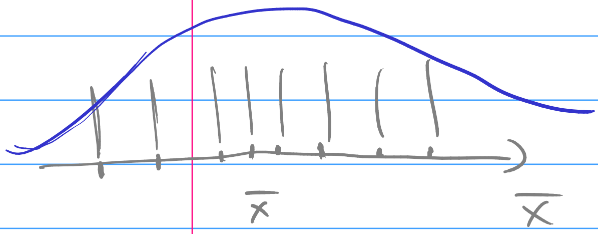
$$\bar{\theta}_{ML} = \arg\max \sum_{\bar{x} \in D} \log p(\bar{x}|\bar{\theta}) + \log p(\bar{\theta}) \longrightarrow \max$$

$N = |D|$

$$= \arg\min \left(- \sum_{\bar{x} \in D} \underset{= p_{data}(\bar{x})}{\left(1/N\right)} \cdot \log \frac{p(\bar{x}|\bar{\theta})}{(1/N)} \right) =$$



$$\boxed{KL(p\|q) = -\int p \ln \frac{q}{p} \, d\bar{x}}$$

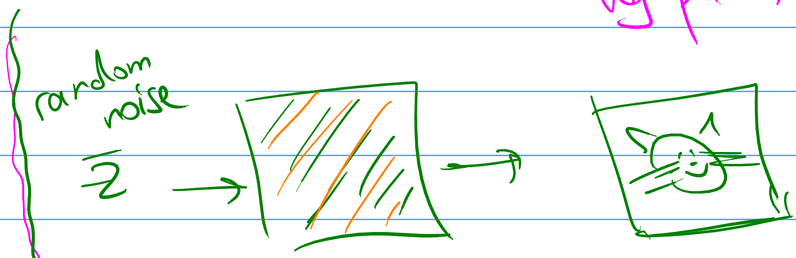


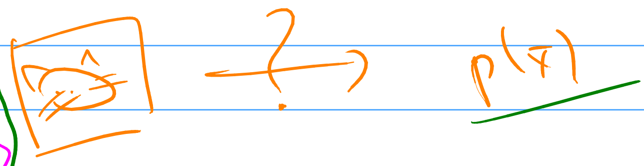
$$= \arg\min \left[KL\left(p_{data} \| p_{model}\right) - \log p(\theta) \right]$$

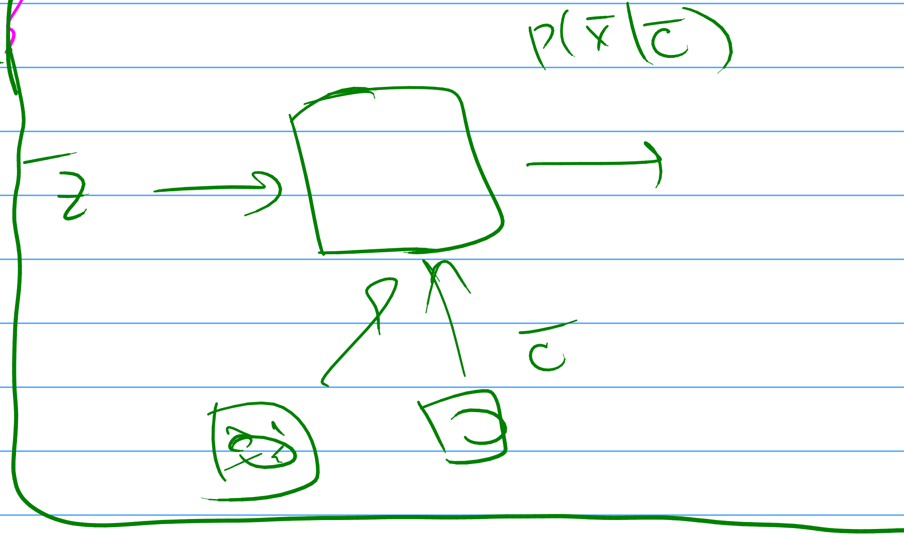
MAP:

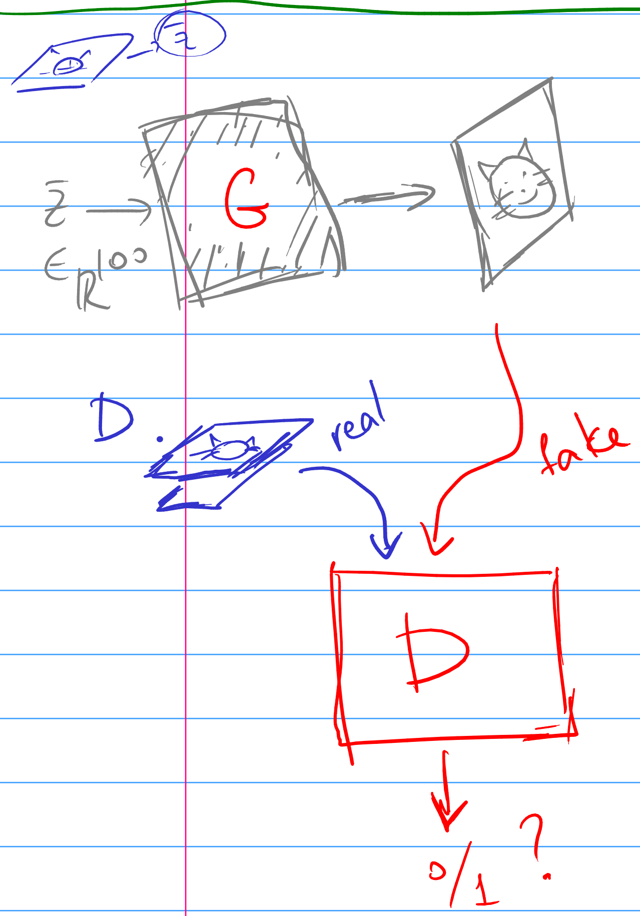
$$\theta_{MAP} = \arg\max p(\theta)\, p(D|\theta)$$

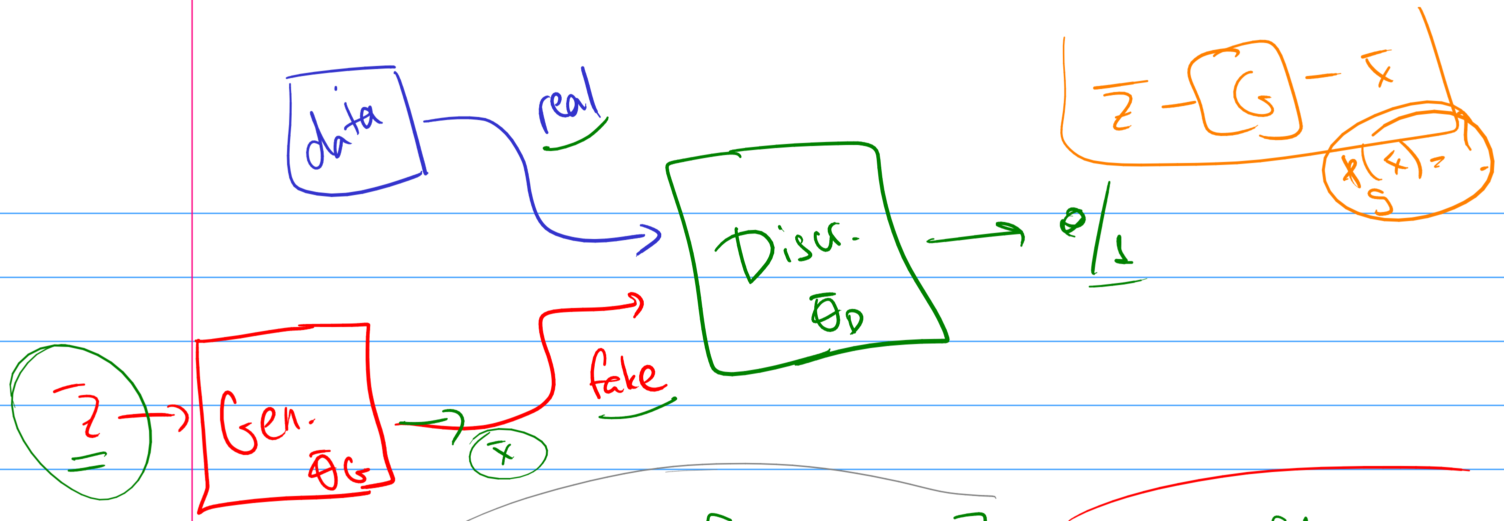
$$= \arg\min \left[-\log p(D|\theta) - \log p(\theta) \right]$$











max D

$$\mathbb{E}_{\bar{x}\sim p_{data}(\bar{x})}\left[\log D(\bar{x})\right] + \mathbb{E}_{\bar{x}\sim p_g(\bar{x})}\left[\log(1-D(\bar{x}))\right]$$

min G

$\bar{x} \sim G(\bar{z})$, ge $\bar{z} \sim p_z$

$$G: \mathbb{E}_{\bar{z}}\left[\log(1-D(G(\bar{z})))\right]$$

G → min

GAN:

$$\min_G \max_D \left[\mathbb{E}_{\bar{x}\sim p_{data}}\left[\log D(\bar{x})\right] + \mathbb{E}_{\bar{z}}\left[\log(1-D(G(\bar{z})))\right]\right]$$