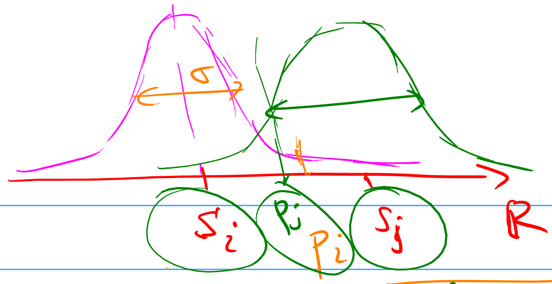
$X = \{x_1 \ldots x_N\} \longrightarrow \mathbb{R}$



Bradley-Terry models

$i \longrightarrow \gamma_i$

$$\boxed{p(i \succ j) = \frac{\gamma_i}{\gamma_i + \gamma_j}}$$

$p_i > p_j + \varepsilon \Rightarrow$ i выигр.

$p_j > p_i + \varepsilon \Rightarrow$ j - -

$|p_i - p_j| \le \varepsilon \Rightarrow$ ничья

TrueSkill

Expectation Propagation

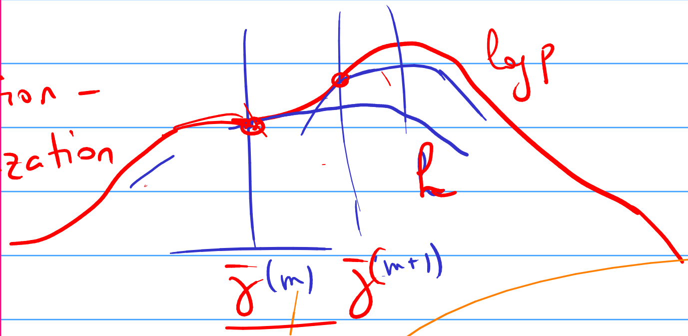
$$p(D|\bar\gamma) = \prod_{(i,j)\in D} \frac{\gamma_i}{\gamma_i+\gamma_j} \xrightarrow{\bar\gamma} \max$$

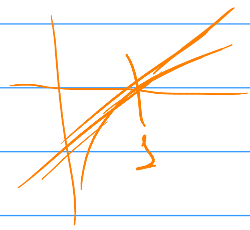
$$= \prod_i \prod_j \left(\frac{\gamma_i}{\gamma_i+\gamma_j}\right)^{w_{ij}}$$

$$\log p(D|\bar\gamma) = \sum_i \sum_j \left(w_{ij}\log\gamma_i - w_{ij}\log(\gamma_i+\gamma_j)\right) \xrightarrow[\bar\gamma]{} \max$$

MM-alg

minorization - maximization





$\boxed{\log a \le a - 1}$

$a = 1$

$$a = \frac{\gamma_i+\gamma_j}{\gamma_i^{(m)}+\gamma_j^{(m)}}$$

$$\log \frac{\gamma_i+\gamma_j}{\gamma_i^{(m)}+\gamma_j^{(m)}} \le \frac{\gamma_i+\gamma_j}{\gamma_i^{(m)}+\gamma_j^{(m)}} - 1$$

$$\log\frac{\gamma_i}{\gamma_i+\gamma_j} = \log\gamma_i - \log(+) \ge \log\gamma_i - \frac{\gamma_i+\gamma_j}{\gamma_i^{(m)}+\gamma_j^{(m)}} + 1 - \log(\gamma_i^{(m)}+\gamma_j^{(m)})$$

$$L(\bar\gamma, \bar\gamma^{(m)}) \le \log p(D|\bar\gamma)$$

$$\gamma^{(m+1)} := \arg\max L(\gamma, \gamma^{(m)})$$

$$= \sum_i \sum_j w_{ij}\left(\log\gamma_i - \frac{\gamma_i+\gamma_j}{\gamma_i^{(m)}+\gamma_j^{(m)}} + 1 - \log(\gamma_i^{(m)}+\gamma_j^{(m)})\right)$$

$$\frac{\partial L}{\partial \gamma_k} = \sum_i \sum_j w_{ij} \left(\frac{[i=k]}{\gamma_k} - \frac{[i=k]+[j=k]}{\gamma_i^{(m)} + \gamma_j^{(m)}} \right) = 0$$

$$\frac{1}{\gamma_k} \underbrace{\left(\sum_i w_{ki}\right)}_{= W_k} - \sum_{\text{пары k-го}} \frac{\overbrace{N_{kj}}^{= w_{kj} + w_{jk}}}{\gamma_k^{(m)} + \gamma_j^{(m)}} = 0$$

$$\gamma_k^{(m+1)} = W_k \cdot \left(\sum_j \frac{N_{kj}}{\gamma_k^{(m)} + \gamma_j^{(m)}} \right)^{-1}$$

$$p(i>j) = \frac{\gamma_i}{\gamma_i + \theta\gamma_j} \qquad p(j>i) = \frac{\gamma_j}{\gamma_j + \theta\gamma_i}$$

$$p(i=j) = 1 - \ldots - \ldots = \frac{(..)(..) - \gamma_i(..) - \gamma_j(..)}{(\gamma_j + \theta\gamma_i)(\gamma_i + \theta\gamma_j)}$$

$$p(\pi(\{1 .. n\}))$$

Generalized EM

$$\theta^{(m+1)}: \; Q(\theta^{(m+1)}, \theta^{(m)}) > Q(\theta^{(m)}, \theta^{(m)})$$

Monte Carlo EM

$$Q(\theta, \theta^{(m)}) = \mathbb{E}_{z \sim p(z|x, \theta^{(m)})} \left[\log p(x, z|\theta) \right] \approx$$

$$\approx \frac{1}{R} \sum_{r=1}^{R} \log p(x, z^{(r)} | \theta) \xrightarrow{\theta} \max \qquad z^{(r)} \sim p(z|x, \theta^{(m)})$$

EM с апр. распр. $\cancel{\theta_{ML}}$

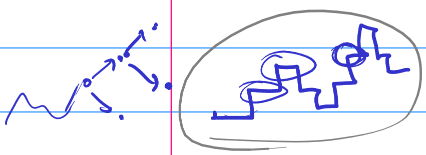
$$\theta_{MAP} = \arg\max\, p(\theta|X) = \arg\max\left(\underline{\log p(X|\theta) + \log p(\theta)}\right)$$

E-шаг $\quad Q(\theta, \theta^{(m)}) = E_Z\left[\log p(X, Z|\theta)\right]$

M-шаг $\quad \theta^{(m+1)} = \arg\max\left(Q(\theta, \theta^{(m)}) + \log p(\theta)\right)$

Hidden Markov models

Markov chain: $x_1 \to x_2 \to x_3 \to \dots \to x_{T-1} \to x_T$



f → ə → n → e: → m → ə

$$p(x_t | x_1 \dots x_{t-2}, x_{t-1}) = p(x_t | x_{t-1})$$

Language model: $\boxed{p(w_n | w_{n-1})}$

$p(w_n | w_{n-1}, w_{n-2})$

$p(x_1)$ π $\quad \boxed{p(x_t|x_{t-1})}$

$$x_1 \to x_2 \to x_3 \to \dots \to x_{t-1} \to x_t \to \dots \to x_T$$

$x_t \in \{1 \dots n\}$

$p(x_1 = i) = \pi_i \qquad p(x_t = j | x_{t-1} = i) = a_{ij}$

$$\boxed{\pi_i^* = \frac{[x_1 = i] + 1}{N + n}}$$

(π, A)

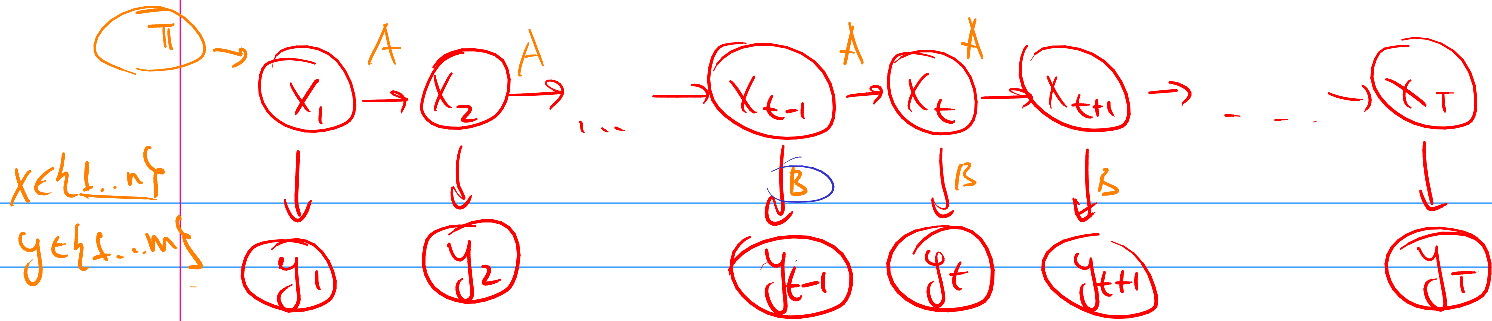
$$A = \begin{array}{cc} & j \\ i & \begin{pmatrix} & & \\ & a_{ij} & \\ & & \end{pmatrix} \end{array}$$

$$\boxed{a_{ij}^* = \frac{[x_t = j, x_{t-1} = i] + 1}{[x_{t-1} = i] + n}}$$

CG-content

$x_1\ x_2\ x_3 \qquad x_{t-1} x_t \qquad x_{t-1} \to x_t$

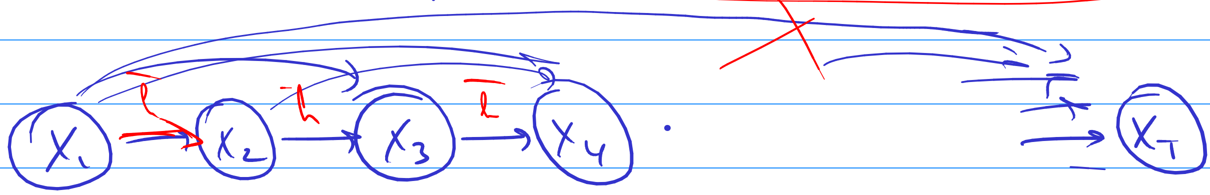
ATCGCCGAGCAGTTATGTACCGGCTGC



$$\pi_i = p(x_1 = i) \qquad p(x_t = j \mid x_{t-1} = i) = a_{ij}$$

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

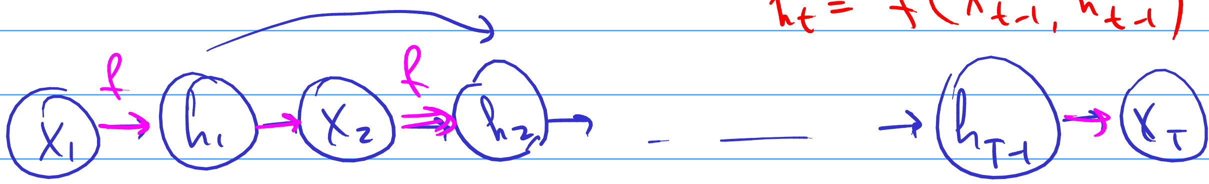
$$p(x_1, x_2, \ldots, x_{T-1}, x_T) = p(x_1)\, p(x_2 \mid x_1)\, p(x_3 \mid x_1, x_2) \ldots p(x_T \mid x_1 \ldots x_{T-1})$$



Markov chain $\quad p(x_t \mid x_1 \ldots x_{t-1}) = p(x_t \mid x_{t-1})$

RNN: $\quad p(x_t \mid x_1 \ldots x_{t-1}) = p(x_t \mid h_t)$

$$h_t = f(x_{t-1}, h_{t-1})$$



$D = d_1 \ldots d_T$ $\qquad p(D, Q \mid \pi, A, B) = \pi_{q_1} \cdot b_{q_1}(d_1) \cdot a_{q_1 q_2} \cdot$

$\boxed{Q = q_1 \ldots q_T}$ $\qquad \cdot\, b_{q_2}(d_2)\, a_{q_2 q_3} \ldots a_{q_{T-1} q_T}\, b_{q_T}(d_T)$

$$p(D \mid \pi, A, B) = \sum_{q_1 \ldots q_T} \pi_{q_1} b_{q_1}(d_1)\, a_{q_1 q_2} \ldots a_{q_{T-1} q_T}\, b_{q_T}(d_T) \to \max_{\pi, A, B}$$

π $1 \times n$ $\qquad A$ $n \times n$ $\qquad B$ $n \times m$

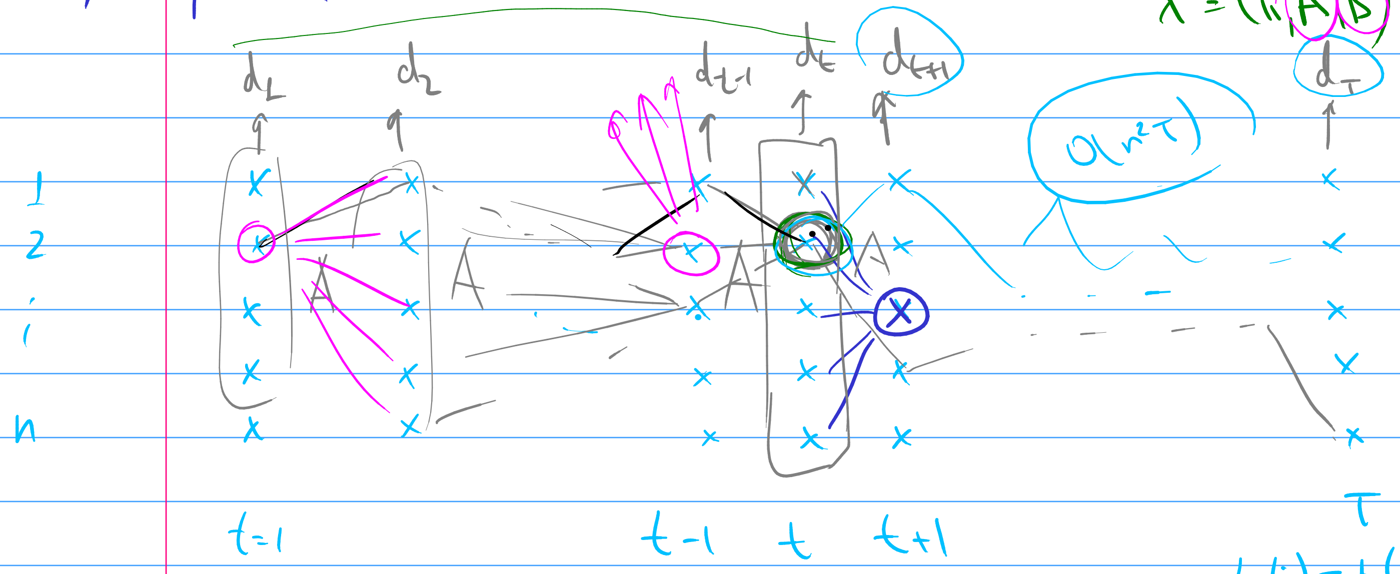
$x \in \{1..n\}$ $\qquad y \in \{1..m\}$

1) $p(D \mid \pi, A, B) = ?$

2) $p(Q \mid D, \pi, A, B) \xrightarrow{Q} \max$

3) $p(D \mid \pi, A, B) \rightarrow \max$

$\lambda = (\pi, A, B)$



$\alpha_T(i) = p(D, q_T = i \mid \lambda)$

1) $p(D \mid \pi, A, B)$ $\qquad \alpha_t(i) = p(d_1, \dots, d_t, q_t = i \mid \lambda)$

$\alpha_1(i) = p(d_1, q_1 = i \mid \lambda) = \pi_i\, b_i(d_1)$

$\alpha_{t+1}(i) = p(d_1 \dots d_t, d_{t+1}, q_{t+1} = i \mid \lambda) =$

$= \sum_j p(d_1 \dots d_t, d_{t+1}, q_{t+1} = i, q_t = j \mid \lambda) =$

$= \sum_j p(d_1 \dots d_t, q_t = j \mid \lambda)\, p(d_{t+1}, q_{t+1} = i \mid q_t = j, \cancel{d_1 \dots d_t}, \lambda)$

$p(q_{t+1} = i \mid q_t = j, \lambda) \cdot p(d_{t+1} = d_{t+1} \mid q_{t+1} = i, \cancel{q_t}, \lambda)$

$= \sum_j \alpha_t(j)\, a_{ji}\, b_i(d_{t+1})$

$$\boxed{\begin{aligned} \alpha_1(i) &= \pi_i b_i(d_1) \\ \alpha_{t+1}(i) &= \sum_j \alpha_t(j) a_{ji} b_i(d_{t+1}) \end{aligned}}$$

$p(D \mid \lambda) = \sum_i \alpha_T(i)$

$\beta_t(i) = p(d_{t+1}, \dots, d_T \mid q_t = i, \lambda)$

$$\beta_T(i) = 1$$

$$\beta_t(i) = \sum_j p(d_{t+1}, d_{t+2}, \dots, d_T, q_{t+1}=j \mid q_t = i, \lambda) =$$

$$= \sum_j p(d_{t+2}, \dots, d_T \mid q_{t+1} = j, \cancel{d_{t+1}}, \cancel{q_t = i}, \lambda) \cdot p(d_{t+1}, q_{t+1} = j \mid q_t = i, \lambda) =$$

$$\boxed{\begin{aligned} &\beta_T(i) = 1 \\ &\beta_t(i) = \sum_{j=1}^{n} \beta_{t+1}(j)\, a_{ij}\, b_j(d_{t+1}) \end{aligned}}$$

$$\beta_1(i) = p(d_2 \dots d_T \mid q_1 = i, \lambda)$$

$$p(D \mid \lambda) = \sum_i p(d_1, d_2 \dots d_T, q_1 = i \mid \lambda) =$$

$$= \sum_i \pi_i\, b_i(d_1)\, \beta_1(i)$$

---

2) $p(Q \mid D, \lambda) \xrightarrow{Q} \max$ inference в HMM

$p(q_t \mid D, \lambda) \xrightarrow{q_t} \max$

$$p(q_t = i \mid D, \lambda) = \frac{p(D, q_t = i \mid \lambda)}{p(D \mid \lambda)} = \frac{p(d_1 \dots d_t, q_t = i, d_{t+1} \dots d_T \mid \lambda)}{p(D \mid \lambda)}$$

$$= \frac{p(d_1 \dots d_t, q_t = i \mid \lambda) \cdot p(d_{t+1} \dots d_T \mid q_t = i, \cancel{d_1 \dots d_t}, \lambda)}{p(D \mid \lambda)}$$

$$\gamma_t(i) = p(q_t = i \mid D, \lambda) \propto \alpha_t(i) \cdot \beta_t(i)$$

Viterbi algorithm

$$Q^* = \arg\max p(q_1 \ldots q_T \mid D, \lambda) = \arg\max p(q_1 \ldots q_T, D \mid \lambda)$$

$$\delta_t(i) = \max_{q_1 \ldots q_{t-1}} p(q_1 q_2 \ldots q_{t-1}, q_t = i, d_1 \ldots d_t \mid \lambda)$$

$$\delta_1(i) = \pi_i b_i(d_1)$$

$$\delta_{t+1}(i) = \max_{q_1 \ldots q_t = j} p(q_1 \ldots q_{t-1}, q_t^{=j}, q_{t+1} = i, d_1 \ldots \underline{d_{t+1}} \mid \lambda) =$$

$$= \max_j \left[\max_{q_1 \ldots q_{t-1}} p(q_1 \ldots q_{t-1}, q_t = j, d_1 \ldots d_t \mid \lambda) \cdot a_{ji} b_i(d_{t+1}) \right]$$

$$\delta_{t+1}(i) = \max_j \left[\delta_t(j) a_{ji} b_i(d_{t+1}) \right]$$

---

3) $p(D \mid \lambda) \xrightarrow[\lambda = (\pi, A, B)]{} \max$ $\quad Q$

$$p(D, Q \mid \lambda) = \pi_{q_1} a_{q_1 q_2} b_{q_1}(d_1) \ldots a_{q_{T-1} q_T} b_{q_T}(d_T)$$

$$L'(\lambda, \lambda^{(m)}) = E_{Q \sim \lambda^{(m)}} [\ln p(D, Q \mid \lambda)] =$$

$$= E_{Q \sim \lambda^{(m)}} \left[\ln \pi_{q_1} + \ln a_{q_1 q_2} + \ln b_{q_1}(d_1) + \ldots \right] \rightarrow \max$$

$\ln p(y = d_1 \mid x = q_1)$

$$\pi_i^{(m+1)} = \frac{E[\#[q_1 = i]]}{N} = p(q_1 = i \mid D, \lambda^{(m)}) = \gamma_1(i)$$

$$a_{ij}^{(m+1)} = \frac{\sum_t p(q_t = i, q_{t+1} = j \mid D, \lambda^{(m)})}{\sum_t p(q_t = i \mid D, \lambda^{(m)})} = \frac{\sum_t \xi_t(i, j)}{\sum_t \gamma_t(i)}$$

$$\underset{\xi_t(i,j)}{\underline{p}}(q_t=i, q_{t+1}=j \mid D, \lambda^{(m)}) \propto p(q_t=i, q_{t+1}=j, D \mid \lambda^{(m)}) =$$

$$= p(d_1 - d_t, q_t=i \mid \lambda)\, p(q_{t+1}=j \mid q_t=i, \lambda)\, p(d_{t+1} \mid q_{t+1}=j)\, p(d_{t+2} -- d_T \mid q_{t+1}=j, \lambda) =$$

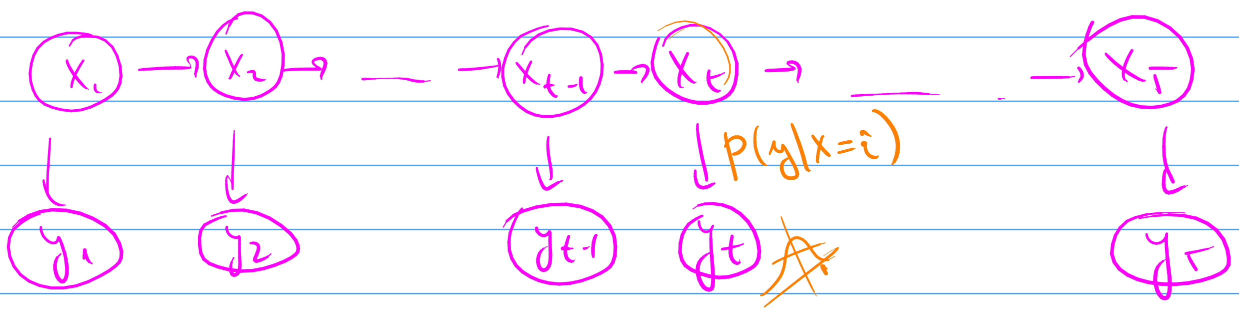
$$= \alpha_t(i)\, a_{ij}\, b_j(d_{t+1})\, \beta_{t+1}(j)$$

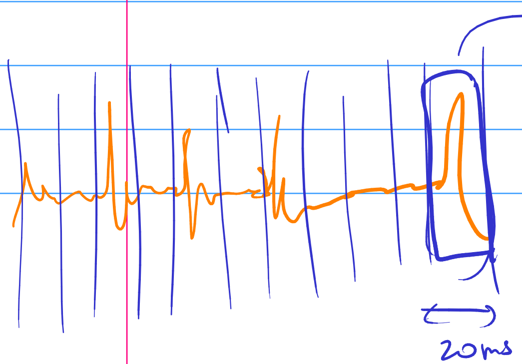
$$\boxed{b_i^{(m+1)}(k) = \frac{\sum_t p(q_t=i, d_t=k \mid D, \lambda^{(m)})}{\sum_t p(q_t=i \mid D, \lambda^{(m)})} = \frac{\sum_{t:d_t=k} \gamma_t(i)}{\sum_t \gamma_t(i)}}$$

$p(y|x=i)$

$$\lambda^{(0)} \xrightarrow{E} \alpha, \beta, \gamma, \xi$$
$$M$$
$$\lambda^{(1)} \xrightarrow{E} \alpha, \beta, \gamma, \xi$$
$$M$$
$$\lambda^{(2)}$$

Baum-Welch algorithm

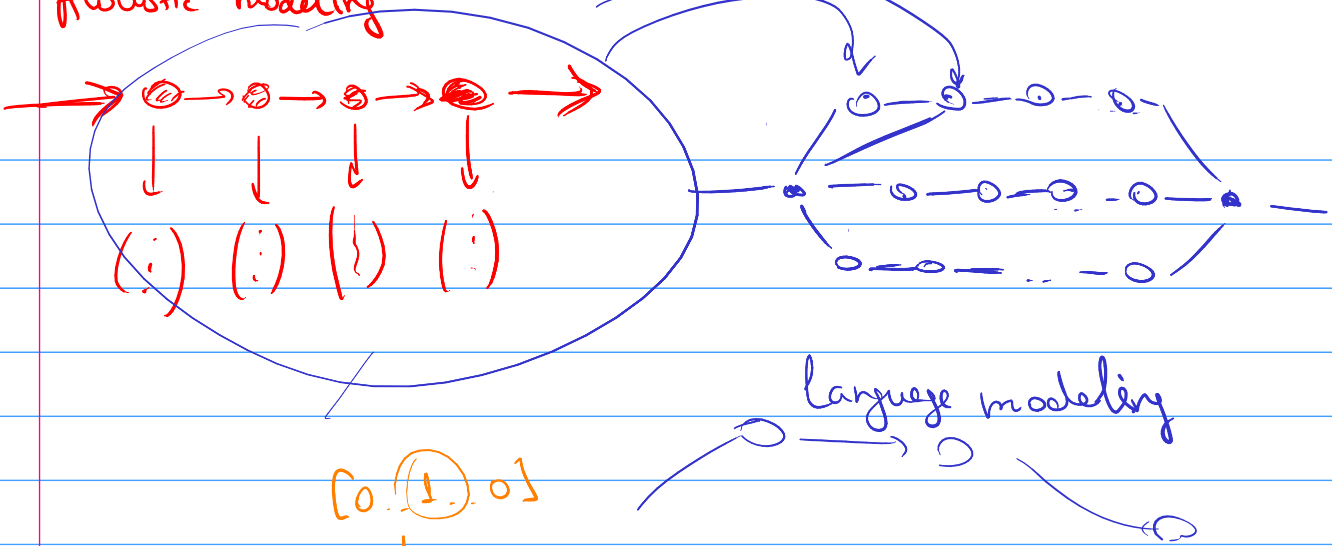




feat extr $\longrightarrow \mathbb{R}^{40}$

MFCC
mel-frequency
cepstral coeff.

Acoustic modeling



$[0 \ldots 1 \ldots 0]$

$$CE \quad \sum^{N} p_i \log q_i$$

$$H(p,q) = KL(p \| q) + H(q)$$

$$\ln \frac{e^{a_i}}{\sum_j e^{a_j}} \qquad a_i - \boxed{\ln\left(\sum_j e^{a_j}\right)}$$

softmax

logaddexp

exp1m