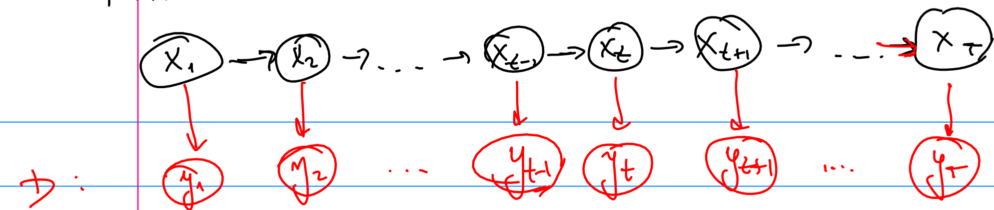
Markov chain

Hidden Markov model, HMM



$p(x_{t+1}|x_t \dots x_1) = p(x_{t+1}|x_t)$

$A = (a_{ij})$; $a_{ij} = p(x_t = j | x_{t-1} = i)$

$\pi = (\pi_i)$; $\pi_i = p(x_1 = i)$ $i = 1, \dots, n$

$p(q_1 q_2 \dots q_T | \pi, A) = \pi_{q_1} a_{q_1 q_2} a_{q_2 q_3} \dots a_{q_{T-1} q_T}$

$D = \{q_1, \dots q_T\}$ $a_{ij} = \frac{\#\{i \to j\}}{\#\{i\}}$ $\pi_i = \frac{\#\{q_1 = i\}}{|D|}$

$y_t \in \{1, \dots, m\}$ $p(y_t | x_t) = p(y_t | x_t, x_1 \dots x_{t-1}, y_1 \dots y_{t-1})$

$B = (b_i(k))$ $b_i(k) = p(y_t = k | x_t = i)$ $\lambda = (\pi, A, B)$

фонема Q : f → a → n → e → m → a

D : ...

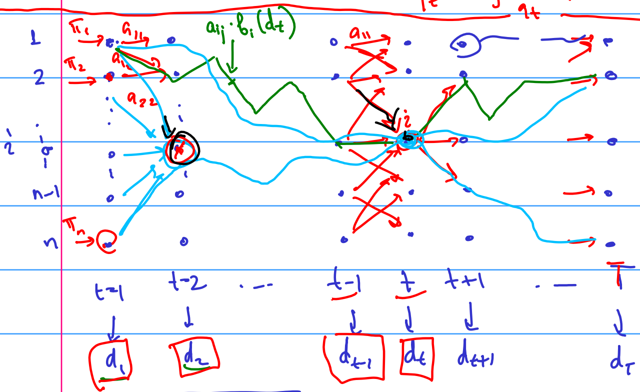
$b_i(k) = \frac{\#\{i : y = k\}}{\#\{i\}}$

$p(Q, D | \lambda) = p(q_1 \dots q_T, d_1, \dots, d_T | \pi, A, B) = \pi_{q_1} b_{q_1}(d_1) a_{q_1 q_2} b_{q_2}(d_2) \dots a_{q_{T-1} q_T} b_{q_T}(d_T)$

$p(D|\lambda) = p(d_1 \dots d_T | \pi, A, B) = \sum_{q_1 \dots q_T} p(Q, D | \lambda) = \sum_{q_1 \dots q_T} \pi_{q_1} b_{q_1}(d_1) \dots a_{q_{T-1} q_T} b_{q_T}(d_T)$

(1) $p(D|\lambda) = ?$ (2) $Q^* = \arg\max_Q p(Q|D, \lambda)$ (3) $\lambda^* = \arg\max_\lambda p(D|\lambda)$

$q_t^* = \arg\max_{q_t} p(q_t | D, \lambda)$



$Q = q_1 \dots q_T$ – путь слева направо

$p(Q|\lambda) = \pi_{q_1} a_{q_1 q_2} \dots a_{q_{T-1} q_T}$

$p(D|\lambda) = \sum_Q p(Q, D | \lambda)$

$\alpha_t(i) = p(d_1 \dots d_t, q_t = i | \lambda) =$

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

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

a_{ji} $p(q_t = i | q_{t-1} = j, \lambda) \cdot p(d_t | q_t = i, \lambda)$ $b_i(d_t)$

$\alpha_t(i) = p(d_1 \dots d_t, q_t = i | \lambda)$ $O(n^2 T)$

$\alpha_1(i) = \pi_i b_i(d_1)$

$\alpha_t(i) = \sum_j \alpha_{t-1}(j) \cdot a_{ji} \cdot b_i(d_t)$

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

$\beta_t(i) = p(d_{t+1} \dots d_T | q_t = i, \lambda)$ $O(n^2 T)$

$\beta_T(i) = 1$

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

$p(D|\lambda) = \sum_i \beta_1(i) \pi_i$

$\gamma_t(i) = p(q_t = i | D, \lambda)$

$\gamma_t(i) \propto \alpha_t(i) \beta_t(i)$

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

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

$a_{ij} b_j(d_{t+1})$

(2) $q_t^* = \arg\max_{q_t} p(q_t | D, \lambda)$

$\gamma_t(i) = p(q_t = i | D_{1..t}, D_{t+1..T}, \lambda) = \frac{p(q_t = i, D_{1..t}, D_{t+1..T} | \lambda)}{p(D|\lambda)}$

sum-product

$= \frac{p(D_{t+1..T} | q_t = i, D_{1..t}, \lambda) \cdot p(q_t = i, D_{1..t} | \lambda)}{p(D|\lambda)} = \frac{\beta_t(i) \cdot \alpha_t(i)}{p(D|\lambda)} = \frac{\alpha_t(i) \beta_t(i)}{\sum_j \alpha_t(j) \beta_t(j)}$

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

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

$\delta_t(i) = \max_{q_{t-1}} \left[\max_{q_1 \dots q_{t-2}} p(\dots) \right] =$

$\max_j \left[\delta_{t-1}(j) \cdot a_{ji} \cdot b_i(d_t) \right]$

$Q^* = \arg\max_Q p(Q|D, \lambda)$ – Viterbi, max-product

(3) $p(D|\lambda) \xrightarrow[\pi, A, B]{} \max$ $p(D, Q | \lambda) \xrightarrow[\pi, A, B]{} \max$ – простая задача   Baum–Welch

сложная задача $= \sum_Q p(D, Q | \lambda)$

$\#\{i\} = \mathbb{E}[\#\{q_t = i\}] = \sum_{t=1}^T \gamma_t(i)$

E-step: α, β, γ

M-step: π, A, B

$\xi_t(i, j) = p(q_t = i, q_{t+1} = j | D, \lambda)$

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

$\lambda^{(0)} = (\pi^{(0)}, A^{(0)}, B^{(0)})$ → $\alpha, \beta, \gamma, \xi$

$p(D|\lambda) \uparrow$ $\lambda^{(1)}$ ←

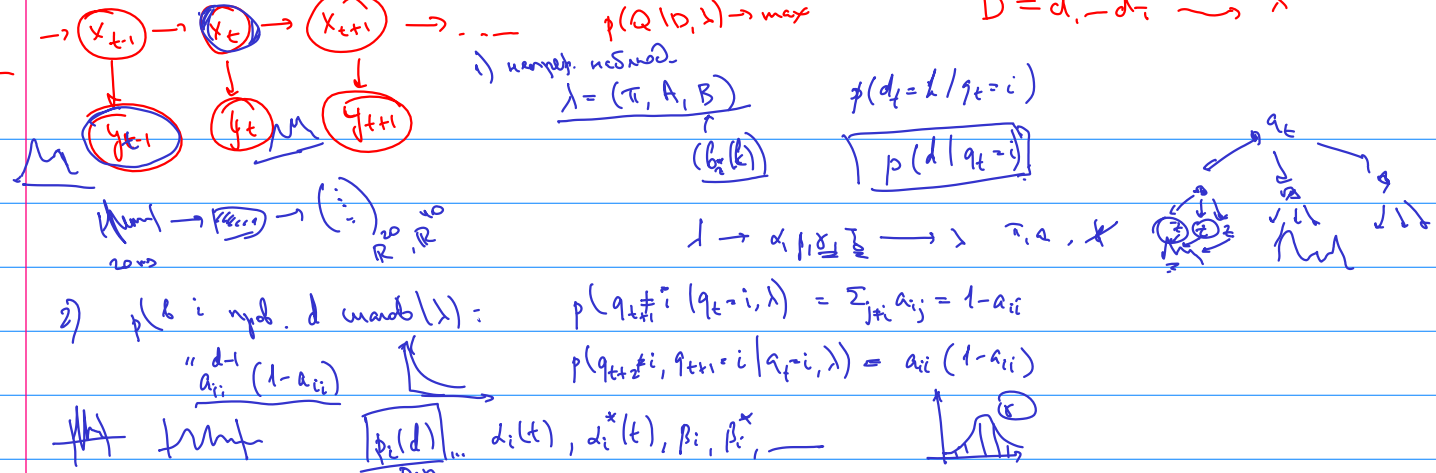
$\lambda^{(2)}$ ← → ....

$D = d_1 \dots d_T$

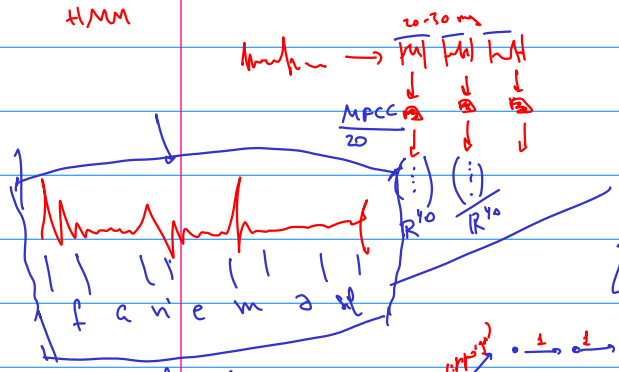
$\pi_i := \frac{\mathbb{E}[\#\{q_1 = i\}]}{|D|} = \frac{\sum_t \gamma_1(i)}{|D|}$

$a_{ij} := \frac{\mathbb{E}[\#\{i \to j\}]}{\mathbb{E}[\#\{i\}]} = \frac{\sum_t \xi_t(i, j)}{\sum_t \gamma_t(i)}$

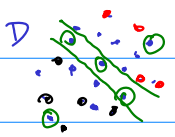
$b_i(k) = \frac{\mathbb{E}[\#_t\{i : d_t = k\}]}{\mathbb{E}[\#\{i\}]} = \frac{\sum_{t : d_t = k} \gamma_t(i)}{\sum_t \gamma_t(i)}$

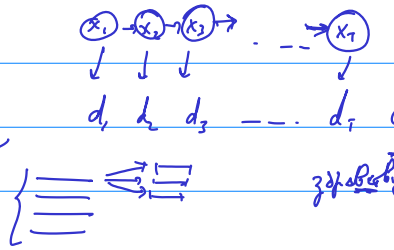


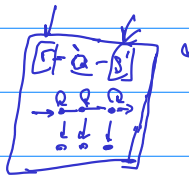
HMM

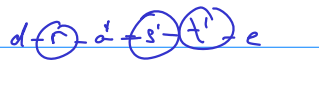


Active learning



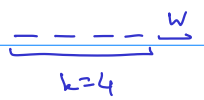




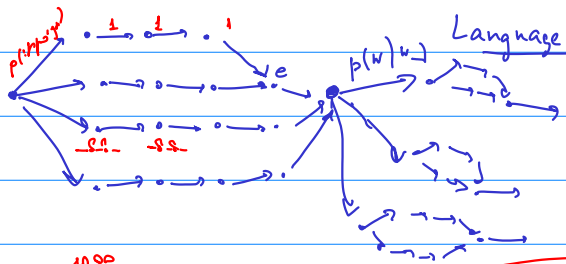


Language modeling $p(w_n | w_{n-1}, \ldots, w_1) = ?$

$p(w_n | w_{n-1} \ldots w_1) = p(w_n | w_{n-1} \ldots w_{n-k})$

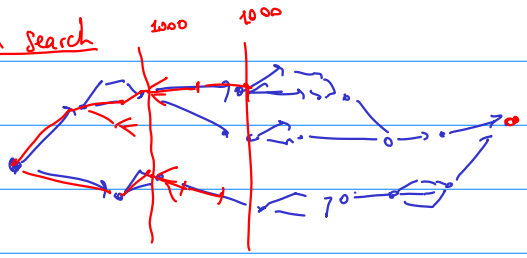


$\approx \frac{\#\{w_- \ldots w\} + 1}{\#\{w_-\} + |V|}$



Viterbi

beam search



H – S – H

RNN language models

WER word error rate

$\approx 5\%$

Google Translate – GNMT

HMM – GMM

HMM – DBN