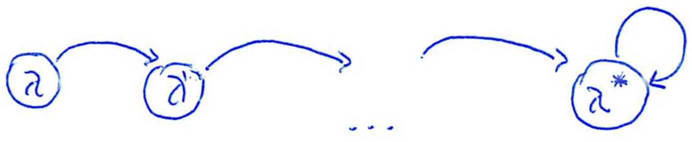
① $P(\vec{O} \mid \lambda)$

② $\vec{q}^{\,\lambda} = \arg\max P(\vec{q} \mid \vec{O}, \lambda)$ speech recognition

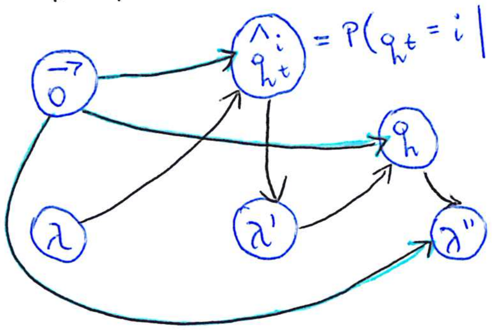
③ $\lambda^* = \arg\max_{\lambda} P(\lambda \mid \vec{O})$ build model given data

Build up a model



till any small changes will make it only worst

$\vec{q}$ - seq. of states that machine went through



$= P(q_t = i \mid \vec{O}, \lambda)$

$$\gamma_t(i) = P(q_t = i \mid \vec{O}, \lambda)$$

$$b_i(k) = P(O_t = k \mid q_t = i)$$

$$b_i(k) = \frac{\#\text{ in state } i \;\&\; \text{output } k}{\#\text{ in state } i}$$ ← time spent in a particular state

$$\hat{b}_i(k) = \frac{\sum_{t=1}^{T} \gamma_t(i)\,[O_t = k]}{\sum_t \gamma_t(i)}$$ ← total expected time to stay in i

$$\varepsilon_t(i,j) = P(q_t = i, q_{t+1} = j \mid \vec{O}, \lambda)$$

from later $$= \frac{\alpha_t(i)\, a_{ji}\, b_i(O_{t+1})\, \beta_{t+1}(j)}{P(\vec{O} \mid \lambda)}$$

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

what state it was in at the begin. of time

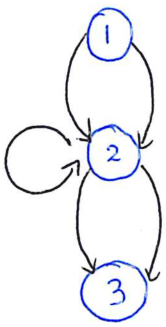
$\hat{\pi}_i = \gamma_1(i)$

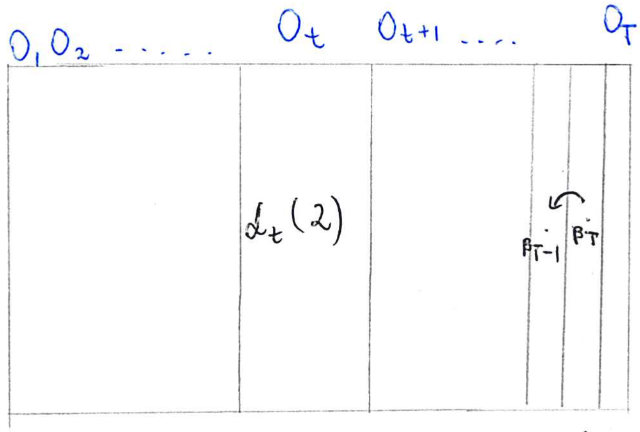
In state j transition to i:

$$\hat{a}_{ij} = \frac{\#\ \text{transitions } j \mapsto i}{\#\ \text{transition } j \to \text{something}} = \frac{\sum_{t=1}^{T-1} \xi_t(j,i)}{\sum_{t=1}^{T-1} \cancel{\sum_i \xi_t(j,i)}\ \gamma_t(j)}$$

$$\alpha_t(i) = P(O_1, \dots, O_t, q_t = i \mid \lambda)$$

output string





$\alpha_t(2)$, what is the prob., that if you have string $O_1 \dots O_t$ it ended in 2.

$$\beta_t(i) = P(O_{t+1}, \dots, O_T \mid q_t = i)$$

$\beta_T(i) = 1$ ← state after last entry

$$\beta_{t-1}(i) = b_i(O_t) \sum_j \beta_t(j)\, a_{jt}$$

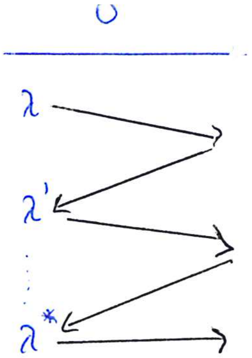
"the right output"

$$\alpha_t(i)\, \beta_i(t) = P(\vec{O}, \vec{q_t} = i \mid \lambda)$$

$$\sum_i \alpha_t(i)\, \beta_t(i) = P(\vec{O} \mid \lambda)$$ ← many different ways to solve the problem

$$\gamma_t(i) = P(q_t = i \mid \vec{O}, \lambda) = \frac{P(q_t = i, \vec{O} \mid \lambda)}{P(\vec{O}, \lambda)} = \frac{\alpha_t(i)\, \beta_t(i)}{P(\vec{O} \mid \lambda)}$$

$\hat{q}$

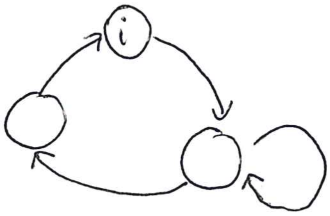


Related to speech recognision.

HMM - supervise learning

- can measure MM for heart

I/O HMM



Unsupervised learning

$$b_i(k) = P(O_t = k \mid q_t = i)$$

Reg. HMM = vanila HMM

$$b_i(k,l) = P(O_t = k \mid c_t = k, q_k = i)$$

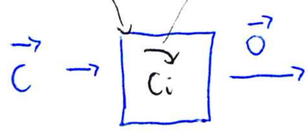
I/O HMM

Input seq & output seq.

I/O HMM - supervised + unsupervised

a whole machine

number of transitions



→ discrete output → continous outputs

- Gaussian HMM
- Mixture of Gaussian HMM

Eg. security system on a computer

- gets a seq. of normal calls

↓

produce HMM
(server usage)

↓ if unusuall responde to it

All of that for discrete (time, symbol)