So far in Class: (2012-11-16)

- Viterbi Alg. $\arg\max_{\bar{S}} P(\bar{S} \mid \bar{O})$
  - Inference

- Forward Alg. $P(\bar{O}) = \sum_{\{\bar{S}\}} P(\bar{O}, \bar{S})$
  -

In order to get to Learning, we also need the backward Formulation.

- Backward Alg. $P(\bar{O} \mid S_1 = i)$

- Posterior Decoding: $P(S_t = i \mid \bar{O})$

$$= \frac{P(\bar{O}_{1:t}, S_t = i) P(\bar{O}_{t+1:T} \mid S_t = i)}{P(\bar{O}_{1:T})}$$

Learning: Baum-Welch (EM)

Restate:

$\pi_i = P(S_1 = i)$ — n possible states

$a_{ij} = P(S_{t+1} = j \mid S_t = i)$

$b_{ik} = P(O_t = k \mid S_t = i)$ — $[b_i(k) = b_{ik}]$

Forward: $\alpha_{it} \equiv P(\overline{O}_{1:t}, S_t = i)$

$$\alpha_{j(t+1)} = \sum_{i=1}^{n} \alpha_{it}\, a_{ij}\, b_j(O_{t+1})$$

$$\alpha_{i1} = \pi_i\, b_i(O_1) \quad \text{(Basecase)}$$

Backward:

$$\beta_{it} \equiv P(O_{(t+1):T} \mid S_t = i)$$

① Basecase: $P(\text{nothing} \mid S_T = i) = ?$

$$\beta_{iT} = 1$$

② Backwards:

$$\beta_{it} = P(O_{(t+1):T} \mid S_t = i)$$

$$= \sum_{j=1}^{n} P(O_{(t+1:T)}, S_{t+1} = j \mid S_t = i)$$

$$= \sum_{j=1}^{n} P(O_{(t+1:T)} \mid S_{t+1} = j, S_t = i) P(S_{t+1} = j \mid S_t = i)$$

$$= \sum_{j=1}^{n} P(O_{(t+1:T)} \mid S_{t+1} = j) P(S_{t+1} = j \mid S_t = i)$$

$$= \sum_{j=1}^{n} \underbrace{P(O_{(t+2:T)} \mid S_{t+1} = j)}_{\beta_{j(t+1)}} \underbrace{P(O_{t+1} \mid S_{t+1} = j)}_{b_j(O_{t+1})} \underbrace{P(S_{t+1} = j \mid S_t = i)}_{a_{ij}}$$

So: $P(S_t = i \mid \bar{O})$

$$= \frac{P(\bar{O}_{1:t}, S_t = i)\, P(\bar{O}_{(t+1):T} \mid S_t = i)}{P(\bar{O}_{1:T})} = \frac{P(\bar{O}_{1:T}, S_t = i)}{P(\bar{O}_{1:T})}$$

$$= \frac{\alpha_{it}\, \beta_{it}}{\sum_i \alpha_{iT}}$$

Q: How do you compute

$$P(S_t = i, S_{t+1} = j \mid \bar{O}) \quad ?$$

A:

$$P(S_t = i, S_{t+1} = j \mid \bar{O}) = \frac{P(\bar{O}, S_t = i, S_{t+1} = j)}{P(\bar{O})}$$

$$= \frac{P(\bar{O}_{1:t}, S_t = i)\, P(S_{t+1} = j \mid S_t = i)\, P(O_{t+1} \mid S_{t+1} = j)\, P(\bar{O}_{(t+2):T} \mid S_{t+1} = j)}{P(\bar{O})}$$

Aside:

$$\sum_i \underbrace{\pi_i\, b_i(O_1)}_{\alpha_{i1}} \beta_{i1}$$

generally: $\sum_i \alpha_{i1} \beta_{i1}$