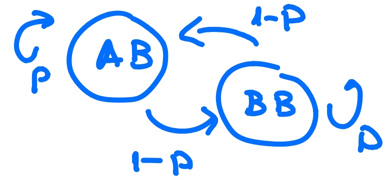
W07 - Probabilistic Models - Expectation Maximization

HMM - parameter $= \underset{\sim}{p}$ $= P(AB|AB)$

$= P(BB|BB)$



$P(X_1 \dots X_L | p) \propto P(p | X_1 \dots X_L)$

We need "labeled" data : $\{\hat{X}_1 \hat{Z}_1, \dots \hat{X}_L \hat{Z}_L\}_{n=1}^{N}$

can calculate

$\hat{c}_i(AB \to AB)$

$\hat{c}_i(BB \to BB)$

$\hat{c}_i(AB \to BB)$

$\hat{c}_i(BB \to AB)$

$$p^* = \frac{\sum_n \sum_i \left[c_i^n(AB \to AB) + c_i^n(BB \to BB) \right]}{\sum_n \sum_i \left[c_i^n(AB \to AB) + c_i^n(BB \to BB) + c_i^n(AB \to BB) + c_i^n(BB \to AB) \right]}$$

Maximum likelihood estimate of the parameters

ML Derivation

$$D = \{ x_1^m z_1^m, \ldots, x_L^m z_L^m \}_{m=1}^{M}.$$

$$P(D \mid p) = \prod_{m=1}^{M} P(x_1^m z_1^m \ldots x_L^m z_L^m \mid p)$$

$$= \prod_{m=1}^{M} P(x_1^m z_1^m) \prod_{i=2}^{L} P(z_i^m \mid z_{i-1}^m) \cdot P(x_i^m \mid z_i^m)$$

$$= \prod_{m=1}^{M} P(x_1^m z_1^m) \cdot \prod_{i=2}^{L} P(x_i^m \mid z_i^m) \cdot p^{c_i^m(AB \to AB) + c_i^m(BB \to BB)} \cdot (1-p)^{c_i^m(AB \to BB) + c_i^m(BB \to AB)}$$

$$\log P(D \mid p) \propto \sum_m \sum_i \left[c_i^m(\text{same}) \log p + c_i^m(\text{break}) \log(1-p) \right]$$

$$\frac{\delta \log P(D \mid p)}{\delta p} = \frac{c(\text{same})}{p} - \frac{c(\text{break})}{1-p} \Big|_{p^*} = 0$$

$$\boxed{p^* = \frac{c(\text{same})}{c(\text{same}) + c(\text{break})}}$$

Expectation Maximization

Incomplete data : $D = \{ x_1^m \dots x_L^m \}_{m=1}^M$

EM: • start w/ $p^{(0)}$

• Expectation:

$$E_i(AB \to AB) = P(x_1 .. x_i\, z_i = AB\, x_{i+1}\, z_{i+1} = AB \dots x_L \mid p^0)$$

$$E_i(BB \to AB) = P(\dots\ z_i = BB \dots z_{i+1} = AB \dots \mid p^{(0)})$$

• Maximization

$$p^{(1)} = \frac{\sum_i \sum_m \left[E_i^m(AB \to AB) + E_i^m(BB \to BB) \right]}{\sum_i \sum_m \left[E_i^m(\text{same}) + E_i^m(\text{break}) \right]}$$

How to calculate

$$E_i(AB \to AB) = P(x_1 \dots x_i\, z_i = AB\, x_{i+1}\, z_{i+1} = AB\, x_{i+2} \dots x_L)$$

$$= P(x_1 \dots x_i\, z_i = AB)\, P(x_{i+2} \dots x_L \mid z_{i+1} = AB)\, P\left(\substack{x_{i+1} \\ z_{i+1} = AB} \mid z_i = AB\right)$$

$$= f_{AB}(i) \cdot p^{(i)} \cdot P(x_{i+1} \mid AB)\, b_{AB}(i+1)$$

$$E_i(AB \to AB) = f_{AB}(i)\, p\, P(x_{i+1} | AB)\, b_{AB}(i+1)$$

$$E_i(BB \to BB) = f_{BB}(i)\, p\, P(x_{i+1} | BB)\, b_{BB}(i+1)$$

$$E_i(AB \to BB) = f_{AB}(i)\,(1-p)\, P(x_{i+1} | BB)\, b_{BB}(i+1)$$

$$E_i(BB \to AB) = f_{BB}(i)\,(1-p)\, P(x_{i+1} | AB)\, b_{AB}(i+1)$$

EM estimation for other conditional probabilities

$$E_i(x | AB) = E_{i-1}(AB \to AB) + E_{i-1}(BB \to AB)$$

$$= P(x_i | AB)\left[f_{AB}(i-1)\, p\, b_{AB}(i) + f_{BB}(i-1)\,(1-p)\, b_{AB}(i) \right]$$

$$E_i(x | BB) = E_{i-1}(BB \to BB) + E_{i-1}(AB \to BB)$$

$$= P(x_i | BB)\left[f_{BB}(i-1) \cdot p\, b_{BB}(i) + f_{AB}(i-1)(1-p)\, b_{BB}(i) \right]$$

Why does EM work?

EM as a variational method.

$$D = \{x_1^m \ldots x_L^m\}_{m=1}^M$$

Unobserved data:

$$Z = \{z_1^m \ldots z_L^m\}_{m=1}^M.$$

$$P(D \mid p) = \sum_z P(D, z \mid p)$$

In a variational method, you introduce an additional and arbitrary probability distribution on the new variable $Q(z)$

and use it as

$$P(D \mid p) = \sum_z Q(z) \frac{P(D, z \mid p)}{Q(z)}$$

$\sum_z Q(z) = 1$.

The variant distribution Q, could also be conditioned on D or on a completely different set of parameters θ but not the parameter p.

$$\log P(D|p) = \log \sum_z Q(z) \frac{P(Dz|p)}{Q(z)}$$

Using Jensen's inequality

$$\log P(D|p) \geq \sum_z Q(z) \log \frac{P(Dz|p)}{Q(z)}$$

To describe the EM algorithm,

we select $Q_n(z|D) = P(z \mid D, p^{(n)})$

introduce

$$G_n(D|p) = \sum_z P(z|D\, p^{(n)}) \log \frac{P(D, z|p)}{P(z|D\, p^{(n)})}$$

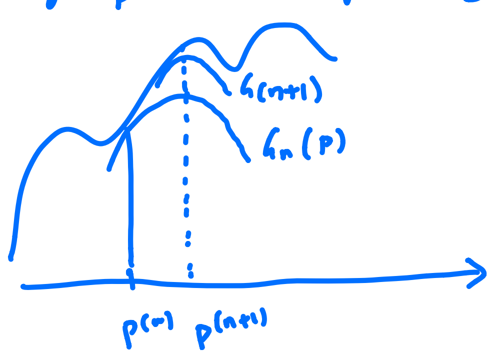
then $\log P(D|p) \geq G_n(D, p)$

- the 2 functions take same value at $p = p^{(n)}$

$$G_n(D, p = p^{(n)}) = \log P(D|p^{(n)})$$

- optimizing $G_n(D, p)$ is equivalent to optimizing.

$$\operatorname{argmax}_p G_n(D, p) = \operatorname{argmax}_p \sum_z P(z|D\, p^{(n)}) \log P(Dz|p)$$



At each iteration in the EM algorithm

E-step calculate $P(z \mid D\, p^{(n)})$ for all possible realizations of the unobservable variables

in our particular case

$$E_i^m(AB \to AB \mid p^{(n)})$$
$$E_i^m(BB \to BB \mid p^{(n)})$$
$$E_i^m(AB \to BB \mid p^{(n)})$$
$$E_i^m(BB \to AB \mid p^{(n)})$$

M-step optimize p in

$$\sum_z P(z \mid D p^{(n)}) \log P(D z \mid p)$$

see "what's the expectation maximization algorithm"
Do & Batzoglou

EM guarantees a local maxima.