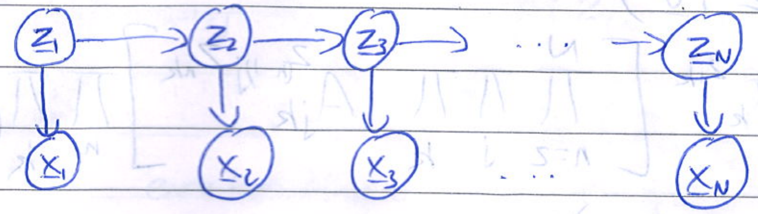
HMMs



$$p(\underline{x}_1, \dots \underline{x}_N, \underline{z}_1, \dots \underline{z}_N) \qquad [p(X, Z)]$$

$$= p(\underline{z}_1) \cdot \left[\prod_{n=2}^{N} p(\underline{z}_n | \underline{z}_{n-1})\right] \prod_{n=1}^{N} p(\underline{x}_n | \underline{z}_n) \qquad -(1)$$

Parameters: $\underline{\pi}$ (initial state), A (transitions), $\underline{\Phi}$ (emissions) } $\underline{\theta}$ (all parameters combined)

EM algorithm

Initial parameter settings $\underline{\theta}^{old}$

E step

Evaluate $p(Z | X, \underline{\theta}^{old})$ posterior on latent var.

Define Q, expectation of complete-data log likelihood with respect to posterior:

$$Q(\underline{\theta}, \underline{\theta}^{old}) = \sum_{Z} p(Z | X, \underline{\theta}^{old}) \ln p(X, Z | \underline{\theta}).$$

In (1), we have:

$$p(\underline{z}_1) = \prod_{k=1}^{K} \pi_k^{z_{1k}}$$

$$p(\underline{z}_n | \underline{z}_{n-1}) = \prod_{j=1}^{K} \prod_{k=1}^{K} A_{jk}^{z_{(n-1)j} z_{nk}}$$

$$p(\underline{x}_n | \underline{z}_n) = \prod_{k=1}^{K} p(\underline{x}_n | \underline{\Phi}_k)^{z_{nk}}$$

$\Rightarrow p(X, Z | \underline{\theta})$

$$= \prod_k \pi_k^{z_{1k}} \left[\prod_{n=2}^{N} \prod_j \prod_k A_{jk}^{z_{(n-1)j} z_{nk}} \right] \prod_n \prod_k p(x_n | \phi_k)^{z_{nk}}$$

$\log p(X, Z | \underline{\theta})$

$$= \sum_k z_{1k} \ln \pi_k + \sum_{n=2}^{N} \sum_j \sum_k (z_{(n-1)j} z_{nk}) \ln A_{jk}$$

$$+ \sum_n \sum_k z_{nk} \ln p(\underline{x}_n | \underline{\phi}_k)$$

∴ The terms in Z that appear are individual and pairwise (for successive data points)

If we take the expectation:

$$Q(\underline{\theta}, \underline{\theta}^{old}) = \sum_Z p(Z | X, \underline{\theta}^{old}) \ln p(X, Z | \underline{\theta})$$

$$= \sum_k E[z_{1k}] \ln \pi_k + \sum_{n=2}^{N} \sum_j \sum_k E[z_{(n-1)j} z_{nk}]$$

$$\ln A_{jk} + \sum_n \sum_k E[z_{nk}] \ln p(\underline{x}_n | \underline{\phi}_k)$$

Where $E[\,]$ denotes expectation with respect to the posterior

Now, $E[z_{nk}] = 0.\, p(z_{nk}=0) + 1.\, p(z_{nk}=1)$

(given $X, \underline{\theta}^{old}$) $= p(z_{nk}=1)$

[Define] $\triangleq \gamma(z_{nk})$

Similarly: Joint probability

$$E[z_{(n-1)j} z_{nk}] = p(z_{(n-1)j}=1, z_{nk}=1)$$

$$\triangleq \xi(z_{(n-1)j}, z_{nk})$$

So evaluating γ and ξ values will give us $Q(\underline{\theta}, \underline{\theta}^{old})$ from the E step

<u>Forward - Backward Algorithm</u>

[a.k.a. Baum - Welch Algorithm]

Via Bayes' theorem

$$\gamma(z_{nk}) = p(z_{nk}=1 \mid X, \underline{\theta}^{old})$$

we will not write this term explicitly; fixed in E-step

$$= \frac{p(X \mid z_{nk}=1)\, p(z_{nk}=1)}{p(X)}$$

Now, given $\underline{z}_n$, $\underline{x}_1 \ldots \underline{x}_n$ are <u>conditionally independent</u> of future data points

$$\Rightarrow \gamma(z_{nk}) = \frac{p(\underline{x}_1, \ldots \underline{x}_n \mid z_{nk}=1)\, p(\underline{x}_{n+1}, \ldots \underline{x}_N \mid z_{nk}=1)\, p(z_{nk}=1)}{p(X)}$$

$\boxed{\text{Define } \alpha, \beta}$

$$= \frac{\overbrace{p(\underline{x}_1, \ldots \underline{x}_n, z_{nk}=1)}^{\triangleq \alpha(z_{nk})}\ \overbrace{p(\underline{x}_{n+1}, \ldots \underline{x}_N \mid z_{nk}=1)}^{\triangleq \beta(z_{nk})}}{p(X)}$$

$$= \frac{\alpha(z_{nk})\, \beta(z_{nk})}{p(X)} \quad (2)$$

So how to evaluate α, β?

$$\alpha(z_{nk}) = p(\underline{x}_1, \dots \underline{x}_n \mid z_{nk}=1) \cdot p(z_{nk}=1)$$

$$= p(\underline{x}_n \mid z_{nk}=1)\, p(\underline{x}_1, \dots \underline{x}_{n-1} \mid z_{nk}=1) \cdot p(z_{nk}=1)$$

$$= p(\underline{x}_n \mid z_{nk}=1) \cdot \sum_{k'} p(\underline{x}_1, \dots \underline{x}_{n-1}, z_{(n-1)k'}=1, z_{nk}=1)$$

$$= p(\underline{x}_n \mid z_{nk}=1) \cdot \sum_{k'} p(\underline{x}_1, \dots \underline{x}_{n-1}, z_{nk}=1 \mid z_{(n-1)k'}=1) \cdot p(z_{(n-1)k'}=1)$$

$$= p(\underline{x}_n \mid z_{nk}=1) \cdot \sum_{k'} p(\underline{x}_1, \dots \underline{x}_{n-1} \mid z_{(n-1)k'}=1)\, p(z_{nk}=1 \mid z_{(n-1)k'}=1)\, p(z_{(n-1)k'}=1)$$

$$= p(\underline{x}_n \mid z_{nk}=1) \sum_{k'} p(\underline{x}_1, \dots \underline{x}_{n-1}, z_{(n-1)k'}=1) \cdot p(z_{nk}=1 \mid z_{(n-1)k'}=1)$$

$$= \underbrace{p(\underline{x}_n \mid z_{nk}=1)}_{\substack{\text{emission} \\ = p(\underline{x}_n \mid \Phi_k)}} \sum_{k'} \underbrace{\alpha(z_{(n-1)k'})}_{\substack{\text{previous} \\ \alpha}} \underbrace{p(z_{nk}=1 \mid z_{(n-1)k'}=1)}_{\substack{\text{transition} \\ = A_{k'k}}}$$

So this gives a recursive definition for $\alpha(z_{nk})$

(<u>Forward</u> message-passing)

Starting condition:

$$\alpha(z_{1k}) = p(\underline{x}_1, z_{1k}=1)$$
$$= p(z_{1k}=1) \cdot p(\underline{x}_1 | z_{1k}=1)$$
$$= \pi_k \cdot p(\underline{x}_1 | \underline{\phi}_k)$$

For $\beta(z_{nk})$

$$= p(\underline{x}_{n+1}, \ldots \underline{x}_N | z_{nk}=1)$$

$$= \sum_{k'} p(\underline{x}_{n+1}, \ldots \underline{x}_N, z_{(n+1)k'}=1 | z_{nk}=1)$$

$$= \sum_{k'} p(\underline{x}_{n+1}, \ldots \underline{x}_N | z_{(n+1)k'}=1, z_{nk}=1) \cdot p(z_{(n+1)k'}=1 | z_{nk}=1)$$

(cond. indep.)

$$= \sum_{k'} p(\underline{x}_{n+1} | z_{(n+1)k'}=1) \cdot p(\underline{x}_{n+2}, \ldots \underline{x}_N | z_{(n+1)k'}=1) \cdot p(z_{(n+1)k'}=1 | z_{nk}=1)$$

$$\Rightarrow \beta(z_{nk}) = \sum_{k'} p(\underline{x}_{n+1} | z_{(n+1)k'}=1) \cdot \beta(z_{(n+1)k'}) \cdot p(z_{(n+1)k'}=1 | z_{nk}=1)$$

$$= \sum_{k'} \underbrace{p(\underline{x}_{n+1} | \underline{\phi}_{k'})}_{\text{emission}} \cdot \underbrace{\beta(z_{(n+1)k'})}_{\text{next } \beta} \cdot \underbrace{A_{kk'}}_{\text{transition}}$$

So we have a <u>backward</u> recursion for evaluating β values (Backward message-passing)

Starting condition:

By defn. in Eq. 2,

$$\gamma(z_{Nk}) = \frac{\alpha(z_{Nk})\,\beta(z_{Nk})}{p(X)}$$

$$= \frac{p(\underline{x}_1, \ldots \underline{x}_N, z_{Nk}=1)\cdot\beta(z_{Nk})}{p(X)}$$

$$= \frac{p(X, z_{Nk}=1)\cdot\beta(z_{Nk})}{p(X)}$$

But $\gamma(z_{Nk}) = p(z_{Nk}=1 \mid X)$

$$\Rightarrow \beta(z_{Nk}) = 1 \quad \forall k$$

For the ξ values:

$\xi(z_{(n-1)j}, z_{nk})$

[Bayes' Thm.]

$$= \frac{p(X \mid z_{(n-1)j}=1, z_{nk}=1)\,p(z_{(n-1)j}=1, z_{nk}=1)}{p(X)}$$

$$= \frac{p(\underline{x}_1, \ldots \underline{x}_{n-1} \mid z_{(n-1)j}=1)\,p(\underline{x}_n \mid z_{nk}=1)\,p(\underline{x}_{n+1}, \ldots \underline{x}_N \mid z_{nk}=1)\cdot p(z_{nk}=1 \mid z_{(n-1)j}=1)\cdot p(z_{(n-1)j}=1)}{p(X)}$$

$$= \frac{\alpha(z_{(n-1)j}) \cdot p(x_n \mid z_{nk}=1) \cdot \beta(z_{nk}) \cdot p(z_{nk}=1 \mid z_{(n-1)j}=1)}{p(X)}$$

$$\Rightarrow \xi(z_{(n-1)j}, z_{nk}) = \frac{\alpha(z_{(n-1)j}) \cdot p(\underline{x}_n \mid \Phi_k) \cdot \beta(z_{nk}) \cdot A_{jk}}{p(X)}$$

So computing α, β values also gives us the joint posteriors ξ

Also note that $p(X)$ — the likelihood of the data — can be obtained as:

$$p(X) = \sum_k p(X, z_{Nk}=1)$$

$$= \sum_k \alpha(z_{Nk})$$

This completes the E-step.

<u>M-step</u>

Find $\underline{\theta}$ that maximises $Q(\underline{\theta}, \underline{\theta}^{old})$; this can be done as usual to give us updated values of $\underline{\pi}, A, \underline{\Phi}$.

<u>Forecasting:</u>

$$p(\underline{x}_{N+1} \mid X)$$

$$= \frac{1}{p(X)} \sum_k p(\underline{x}_{N+1} \mid \Phi_k) \sum_{k'} A_{k'k} \, \alpha(z_{Nk'})$$

$$= \sum_k p(\underline{x}_{N+1} \mid \Phi_k) \sum_{k'} A_{k'k} \, \gamma(z_{Nk'})$$

- HMMs for discrimination / classification
  $\rightarrow p(X)$ from E-step