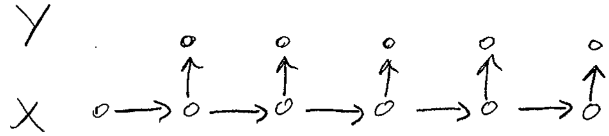
Application - Hidden Markov Models (HMM)



X - Unobserved Markov Chain

Y - Observed Signal

X - as in previous examples

$$P\{X_0=0\} = P\{X_0=1\} = 1/2 \qquad p(x_n|x_{n-1}) = \begin{cases} \rho & x_n \neq x_{n-1} \\ 1-\rho & x_n = x_{n-1} \end{cases}$$

$$p_y(y_n \mid x_n) = \begin{cases} N(0, \sigma_1^2) & x_n = 0 \\ N(0, \sigma_2^2) & x_n = 1 \end{cases}$$

$$p(x_n|x_{n-1}) = \delta(x_n, x_{n-1})(1-\rho) + (1-\delta(x_n, x_{n-1}))\rho$$

$$\log p_{x,y}(x,y) =$$

$$\sum_{i=1}^{N}\left\{ -\frac{1}{2\sigma_{x_i}^2} y_i^2 - \frac{1}{2}\log(\sigma_{x_i}^2) + \log\left[\delta(x_i, x_{i-1})(1-\rho) + (1-\delta(x_i, x_{i-1}))\rho\right]\right\}$$

$$+ C$$

Define: $t(x_i, x_{i-1}) = 1 - \delta(x_i, x_{i-1})$

Notice

$$\log\left[\delta(x_i, x_{i-1})(1-\rho) + (1-\delta(x_i, x_{i-1}))\rho\right]$$

$$= t(x_i, x_{i-1}) \log\left(\frac{\rho}{1-\rho}\right) + \log(1-\rho)$$

Defn

$$\beta \triangleq \log\left(\frac{1-\rho}{\rho}\right) = -\log\left(\frac{\rho}{1-\rho}\right)$$

$$l(y|k) \triangleq \frac{1}{2\sigma_k^2} y^2 + \frac{1}{2}\log(\sigma_k^2)$$

$$\log p_{x,y}(x,y) = \sum_{i=1}^{N} \left\{ -l(y_i | x_i) - t(x_i, x_{i-1})\beta \right\} + C''$$

likelihood of data

transition probability

$$\hat{X}_{MAP} = \arg\max_{X} \left\{ \sum_{i=1}^{N} -l(y_i | x_i) - t(x_i, x_{i-1})\beta \right\}$$

$$\hat{X}_{ML} = \arg\max_{X} \left\{ \sum_{i=1}^{N} -l(y_i | x_i) \right\}$$

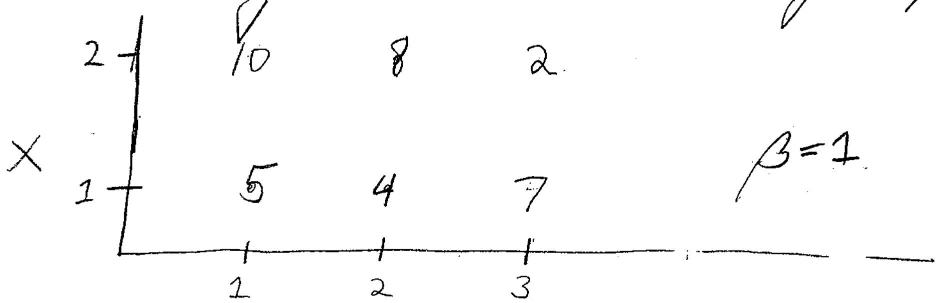
$$(\hat{X}_{ML})_i = \arg\max_{x_i} \left\{ -l(y_i | x_i) \right\}$$

Lecture 14

$$\hat{X}_{MAP} = \underset{X}{\arg\max}\left\{ \sum_{i=1}^{N} \underbrace{-l(y_i|x_i)}_{\text{data fitting term}} \underbrace{-t(x_i,x_{i-1})\beta}_{\text{prior model term}} \right\}$$

Notice

1) Choosing X is like choosing a "path"



2) $l(y_i|x_i)$ is the "cost" of each node

3) $t(x_i,x_{i-1})\beta$ is the "cost" of each transition

A) Dynamic Programming

Defn

$L(K,n)$ - Cost of best path starting at n with value K.

$L(1,3) = 7 \quad 2$

$L(2,3) = 2$

$$L(K,n) = \underset{l}{\operatorname{argmin}}\{L(l,n+1) + t(K,l)\} + l(y_n|K)$$

$$L(1,2) = \min\{L(1,3),\ L(2,3)+1\} + 4$$
$$= \min\{7,\ 3\} + 4$$
$$= 7 \quad (2)$$

$$L(2,2) = \min\{L(1,3)+1,\ L(2,3)\} + 8$$
$$= \min\{8,2\} + 8$$
$$= 10 \quad (2)$$

$$L(1,1) = \min\{L(1,2),\ L(2,2)+1\} + 5$$
$$= \min\{7,\ 11\} + 5$$
$$= 12 \quad (1)$$

$$L(2,1) = \min\{L(1,2)+1,\ L(2,2)\} + 10$$
$$= \min\{8,\ 10\} + 10$$
$$= 18 \quad (1)$$

$L(1,1) < L(2,1) \Rightarrow$ Best path $1, 1, 2$

B. Coordinate decent, Gauss-Seidel / ICM

$$\hat{X}_n = \arg\min_{x_n} \left\{ \sum_{i=1}^{N} l(y_i|x_i) + t(x_i, x_{i-1})\beta \right\}$$

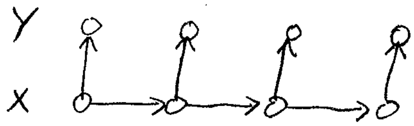
$$= \arg\min_{x_n} \left\{ l(y_n|x_n) + \beta\left(t(x_n, x_{n-1}) + t(x_{n+1}, x_n) \right) \right\}$$

- Iterative application converges to local minimum
- Does not generally converge to global minimum

C. Gradient decent $\Rightarrow$ doesn't work!

X discrete!

Extension



$$P\{X_0 = i\} = \pi_i \qquad i = 1, \ldots, M$$

$$p(y_s | x_s) \sim N(\mu_{x_s}, \sigma^2_{x_s})$$

$$p(y|x) = \prod_{s=1}^{N} p(y_s | x_s)$$

$$P_{ij} = P\{X_s = j \,|\, X_{s-1} = i\}$$

Each element is a vector

$$\theta = [\pi, \underset{\text{matrix}}{P}, \mu, \gamma]$$

$$\hat{\theta} = \underset{\theta \in \Omega}{\operatorname{argmax}}\ P\{Y = y | \theta\}$$

$$\Omega = \left\{ \theta : \sum_{i=1}^{M} \pi_i = 1,\ \sum_{j=1}^{M} P_{ij} = 1 \quad \forall i \right\}$$

$\boxed{\text{Lecture 31}}$

$$\log p(y\,x|\theta) = \log p(y|x,\theta) + \log p(x|\theta)$$

$$Q(\theta',\theta) = E\left[\log p(Y,X|\theta) \,|\, Y=y,\theta\right]$$

$$= E\left[\log p(Y|X\theta') | Y=y,\theta\right] \quad \text{1st term}$$

$$+ E\left[\log p(X|\theta') \,|\, Y=y,\theta\right] \quad \text{2nd term}$$

1st term same as previous exampl

$$= \sum_{j=1}^{M} \sum_{s=1}^{N} \left\{ \frac{1}{2\gamma_j'}(y_s - \mu_j')^2 + \frac{1}{2}\log(2\pi\gamma_j')\right\} P\{X_s=j|Y=y,\theta\}$$

$$f_s(j|\theta) \triangleq P\{X_s=j\,|\,Y=y,\theta\}$$

2nd term = ?

$$\log p(X|\theta') = \log\left\{ \pi_{X_1}' \prod_{s=1}^{N-1} P_{X_s X_{s+1}}' \right\}$$

Define

$$\tau_i = \delta(X_1 - i)$$
$$= \#\text{ of } X_1\text{'s} = i$$
$$K_{ij} = \#\text{ of } (X_s=i, X_{s+1}=j) \text{ pairs}$$

$$\log p(X|\theta') = \log\left(\left(\prod_{i=1}^{M} \pi_i'^{\tau_i}\right)\left(\prod_{\substack{i=1\\j=1}}^{M} P_{ij}'^{K_{i,j}}\right)\right)$$

$$= \sum_{i=1}^{M} \tau_i \log \pi_i' + \sum_{i=1}^{M}\sum_{j=1}^{M} K_{ij} \log P_{ij}'$$

$$E[\log p(X|\theta) \,|\, Y=y, \theta]$$

$$= \sum_{i=1}^{M} \bar{\tau}_i \log \pi_i' + \sum_{i=1}^{M}\sum_{j=1}^{M} \bar{K}_{ij} \log P_{ij}'$$

$$\bar{\tau}_i \triangleq P\{X_1 = i \,|\, Y=y, \theta\} = f_1(i|\theta)$$

$$\bar{K}_{i,j} \triangleq E[K(i,j) \,|\, Y=y, \theta]$$

$$= \sum_{s=1}^{N-1} \underbrace{P\{X_s = i, X_{s+1} = j \,|\, Y=y, \theta\}}_{g_s(i,j|\theta)}$$

M-step

$Q(\theta', \theta) =$

$$\sum_{s=1}^{N} \sum_{j=1}^{M} \left\{ \frac{1}{2\gamma_j'} (y_s - \mu_j')^2 + \frac{1}{2} \log(2\pi \gamma_j') \right\} f_s(j|\theta)$$

$$+ \sum_{i=1}^{M} \bar{\tau}_i \log \pi_i' \quad + \sum_{i,j=1}^{M} \bar{R}_{i,j} \log P_{ij}'$$

maximize with respect to π, P, μ, γ

1) From previous example

for $j = 1, \ldots, M$

$$\hat{N}_j = \sum_{s=1}^{N} f_s(j|\theta)$$

$$\hat{\mu}_j = \frac{1}{\hat{N}_j} \sum_{s=1}^{N} y_s f_s(j|\theta)$$

$$\hat{\gamma}_j = \frac{1}{\hat{N}_j} \sum_{s=1}^{N} (y_s - \hat{\mu}_j)^2 f_s(j|\theta)$$

2) maximize P_{ij} subject to $\sum_{j=1}^{M} P_{ij} = 1$ constraint

Use Lagrange multipliers λ_i

$$\frac{\partial}{\partial P_{m,n}} \left\{ \sum_{i,j} \overline{K}_{i,j} \log P_{ij} + \sum_{i=1}^{M} \lambda_i \sum_{j=1}^{M} P_{ij} \right\} \Bigg|_{P=\hat{P}} = 0$$

$$\frac{\overline{K}_{m,n}}{\hat{P}_{m,n}} + \lambda_m = 0$$

$$\hat{P}_{m,n} = \frac{\overline{K}_{m,n}}{-\lambda_m}$$

$$\sum_{n=1}^{N} \hat{P}_{mn} = 1 \Rightarrow \boxed{\hat{P}_{mn} = \frac{\overline{K}_{m,n}}{\sum_{n=1}^{M} \overline{K}_{m,n}}}$$

3) maximize π_i subject to $\sum_{i=1}^{M} \pi_i = 1$

$$\boxed{\hat{\pi}_i = \frac{\overline{\tau}_i}{\sum_{i=1}^{M} \overline{\tau}_i}}$$

Lecture 32

E step

(Details in HMM paper by Rabiner + Juang)

How do we compute

$$f_s(i\,|\,\theta),\ \hat{R}(i,j),\ \hat{\tau}(i)$$

note $\hat{\tau}(i) = f_1(i\,|\,\theta)$

First we must compute the following constants

$$\alpha_s(i) = P\{ Y_r = y_r\ r \le s,\ X_s = i \mid \theta \}$$

$$\beta_s(i) = P\{ Y_r = y_r\ r > s \mid X_s = i, \theta \}$$

Exercise: show that $\alpha_s(i)$ and $\beta_s(i)$ can be computed using the following recursions

Define $p(y_s | j) \triangleq p(y_s | x_s = j) \sim N(\mu_j, \sigma_j^2)$

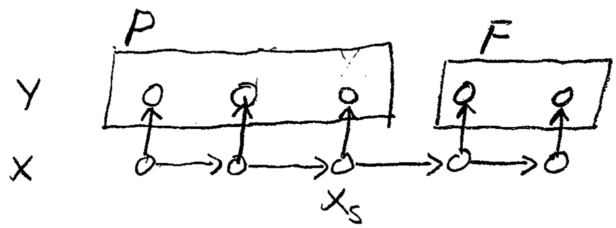
$$\boxed{\begin{aligned} \alpha_1(i) &= \pi_i\, p(y_1 | i) \\ \alpha_{s+1}(i) &= \left[\sum_{i=1}^{M} \alpha_s(i) P_{ij} \right] p(y_{s+1} | j) \end{aligned}}$$

forward recursion

$$\boxed{\begin{aligned} \beta_N(i) &= 1 \\ \beta_s(i) &= \sum_{j=1}^{M} P_{ij}\, p(y_{s+1} \mid j)\, \beta_{s+1}(j) \end{aligned}}$$

backward recursion

<u>Intuition</u>



P - past + present observations
F - future observations

$$\alpha_s(i) = P\{P\ ,\ x_s = i \mid \theta\}$$

$$\beta_s(i) = P\{F \mid x_s = i, \theta\}$$

$$f_s(i \mid \theta) = p(x_s = i \mid y, \theta)$$

$$= \frac{p(x_s, F, P \mid \theta)}{p(F, P \mid \theta)} \quad = p(y \mid \theta) \leftarrow \text{constant}$$

$$\boxed{f_s(i \mid \theta) = \frac{\beta_s(i)\,\alpha_s(i)}{\sum_{i=1}^{M} \beta_s(i)\,\alpha_s(i)}}$$

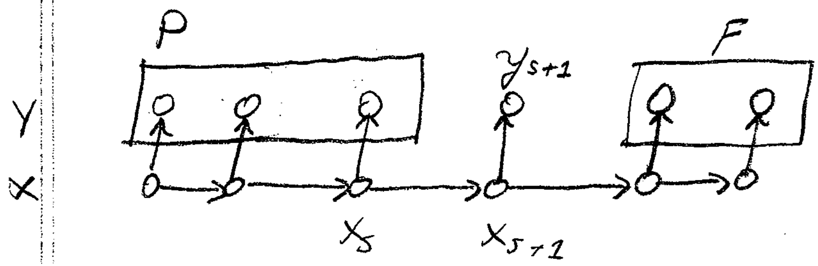
notice: $p(y \mid \theta) = \sum_{i=1}^{M} \beta_s(i)\,\alpha_s(i)$

$$\hat{R}(i,j) = \sum_{s=1}^{N} P\{X_s = i,\ X_{s+1} = j \mid Y = y,\ \theta\}$$

↑ drop θ for convenience

$$= \sum_{s=1}^{N} \frac{P\{X_s = i,\ X_{s+1} = j,\ Y = y\}}{p(y) \leftarrow \text{constant}}$$

$$= \sum_{s=1}^{N} \frac{P\{F \mid X_{s+1} = j\}\, p(y_{s+1} \mid j)\, P_{X_s X_{s+1}}\, P\{X_s = i, P\}}{p(y)}$$



$$= \boxed{\sum_{s=1}^{N-1} \frac{\beta_{s+1}(j)\, p(y_{s+1} \mid j)\, P_{ij}\, \alpha_s(i)}{p(y)} = \hat{R}(i,j)}$$

Exercise: Write all update equations together

Summary of EM for HMM

M step

$$N_j^{(k+1)} = \sum_{s=1}^{N} f_s(j \mid \theta^{(k)})$$

$$\mu_j^{(k+1)} = \frac{1}{N_j^{(k+1)}} \sum_{s=1}^{N} y_s \, f_s(j \mid \theta^{(k)})$$

$$\gamma_j^{(k+1)} = \frac{1}{N_j^{(k+1)}} \sum_{s=1}^{N} (y_s - \mu_j^{(k+1)}) \, f_s(j \mid \theta^{(k)})$$

$$\pi_j^{(k+1)} = f_0(j \mid \theta^{(k)})$$

$$P_{ij}^{(k+1)} = \frac{g(i,j \mid \theta^{(k)})}{\sum_{l=1}^{M} g(i,l \mid \theta^{(k)})}$$

E step

$$\alpha_1^{(k)}(i) = p(y_1 \mid i, \theta^{(k)}) \, \pi_i^{(k)}$$

$$\alpha_{s+1}^{(k)} = \left[\sum_{i=1}^{M} \alpha_s^{(k)}(i) \, P_{ij}^{(k)} \right] p(y_{s+1} \mid j, \theta^{(k)})$$

$$\beta_N^{(k)}(i) = 1$$

$$\beta_{s-1}^{(k)}(i) = \sum_{j=1}^{M} P_{ij}^{(k)} \, p(y_s \mid j, \theta^{(k)}) \, \beta_s^{(k)}(j)$$

$$\gamma_s(i \mid \theta^{(k)}) = \frac{\beta_s^{(k)}(i)\ \alpha_s^{(k)}(i)}{\sum_{\ell=1}^{M} \beta_s^{(k)}(\ell)\ \alpha_s^{(k)}(\ell)}$$

$$g(i,j \mid \theta^{(k)}) = \sum_{s=1}^{N-1} \alpha_s^{(k)}(i)\ P_{ij}^{(k)}\ p(y_{s+1} \mid j; \theta^{(k)})\ \beta_{s+1}^{(k)}(j)$$