e) $P(X_t|X_1,\dots,X_{t-1}) = \sum_{i=1}^{n} P(X_t, S_t=i|X_1,\dots,X_{t-1}) = \sum_i \underbrace{P(X_t|S_t=i)}_{\text{Gaussian}} \cdot \underbrace{P(S_t=i|X_1,\dots,X_{t-1})}_{\text{number}}$

Gaussian mixture with n components

f) $P(X_t) = \sum_{i=1}^{n} \underbrace{P(X_t|S_t=i)}_{\text{Gaussian}} \underbrace{P(S_t=i)}_{\text{number}}$ → Gaussian mixture with n components

g) $\underbrace{P(X_1|S_1,\dots,S_t)\, P(X_2|S_1,\dots,S_t,X_1) \dots P(X_t|S_1,\dots,S_t,X_1,\dots,X_{t-1})}_{\text{d-separation}} = \prod_{i=1}^{t} P(X_i|S_i)$

multivariate Gaussian with (t-dim).

$$\mu = \begin{bmatrix} \mu_{S_1} \\ \mu_{S_2} \\ \vdots \\ \mu_{S_t} \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \sigma_{S_1}^2 & & & 0 \\ & \sigma_{S_2}^2 & & \\ & & \ddots & \\ 0 & & & \sigma_{S_t}^2 \end{bmatrix}$$

→ can be derived by multiplying all univariate Gaussian $P(X_i|S_i)$ in the product

h) $\dfrac{P(X_t, X_{t'}, S_t, S_{t'})}{P(S_t, S_{t'})} \xrightarrow[\text{the case } t'<t \text{ will be symmetric}]{\text{let's assume } t'>t} \dfrac{\cancel{P(S_t)}\cancel{P(S_{t'}|S_t)}P(X_t|S_t)P(X_{t'}|S_{t'})}{\cancel{P(S_t)}\cancel{P(S_{t'}|S_t)}}$

$= P(X_t|S_t)\,P(X_{t'}|S_{t'})$ → 2D Gaussian with $\mu = \begin{bmatrix} \mu_t \\ \mu_{t'} \end{bmatrix}$ and $\Sigma = \begin{bmatrix} \sigma_t^2 & 0 \\ 0 & \sigma_{t'}^2 \end{bmatrix}$

MEHMET KOC
#A53035914

11/21/2013

CSE 250A – Homework 6

4 $\ell_{it}^* = \max_{\{s_1,\ldots,s_{t-1}\}} \log P(o_1, o_2, \ldots, o_t, S_1, \ldots, S_t = i)$

$\ell_{j(t+1)}^* = \max_i \left(\ell_{it}^* + \log a_{ij}\right) + \log b_j(o_{t+1})$

Base case: $\ell_{i1}^* = \log P(o_1, S_1 = i) = \log (S_1 = i) + \log P(o_1 | S_1 = i) = \log \pi_i + \log b_i(o_1)$

At each time step ℓ_t^* is an $n \times 1$ vector $(n = 26)$; $t = 1, \ldots, 152000$

Save $\ell_t^* = \begin{bmatrix} \ell_{1t}^* \\ \ell_{2t}^* \\ \vdots \\ \ell_{26t}^* \end{bmatrix}$ for $\forall t = 1, \ldots T$ in $L = [\ell_1^* \ \ell_2^* \ \ldots \ \ell_T^*]$ ($n \times T$ matrix)

Also, save $\Phi_{t+1}(j) = \arg\max_i \left(\ell_{it}^* + \log a_{ij}\right)$

$\Phi_t = \begin{bmatrix} \Phi_t(1) \\ \vdots \\ \Phi_t(26) \end{bmatrix}$ ($n \times 1$ vector) $\longrightarrow \Phi = [\Phi_1 \ \Phi_2 \ \ldots \ \Phi_T]$ ($n \times (T-1)$ matrix)

Now backtracking:
$s_T^* = \arg\max_i \ell_{iT}^*$, $s_t^* = \Phi_{t+1}(s_{t+1}^*) \longrightarrow$ Sequence found $\rightarrow$ Code & plot on the next page

Lastly, eliminate repetitions and decode the actual sentence.

4 c) $P(X_t | S_{t-1}) = \sum_{i=1}^{n} P(X_t, S_t = i | S_{t-1}) = \sum_i P(X_t | S_t = i, S_{t-1}) P(S_t = i | S_{t-1})$

$= \sum_i \underbrace{P(X_t | S_t = i)}_{\text{Gaussian}} \underbrace{P(S_t = i | S_{t-1})}_{\text{number}} \longrightarrow$ Gaussian mixture with n components

d) $\sum_{(s_1,\ldots,s_t)} P(X_1, \ldots, X_t, S_1 = s_1, \ldots, S_t = s_t) = \sum_{(s_1,\ldots,s_t)} \underbrace{P(X_1, \ldots, X_t | S_1 = s_1 \ldots, S_t = s_t)}_{\text{Gaussian}} \underbrace{P(S_1 = s_1, \ldots, S_t = s_t)}_{\text{number}}$

Gaussian $\rightarrow$ multivariate $\boxed{\prod_{i=1}^{t} P(X_i | S_i)}$

$\downarrow$

Gaussian mixture with n^t components

② $$\frac{P(S_{t-1}=j, S_{t+2}=l, S_t=i, \vec{O})}{P(S_t=i, \vec{O})} = \frac{\sum_{s_{t+1}} P(S_{t+2}=l, S_{t+1}, S_t=i, S_{t-1}=j, \vec{O})}{\underbrace{P(o_1, o_2, \ldots, o_t, S_t=i)}_{\alpha_{it}} \underbrace{P(o_{t+1}, \ldots, o_T | S_t=i)}_{\beta_{it}}}$$

$$\sum_{s_{t+1}} \underbrace{P(S_{t-1}=j, o_1, \ldots, o_{t-1})}_{\alpha_j(t-1)} \underbrace{P(S_t=i|S_{t-1}=j)}_{a_{ji}} P(S_{t+1}|S_t=i) P(S_{t+2}=l|S_{t+1}) P(o_t|S_t=i) P(o_{t+1}|S_{t+1}) P(o_{t+2}|S_{t+2}=l) \cdot \underbrace{P(o_{t+3}, \ldots, o_T | S_{t+2}=l)}_{\beta_{l(t+2)}}$$

$$\frac{\sum_{s_{t+1}} \alpha_j(t-1)\, a_{ji}\, a_{i(s_{t+1})}\, a_{(s_{t+1})l}\, b_i(o_t)\, b_{(s_{t+1})}(o_{t+1})\, b_l(o_{t+2})\, \beta_{l(t+2)}}{\alpha_{it}\beta_{it}}$$

for a particular observation sequence $\{o_1, o_2, \ldots o_T\}$

This expression assumes stationary joint distribution

③ a) $$q_{jt} = P(S_t=j|o_1, \ldots, o_t) = \frac{P(S_t=j, o_1, \ldots, o_t)}{P(o_1, \ldots, o_t)} \quad \text{(Product rule)}$$

$$= \frac{\sum_i P(S_t=j, S_{t-1}=i, o_1, \ldots, o_t)}{\sum_{ij} P(S_t=j, S_{t-1}=i, o_1, \ldots, o_t)} \quad \text{(Sum Rule)}$$

$$= \frac{\sum_i P(S_{t-1}=i, o_1, \ldots, o_{t-1}) P(S_t=j|S_{t-1}=i) P(o_t|S_t=j)}{\sum_{ij} P(S_{t-1}=i, o_1, \ldots, o_{t-1}) P(S_t=j|S_{t-1}=i) P(o_t|S_t=j)} \quad \begin{pmatrix}\text{Product Rule} \\ \text{Conditional Independence}\end{pmatrix}$$

$q_{i(t-1)}\, P(o_1, \ldots, o_{t-1})$ — independent of i,j (they cancel)

$$= \frac{P(o_t|S_t=j) \sum_i P(S_t=j|S_{t-1}=i) P(S_{t-1}=i, o_1, \ldots, o_{t-1})}{\sum_{ij} P(o_t|S_t=j) P(S_t=j|S_{t-1}=i) P(S_{t-1}=i, o_1, \ldots, o_{t-1})} = \frac{b_j(o_t) \sum_i a_{ij}\, q_{i(t-1)}}{\sum_{ij} b_j(o_t)\, a_{ij}\, q_{i(t-1)}}$$

b) $$P(x_t|y_1, \ldots, y_t) = \frac{P(x_t, y_1 \ldots y_t)}{P(y_1, \ldots, y_t)} = \frac{\int P(x_t, x_{t-1}, y_1, \ldots, y_t)\, dx_{t-1}}{\iint P(x_t, x_{t-1}, y_1, \ldots, y_t)\, dx_{t-1}\, dx_t}$$

$$= \frac{\int P(x_{t-1}, y_1, \ldots, y_{t-1}) P(x_t|x_{t-1}) P(y_t|x_t)\, dx_{t-1}}{\iint P(x_{t-1}, y_1, \ldots, y_{t-1}) P(x_t|x_{t-1}) P(y_t|x_t)\, dx_{t-1}\, dx_t} = \frac{P(y_t|x_t) \int P(x_{t-1}, y_1, \ldots, y_{t-1}) P(x_t|x_{t-1})\, dx_{t-1}}{\int P(y_t|x_t)\, dx_t \int P(x_{t-1}, y_1, \ldots, y_{t-1}) P(x_t|x_{t-1})\, dx_{t-1}}$$

$\rightarrow P(x_{t-1}|y_1, \ldots, y_{t-1}) \underbrace{P(y_1, \ldots, y_{t-1})}_{\text{independent of integrals (they cancel)}}$

In the calculations above, the joint distribution $P(x_t, x_{t-1}, y_1, \dots, y_t)$ is marginalized once in the numerator and twice in the denominator. For an arbitrary distribution, the resultant distributions may not be easy to solve. However, if the RVs are Gaussian, then we know its marginalization will also be Gaussian and this makes the computations tractable..

(15) a) $P(y_1|x) = \dfrac{P(x|y_1)P(y_1)}{P(x|y_0)P(y_0)+P(x|y_1)P(y_1)} =$

$$= \frac{\pi_1 (2\pi)^{-d/2}|\Sigma_1|^{-1/2}\exp\left(\frac{-1}{2}(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1)\right)}{\pi_1 (2\pi)^{-d/2}|\Sigma_1|^{-1/2}\exp\left(\frac{-1}{2}(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1)\right) + \underbrace{\pi_0}_{(1-\pi_1)}(2\pi)^{-d/2}|\Sigma_0|^{-1/2}\exp\left(\frac{-1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)\right)}$$

b) $\Sigma_0 = \Sigma_1 = \Sigma$

$$= \frac{\pi_1 \cancel{(2\pi)^{-d/2}|\Sigma|^{-1/2}}\exp\left(\frac{-1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\right)}{\cancel{(2\pi)^{-d/2}|\Sigma|^{-1/2}}\left(\pi_1\exp\left(\frac{-1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\right)+\pi_0\exp\left(\frac{-1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)\right)\right)}$$

$$= \frac{1}{1+\frac{\pi_0}{\pi_1}\exp\left(\frac{-1}{2}\left(\cancel{x^T\Sigma^{-1}x}-\mu_0^T\Sigma^{-1}x-x^T\Sigma^{-1}\mu_0+\mu_0^T\Sigma^{-1}\mu_0-\cancel{x^T\Sigma^{-1}x}+\mu_1^T\Sigma^{-1}x+x^T\Sigma^{-1}\mu_1-\mu_1^T\Sigma^{-1}\mu_1\right)\right)}$$

$$(\mu_1-\mu_0)^T\Sigma^{-1}x + \boxed{x^T\Sigma^{-1}(\mu_1-\mu_0)} \xrightarrow{\text{scalar}} (\mu_1-\mu_0)^T\Sigma^{-1}x$$

$$= \frac{1}{1+\frac{\pi_0}{\pi_1}\exp\left(\frac{-1}{2}\left(2(\mu_1-\mu_0)^T\Sigma^{-1}x+\mu_0^T\Sigma^{-1}\mu_0-\mu_1^T\Sigma^{-1}\mu_1\right)\right)}$$

$$= \frac{1}{1+\exp\left(-\left[\underbrace{(\mu_1-\mu_0)^T\Sigma^{-1}}_{w^T}x+\underbrace{\frac{1}{2}\mu_0^T\Sigma^{-1}\mu_0-\frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1-\ln(\pi_0/\pi_1)}_{b}\right]\right)}$$

where $w = \Sigma^{-1}(\mu_1-\mu_0)$, $b = \frac{1}{2}\mu_0^T\Sigma^{-1}\mu_0 - \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 - \ln\left(\frac{\pi_0}{\pi_1}\right)$