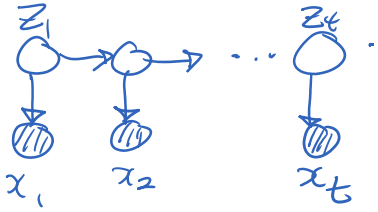
Lecture 15 - scribbles

Tuesday, October 30, 2018 14:35

today: HMM & α-β recursion
ML for HMM (EM)

HMM (hidden Markov model)



$z_t \in \{1, \dots, K\}$ discrete

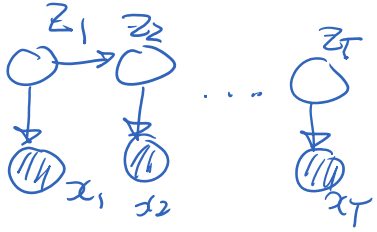
x_t — cts. e.g. speech signal
discrete e.g. DNA sequence

(Often $z_t \sim$ Gaussian in the course → Kalman filter)

HMM → generalization of mixture model

GMM

add dependence on z_t



DGM:

$$p(x_{1:T}, z_{1:T}) = p(z_1) \prod_{t=1}^{T} p(x_t|z_t) \prod_{t=2}^{T} p(z_t|z_{t-1})$$

$p(x_t|z_t)$: emission prob.

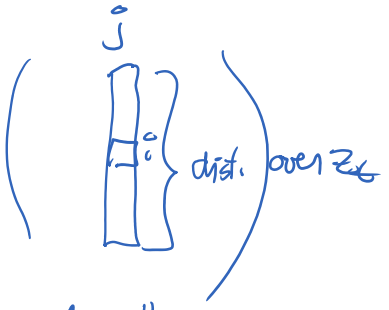
$p(z_t|z_{t-1})$: transition prob.

often, the emission prob & trans. probs are homogeneous i.e. do not depend on t

$$p_t(x_t|z_t) = f(x_t|z_t) \qquad \forall t$$

$p_t(x_t|z_t) = f(x_t|z_t)$

$p_t(z_t = i \mid z_{t-1} = j) = A_{ij}$ — "stochastic matrix"

A 

$\sum_i A_{ij} = 1 \quad \forall j$

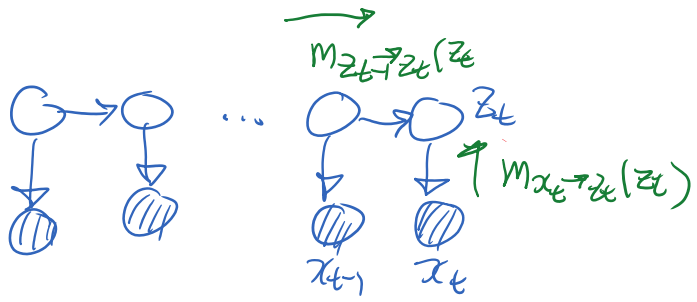
inference tasks:

prediction	$p(z_t \mid x_{1:t-1})$	"where next?"
filtering	$p(z_t \mid x_{1:t})$	"where now?"
smoothing	$p(z_t \mid x_{1:T})$	"where in the past?"

$T > t$

α-recursion:

Let's run sum product here to derive recursions to compute prob.



to compute $p(z_t, \bar{x}_{1:t}) \propto p(z_t \mid \bar{x}_{1:t})$

$p(z_t, \bar{x}_{1:t}) = \frac{1}{Z} 1 \cdot m_{z_{t-1} \to z_t}(z_t)\, m_{x_t \to z_t}(z_t)$ here $Z = 1$

$\alpha_t(z_t)$

$$m_{x_t \to z_t}(z_t) = \sum_{x_t} p(x_t \mid z_t)\, \delta(x_t, \bar{x}_t) = p(\bar{x}_t \mid z_t)$$

$$m_{z_{t-1} \to z_t}(z_t) = \sum_{z_{t-1}} p(z_t \mid z_{t-1})\, m_{z_{t-2} \to z_{t-1}}(z_{t-1})\, m_{x_{t-1} \to z_{t-1}}(z_{t-1})$$

$p(z_{t-1}, \bar{x}_{1:t-1})$

$\alpha_{t-1}(z_{t-1})$

define: $\alpha_t(z_t) \triangleq p(z_t, \bar{x}_{1:t})$

$$\alpha_t(z_t) = \underbrace{p(\bar{x}_t \mid z_t)}_{\text{vector } (z_t)} \sum_{z_{t-1}} \underbrace{p(z_t \mid z_{t-1})}_{\text{matrix}} \underbrace{\alpha_{t-1}(z_{t-1})}_{\text{vector}}$$

α-recursion a.k.a "forward recursion" like the "collect phase" in sum-product alg with z_t as the root

Let $o_t(z_t) \triangleq p(\bar{x}_t \mid z_t)$

$$\alpha_t = o_t \odot A\alpha_{t-1}$$

$\odot$ Hadamard product

initialization: $\alpha_1(z_1) = p(z_1, \bar{x}_1) = p(z_1)\, p(\bar{x}_1 \mid z_1)$

$\tilde{\alpha}_t \triangleq p(z_t \mid \bar{x}_{1:t})$ "filtering distribution"

time complexity:

$O(t \cdot k^2)$

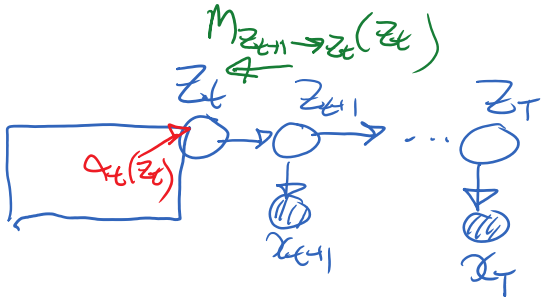
space complexity: $O(k)$ extra storage ($O(k^2)$ for storing A)

Space complexity : $O(K)$ extra storage ($O(K^2)$ for storing A)

$$\sum_{z_t} p(z_t, \bar{x}_{1:t}) = p(\bar{x}_{1:t}) \quad \text{"evidence probability"}$$

$$\alpha_t(z_t)$$

β recursion : smoothing



$$p(z_t, \bar{x}_{1:T}) = \frac{1}{\underbrace{Z}_{(=1)}} \alpha_t(z_t) \cdot \underbrace{m_{z_{t+1} \to z_t}(z_t)}_{\triangleq \beta_t(z_t)}$$

turns out that $\boxed{\beta_t(z_t) \triangleq p(\bar{x}_{t+1:T} | z_t)}$ → why?

$$p(z_t, \bar{x}) = p(\bar{x} | z_t) p(z_t)$$

$$\overset{\text{C.I.}}{=} \underbrace{p(x_{t+1:T} | z_t)}_{\beta_t(z_t)} \underbrace{p(x_{1:t} | z_t) p(z_t)}_{\alpha_t(z_t)} \quad //$$

$$m_{z_{t+1} \to z_t}(z_t) = \sum_{z_{t+1}} p(z_{t+1} | z_t) \, p(\bar{x}_{t+1} | z_{t+1}) \, m_{z_{t+2} \to z_{t+1}}(z_{t+1})$$

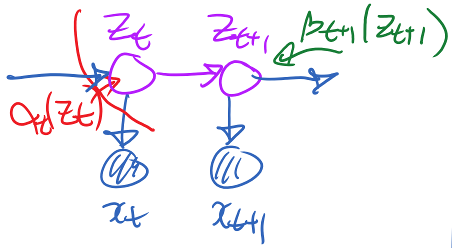
$$\boxed{\beta_t(z_t) = \sum_{z_{t+1}} p(z_{t+1} | z_t) \, p(\bar{x}_{t+1} | z_{t+1}) \, \beta_{t+1}(z_{t+1})}$$

β-recursion (aka backward recursion)

- initialization : $\beta_T(z_T) = 1 \quad \forall z_T$

initialization: $\beta_T(z_T) = 1 \quad \forall z_T$

<u>edge marginal:</u>



$$p(z_t, z_{t+1}, \bar{x}_{1:T}) = \alpha_t(z_t)\,\beta_{t+1}(z_{t+1}) \cdot p(z_{t+1}|z_t) \cdot p(\bar{x}_{t+1}|z_{t+1})$$

15h28

<u>Numerical stability trick:</u>

issue: α_t & β_t can easily go to 1e-100

two possibilities a) (general) store $\log(\alpha_t)$ instead

$$\log\left(\sum_i a_i\right) = \log\left(\tilde{a}\left(\sum_i \frac{a_i}{\tilde{a}}\right)\right)$$

call $\tilde{a} \triangleq \max_i a_i$

$$= \log(\tilde{a}) + \log\left(1 + \sum_{j \neq i_{max}} \exp(\log(a_j) - \log(\tilde{a}))\right)$$

$i_{max} \triangleq \arg\max_i a_i$

b) <u>normalize the messages:</u>

- α-recursion, use $\tilde{\alpha}_t(z_t) = p(z_t|\bar{x}_{1:t})$

before, $\alpha_t = o_t \odot A\alpha_{t-1}$

$$\tilde{\alpha}_t = \frac{o_t \odot A\tilde{\alpha}_{t-1}}{\sum_{z_t}(\quad " \quad)}$$

you can show that $\sum_{z_t}(o_t \odot A\tilde{\alpha}_{t-1})(z_t) = p(\bar{x}_t|x_{1:t-1})$

$$\sum_{z_t} (\alpha_t \odot \tilde{a}_{t+1})(z_t) = p(\bar{x}_t | x_{1:t-1}) \triangleq c_t$$

$$p(\bar{x}_{1:T}) = \prod_{t=1}^{T} p(\bar{x}_t | \bar{x}_{1:t-1}) = \prod_{t=1}^{T} c_t$$

- β-recursion:

define $\tilde{\beta}_t(z_t) \triangleq \frac{\beta_t(z_t)}{p(\bar{x}_{t+1:T} | x_{1:t})} \Big\} \prod_{u=t+1}^{T} c_u$

note $\sum_{z_t} \tilde{\beta}_t(z_t) \overset{\text{rec.}}{\neq} 1$

klico

exercise: derive $\tilde{\beta}$-recursion

ML for HMM:

- suppose $p(x_t | z_t = k) = f(x_t | \eta_k)$ ← some parametric model e.g. Gaussian on x_t $\quad \eta = (\eta_k)_{k=1}^{K}$
- $p(z_{t+1} = i | z_t = j) = A_{ij}$
- $p(z_1 = i) = \pi_i$ $\quad \theta = (\eta, A, \pi)$

want to estimate $\hat{\eta}, \hat{A}$ & $\hat{\pi}$ by ML from data $(x^{(i)})_{i=1}^{N}$

$x^{(i)} = x_{1:T_i}^{(i)}$

⇝ use EM

s^{th} iteration

E step: $q_{s+1}(z) = p(z | x, \theta^{[s]})$

M step : $\hat{\theta}^{[s+1]} = \underset{\theta \in \Theta}{\operatorname{argmax}}\ \mathbb{E}_{q_{s+1}}[\log p(x,z)]$

complete log-likelihood :

$$\log p(x,z|\theta) = \sum_{i=1}^{N}\Big[\log p(z_1^{(i)}) + \sum_{t=1}^{T}\log p(x_t^{(i)}|z_t^{(i)}) + \sum_{t=2}^{T}\log p(z_t^{(i)}|z_{t-1}^{(i)})\Big]$$

$\log p(z_1^{(i)}) = \sum_k z_{1,k}^{(i)} \log \pi_k$

$\log p(x_t^{(i)}|z_t^{(i)}) = \sum_k z_{t,k}^{(i)} \log f(x_t^{(i)}|\theta_k)$

$\log p(z_t^{(i)}|z_{t-1}^{(i)}) = \sum_{l,m} z_{t,l}^{(i)} z_{t-1,m}^{(i)} \log A_{lm}$

q huge variable

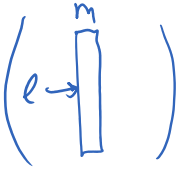
$\mathbb{E}_{q_{s+1}}[\log p(x,z|\theta)] = \dots$

$\mathbb{E}_{q_{s+1}}[z_{t,k}^{(i)}] = q_{s+1}(z_{t,k}^{(i)}=1) \triangleq \gamma_{t,k}^{(i)}$

smoothing distribution $p(z_t^{(i)}|x_{1:T}^{(i)})$

$q_{s+1}(z_{t,l}^{(i)}=1, z_{t-1,m}^{(i)}=1) = p(z_{t,l}^{(i)}=1, z_{t-1,m}^{(i)}=1 | x_{1:T}^{(i)}, \theta^{[s]}) \triangleq \gamma_{t,l,m}^{(i)}$

smoothing edge marginal



maximize with respect to θ

$$\hat{\pi}_k^{[s+1]} = \sum_{i=1}^{N} \gamma_{1,k}^{(i)}$$

$\sum_{i=1}^{N}\sum_{l} \gamma_{1,l}^{(i)} \Big\} N$ (1)

$$\hat{A}_{l,m}^{[s+1]} = \sum_{i=1}^{N}\sum_{t=2}^{T} \gamma_{t,l,m}^{(i)} \Big/ \sum_{u}\sum_{i=1}^{N}\sum_{t=2}^{T} \gamma_{t,u,m}^{(i)}$$

$\hat{\eta}_k \rightarrow$ soft count ML

e.g. Gaussians

similar to GMM

"weighted empirical mean"

"Baum-Welch" alg. — forward-backward alg. (α-β recursion / sum product) + EM for HMM

Viterbi to compute $\underset{z_1,\dots,z_T}{\operatorname{argmax}}\ p(z_{1:T}|x_{1:T})$

(max product)

https://en.wikipedia.org/wiki/Baum%E2%80%93Welch_algorithm