Thursday, 27 April 2023 11:09

AR: - Learn using Pseudo-Inv / Yule-Walker

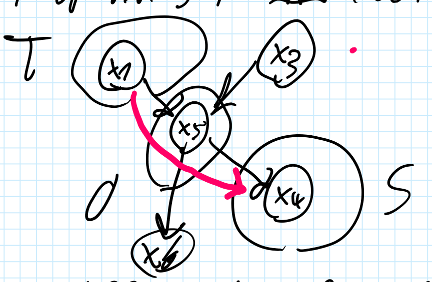
MC: - Markov Property $\forall t: p(z_t | z_{1:t-1}) = p(z_t | z_{t-1})$

- MLE (counting transition frequencies)

HMM: - filtering $p(z_t | x_{1:t})$ => forwards

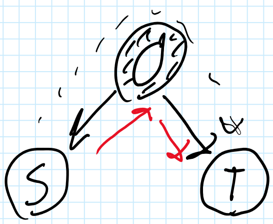
- smoothing $p(z_t | x_{1:T})$ => forwards-Backwards
- MAP $\underset{z_{1:T}}{\arg\max}\ p(z_{1:T} | x_{1:T})$ => Viterbi
- Learning => EM

(Set of r.v. S) $\perp\!\!\!\perp$ (Set of r.v. T) | (set of observed rv O)

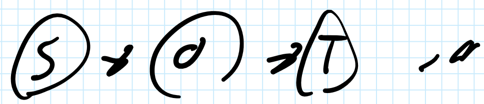


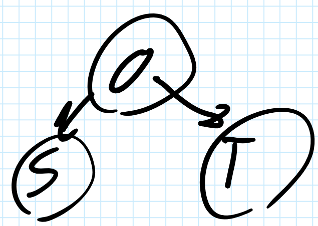
<=> All paths from S to T are "blocked" by O

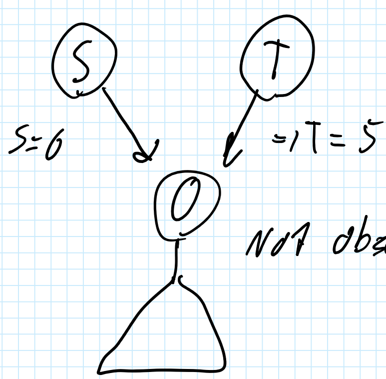
Blocked:



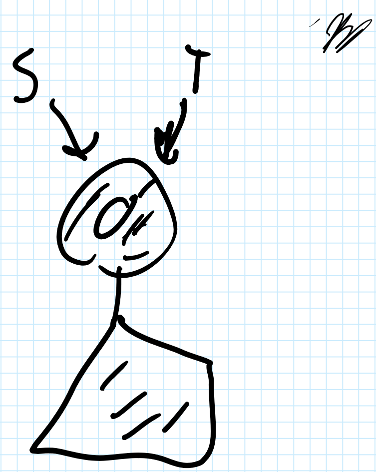
Not blocked







Not observed

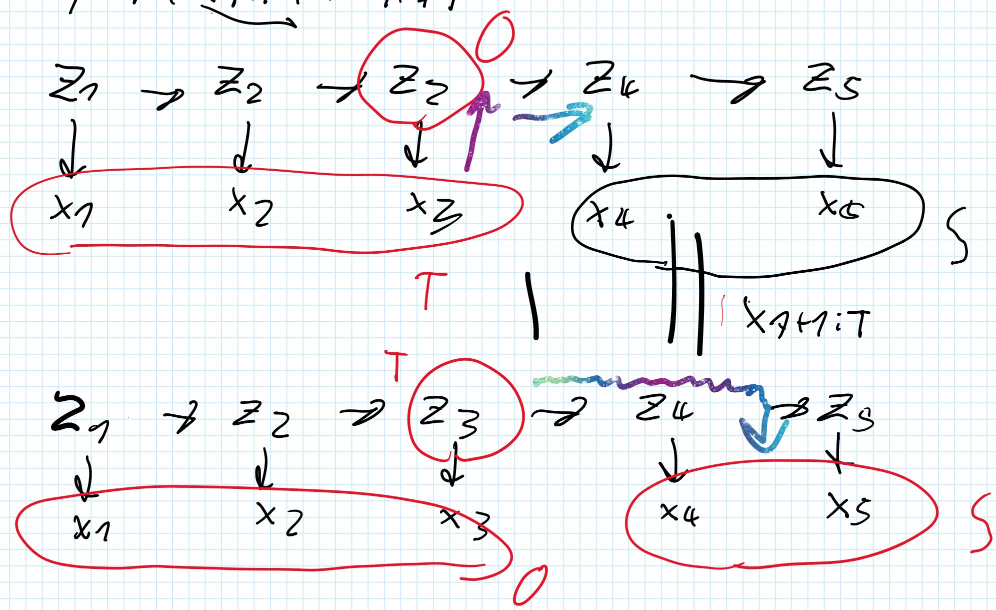


$S \perp\!\!\!\perp T$ when O is not observed

Why $p(X_{t+1:T} \mid Z_t = k, x_{1:t})$

$= p(x_{t+1:T} \mid Z_t = k)$

$\neq p(X_{t+1:T} \mid x_{1:t})$



Problem 2: $(*)$ Let $\mathbf{X}_t$ be a 2-D random vector:

$$\mathbf{X}_t = \begin{bmatrix} u_t \\ v_t \end{bmatrix} \quad \text{where } u_t, v_t \in \{1, 2, \ldots, K\}. \tag{1}$$

Consider the following Markov chain.

$$X_1 \to X_2 \to X_3 \to \cdots \to X_T$$

Model parameters are as follows:

- initial distribution $\boldsymbol{\pi}_x \in \mathbb{R}^{K \times K}$ that parametrizes $\Pr(\mathbf{X}_1)$:

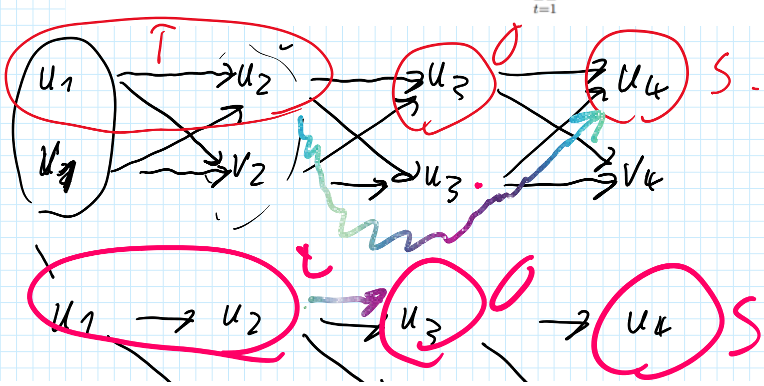
$$\Pr\left(\mathbf{X}_1 = \begin{bmatrix} i \\ j \end{bmatrix}\right) = \boldsymbol{\pi}_x(i, j). \tag{2}$$

- transition probability matrix $\mathbf{A}_x \in \mathbb{R}^{K \times K \times K \times K}$ that parametrizes $\Pr(\mathbf{X}_{t+1} | \mathbf{X}_t)$:

$$\Pr\left(\mathbf{X}_{t+1} = \begin{bmatrix} i_{t+1} \\ j_{t+1} \end{bmatrix} \middle| \mathbf{X}_t = \begin{bmatrix} i_t \\ j_t \end{bmatrix}\right) = \mathbf{A}_x(i_t, j_t, i_{t+1}, j_{t+1}). \tag{3}$$

Because of the Markov property of $\mathbf{X}_t$, the joint probability can be factorized as

$$\Pr(\mathbf{X}_1, \ldots, \mathbf{X}_T) = \Pr(\mathbf{X}_1) \prod_{t=1}^{T-1} \Pr(\mathbf{X}_{t+1} | \mathbf{X}_t).$$



$u_1 \to u_2 \to u_3 \to u_4$ S

$v_1 \quad v_2 \quad v_3 \quad v_4$

$x_1 \qquad x_2$

$u_t \not\perp u_{1:t-2} \mid u_{t-1} \qquad p(u_t \mid u_{1:t-1}) = p(u_t \mid u_{t-1})$

S T O

a) Does the sequence $[u_1, \ldots, u_T]$ (where $u_t \in \{1, 2, \ldots, K\}$ is defined in Eq. (1)) have the first-order Markov property? Why or why not?

b) Let $[Y_1, \ldots, Y_T] \in \{1, 2\}^T$ be a first-order Markov chain with initial probability distribution $\boldsymbol{\pi}_y \in \mathbb{R}^2$ and transition probabilities $\mathbf{A}_y \in \mathbb{R}^{2\times 2}$.

- Briefly explain why the sequence $\begin{bmatrix} Y_2 \\ Y_1 \end{bmatrix}, \begin{bmatrix} Y_3 \\ Y_2 \end{bmatrix}, \ldots, \begin{bmatrix} Y_T \\ Y_{T-1} \end{bmatrix}$ is a 2-D first-order Markov chain.
- Compute initial and transition probabilities, $\boldsymbol{\pi}_x$ and $\mathbf{A}_x$ (defined in Eqs. (2) and (3)) for the sequence $\begin{bmatrix} Y_2 \\ Y_1 \end{bmatrix}, \begin{bmatrix} Y_3 \\ Y_2 \end{bmatrix}, \ldots, \begin{bmatrix} Y_T \\ Y_{T-1} \end{bmatrix}$.

T O

$y_1 \to y_2 \to y_3 \to y_4 \to y_5$ S

$$\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} \perp \begin{pmatrix} y_2 \\ y_1 \end{pmatrix}, \begin{pmatrix} y_3 \\ y_2 \end{pmatrix} \cdots \begin{pmatrix} y_{t-2} \\ y_{t-3} \end{pmatrix} \Bigm| \begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix}$$

S T O

$$p\left(\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} \Bigm| \begin{pmatrix} y_2 \\ y_1 \end{pmatrix} \cdots \begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix} \right)$$

$$= p\left(\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} \Bigm| \begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix} \right)$$

$t \downarrow \cancel{t}1 \qquad \downarrow$

$$\alpha_3 \propto \begin{pmatrix} 0 \\ c \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$\to$

$x_1 = c$

$\alpha_3 \propto | c | (1)$

$\vec{}$

$X_1 = c$

$X_2 = a \quad X_3 = b$

a.) Filtering: $P(Z_3^{,k} | X_{1:3}) \propto P(Z_3 = k, X_{1:3}) := \alpha_3(k)$

$\alpha_1(k) = \pi_k \cdot B_{k, x_1}$

$\alpha_1(1) = B_{1,x_1} \cdot \pi_1$

$\alpha_1(2) = B_{2,x_1} \cdot \pi_2$

$\alpha_1(3) = B_{3,x_1} \; \pi_3$

$\vec{\alpha}_1 = B_{:,x_1} \odot \vec{\pi}$

$\alpha_1 = \begin{pmatrix} 0.8 \\ 0 \end{pmatrix} \odot \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} = \begin{pmatrix} 0.4 \\ 0 \end{pmatrix}$

$\alpha_{t+1}(k) = B_{k,(x_{t+1})} \cdot \sum_j A_{jk} \; \alpha_t(j)$

$\alpha_{t+1}(1) = B_{1,(x_{t+1})} \cdot (A_{:,1})^T \cdot \begin{pmatrix} \alpha_t(1) \\ \alpha_t(2) \\ \alpha_t(3) \end{pmatrix}$

$\alpha_{t+1}(2) = B_{2,(x_{t+1})} \cdot (A_{:,2})^T \cdot$

$\alpha_{t+1}(3) = B_{3,(x_{t+1})} \cdot (A_{:,3})^T$

$B_{:,x_{t+1}} \odot A^T \cdot \vec{\alpha}_t$

$\alpha_2 = \begin{pmatrix} 0.2 \\ 0.4 \end{pmatrix} \odot \begin{pmatrix} 0.2 & 0.5 \\ 0.8 & 0.5 \end{pmatrix} \cdot \begin{pmatrix} 0.4 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.016 \\ 0.128 \end{pmatrix}$

$\alpha_3 = \begin{pmatrix} 0 \\ 0.6 \end{pmatrix} \odot \begin{pmatrix} 0.2 & 0.5 \\ 0.8 & 0.5 \end{pmatrix} \begin{pmatrix} 0.016 \\ 0.128 \end{pmatrix} = \begin{pmatrix} 0 \\ 0.04608 \end{pmatrix}$

b.)

$P(Z_3^{,k} | X_{1:5}) \propto \underbrace{P(Z_3 = k, X_{1:3})}_{\alpha_3(k)} \cdot \underbrace{P(X_{4:T} | Z_3 = k)}_{\beta_3(k)}$

$\beta_T = \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \end{pmatrix}$

$$\beta_T = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$\beta_t = A \cdot \left(B_{:,x_{t+1}} \odot \beta_{t+1} \right)$$

$$A = \begin{pmatrix} 0.2 & 0.8 \\ 0.5 & 0.5 \end{pmatrix} \quad B = \begin{pmatrix} 0.2 & 0 & 0.8 \\ 0.4 & 0.6 & 0 \end{pmatrix} \qquad cabac$$

$$\beta_5 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\beta_4 = \begin{pmatrix} 0.2 & 0.8 \\ 0.5 & 0.5 \end{pmatrix} \cdot \left(\begin{pmatrix} 0.8 \\ 0 \end{pmatrix} \odot \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right) = \begin{pmatrix} 0.16 \\ 0.4 \end{pmatrix}$$

$$\beta_3 = A \left(\begin{pmatrix} 0.2 \\ 0.4 \end{pmatrix} \odot \begin{pmatrix} 0.16 \\ 0.4 \end{pmatrix} \right) = \begin{pmatrix} 0.1344 \\ 0.090 \end{pmatrix}$$

$$P(Z_3 | X_{1:5}) \propto \alpha_3 \odot \beta_3 \propto \begin{pmatrix} 0 \\ c \end{pmatrix} \propto \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

---

c.1

Problem 3: (∗) Consider an HMM where hidden variables are in $\{1, 2\}$ and observed variables are in $\{a, b, c\}$. Let the model parameters be as follows:

$$A = \begin{array}{c} \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{cc} 1 & 2 \end{array} \\ \begin{bmatrix} 0.2 & 0.8 \\ 0.5 & 0.5 \end{bmatrix} \end{array} \qquad B = \begin{array}{c} \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} a & b & c \end{array} \\ \begin{bmatrix} 0.2 & 0 & 0.8 \\ 0.4 & 0.6 & 0 \end{bmatrix} \end{array} \qquad \pi = \begin{array}{c} 1 \\ 2 \end{array} \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$$

Assume that the sequence $X_{1:5} = [cabac]$ is observed.

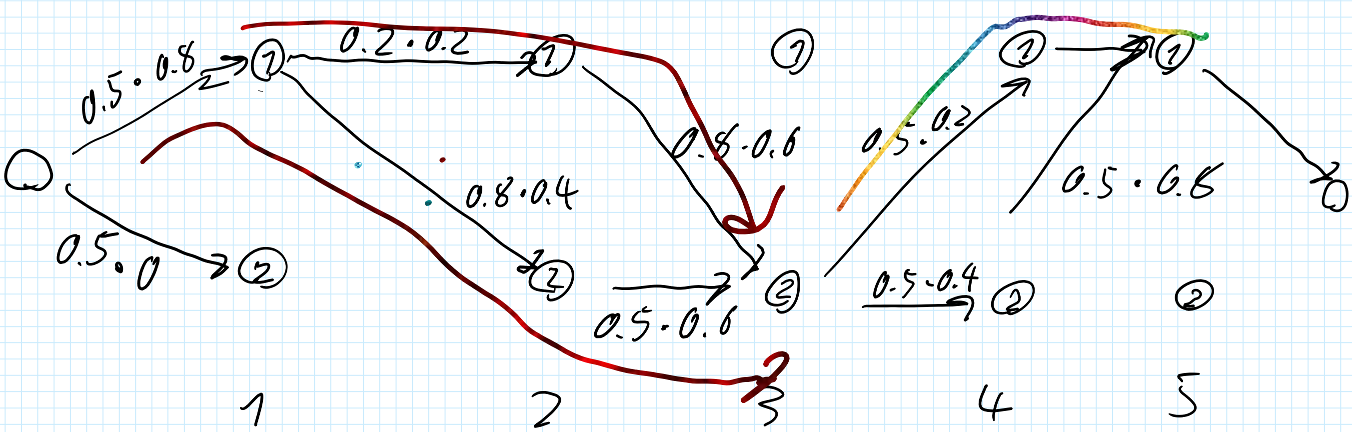
1. Filtering: find the distribution $P(Z_3|X_{1:3})$.
2. Smoothing: find the distribution $P(Z_3|X_{1:5})$.
3. Viterbi algorithm: find the most probable sequence $[Z_1, \ldots, Z_5]$.

$x_1 = c \qquad x_2 = a$

b

$$P(Z_{1:t} | X_{1:t}) \propto P(Z_{1:T}, X_{1:T})$$

$$= P(Z_1) \cdot \prod P(Z_t | Z_{t-1}) \cdot \prod P(X_t | Z_t)$$

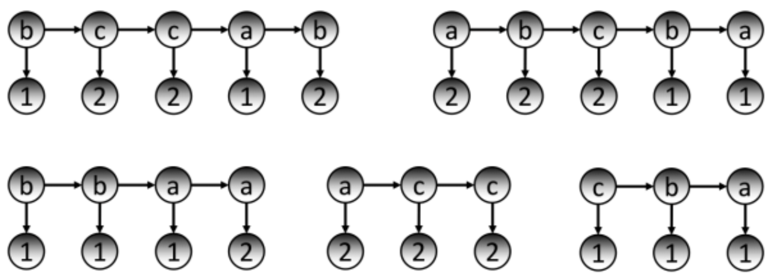


1 2 3 4 5

$$w_{i,j}^{(t)} = P(z_t = j \mid z_{t-1} = i) \cdot P(x_t \mid z_t = j)$$

$[1,2\ 2,2,1]$

Problem 4: Consider an HMM where states Z_t are in $\{a, b, c\}$ and emissions X_t are in $\{1, 2\}$. Given is the following set of fully-observed instances (two sequences of length 5, one sequence of length 4, and two sequences of length 3):



Learn the parameters of the HMM (i.e. $\pi \in \mathbb{R}^3, \mathbf{A} \in \mathbb{R}^{3\times3}$, and $\mathbf{B} \in \mathbb{R}^{3\times2}$) using maximum-likelihood estimation.

$$\min \sum -\log(w)$$

$$\max \prod w$$

$Z =$ Weather $\begin{pmatrix} \text{Rainy} \\ \text{Cloudy} \\ \text{Sunny} \end{pmatrix}$ $\quad X = \#$ Rain

$$\max P(z_{1:T}, x_{1:T})$$

$$\max \log P(z_{1:T}, x_{1:T})$$

$$= \underset{\pi}{\log P(z_1)} + \underset{A}{\sum_{t=2}^{T_1} \log P(z_t \mid z_{t-1})} + \underset{B}{\sum_{t=1}^{T_1} \log p(x_t \mid z_t)}$$

$$\underbrace{\sum_{n=1}^{5} \sum_{k} I(z_1 = k) \cdot \log \pi_k} + \sum_{t=2}^{T_n} \sum_{i,j} I(z_t = j, z_{t-1} = i) \cdot \log(A_{i,j})$$

$$+ \sum_{t=1}^{T_n} \sum_{i',j'} I(x_t = j', z_t = i') \cdot \log B(i', j')$$

$$\max \sum_{k} \sum_{n=1}^{5} I(z_1 = k) \cdot \log \pi_k + \ldots$$

$$\text{s.t.} \sum_{k} \pi_k = 1$$

$$\underbrace{\max_{\lambda} \max_{\pi} \sum_{k} \sum_{n=1}^{5} I(z_1 = k) \cdot \underline{\log \pi_k} + \lambda \cdot \left(\sum_{k} \pi_k - 1 \right)}$$

$$\frac{\partial}{\ldots} = \partial \sum I(z_1 = k) \cdot \frac{1}{\ldots} + \lambda \overset{!}{=} 0$$

$$\frac{\partial}{\partial \pi_k} = \sum_n I(z_1 = k) \cdot \frac{1}{\pi_k} + \lambda \overset{!}{=} 0$$

$$\Leftrightarrow \quad \frac{1}{\pi_k} \cdot \sum_n I(z_1 = k) + \lambda \overset{!}{=} 0$$

$$\Leftrightarrow \quad \sum_n I(z_1 = k) + \lambda \pi_k \overset{!}{=} 0$$

$$\Rightarrow \quad \pi_k = - \frac{\sum_n I(z_1 = k)}{\lambda} \quad \propto \begin{pmatrix} \sum_k I(z_1 = k) / N \\ \sum_k I(z_1 = 2) / N \\ \sum_k I(z_1 = 3) / N \end{pmatrix}$$

Problem 1: Consider the stationary AR(p) process $X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon$ with $\epsilon \sim \mathcal{N}(0, \sigma^2)$. We denote by μ the mean $E[X_t]$ and by γ_i the autocovariance $Cov(X_t, X_{t-i})$. Show:

1. $\mu = \frac{c}{1 - \sum_{i=1}^{p} \phi_i}$, for all t
2. $\gamma_0 = \sum_{j=1}^{p} \phi_j \gamma_{-j} + \sigma^2$
3. $\gamma_i = \sum_{j=1}^{p} \phi_j \gamma_{i-j}$, for all $t, i \in [1, p]$