Machine Learning for Graphs and Sequential Data Exercise Sheet 06

Autoregressive Models, Markov Chains, Hidden Markov Models

Exercises marked with a $(*)$ will be discussed in the in-person exercise session.

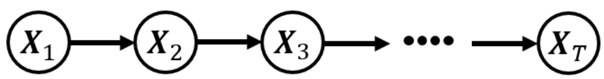
Problem 1: Consider the stationary AR(p) process $X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon$ with $\epsilon \sim \mathcal{N}(0, \sigma^2)$. We denote by μ the mean $E[X_t]$ and by γ_i the autocovariance $Cov(X_t, X_{t-i})$. Show:

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 $t, i \in [1, p]$

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.



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

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

In this task, we refer to this model as *"2-D first-order Markov chain"*.

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 $\begin{bmatrix} Y_2 \\ Y_1 \end{bmatrix}, \begin{bmatrix} Y_3 \\ Y_2 \end{bmatrix}, \dots, \begin{bmatrix} Y_T \\ Y_{T-1} \end{bmatrix}$ 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 $\begin{bmatrix} Y_2 \\ Y_1 \end{bmatrix}, \begin{bmatrix} Y_3 \\ Y_2 \end{bmatrix}, \dots, \begin{bmatrix} Y_T \\ Y_{T-1} \end{bmatrix}$.

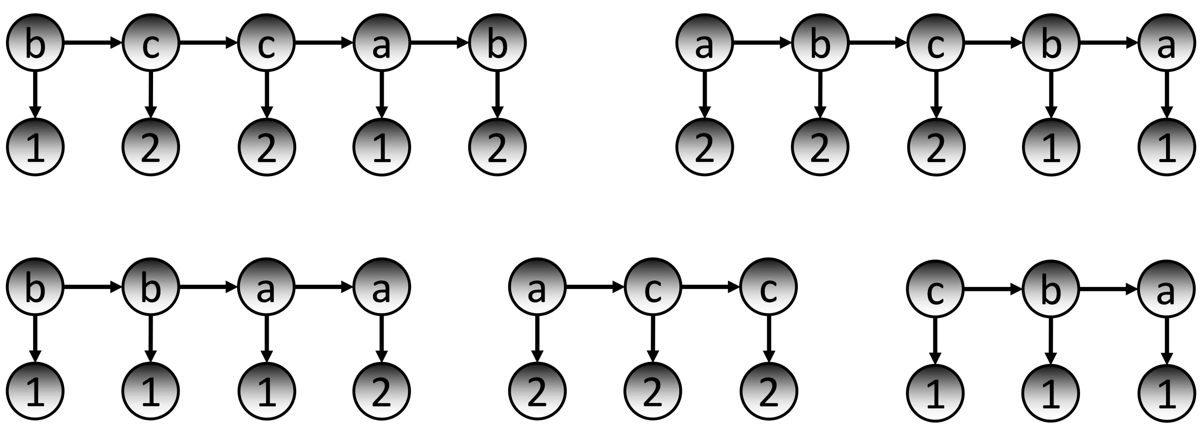
Problem 3: $(*)$ Consider an HMM where hidden variables are in $\{1, 2\}$ and observed variables are in $\{a, b, c\}$. Let the model parameters be as follows:

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

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 $[Z_1, \dots, Z_5]$.

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.