Machine Learning for Graphs and Sequential Data Exercise Sheet 06

Autoregressive Models, Markov Chains

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

The AR(p) process is stationary i.e. $E[X_t] = \mu$ and $Cov(X_t, X_{t-i}) = \gamma_i$.

1. $\underbrace{E[X_t]}_{\mu} = c + \sum_{i=1}^{p} \phi_i \underbrace{E[X_{t-i}]}_{\mu} + \underbrace{E[\epsilon]}_{0} \Rightarrow \mu = \frac{c}{1-\sum_{i=1}^{p} \phi_i}$
2. $Cov(X_t, X_t) = \underbrace{Cov(c, X_t)}_{0} + \sum_{i=1}^{p} \phi_i \underbrace{Cov(X_{t-i}, X_t)}_{\gamma_{-i}} + \underbrace{Cov(\epsilon_t, X_t)}_{\sigma^2} \Rightarrow \gamma_0 = \sum_{j=1}^{p} \phi_j \gamma_{-j} + \sigma^2$
3. $Cov(X_t, X_{t-i}) = \underbrace{Cov(c, X_{t-i})}_{0} + \sum_{j=1}^{p} \phi_i \underbrace{Cov(X_{t-j}, X_{t-i})}_{\gamma_{j-i}} + \underbrace{Cov(\epsilon_t, X_{t-i})}_{0} \Rightarrow \gamma_i = \sum_{j=1}^{p} \phi_j \gamma_{i-j}$

Problem 2:

a) Consider the following AR(1) process $X_t = c + \phi_1 X_{t-1} + \epsilon$ with $\epsilon \sim \mathcal{N}(0, \sigma^2)$. Show that this process is stationary iff the following condition is fulfilled

$$|\phi_1| < 1.$$

AR process is stationary iff the roots of its characteristic polynomial lie outside of the unit circle. The characteristic polynomial is

$$p(x) = 1 - \phi_1 x$$

The unique root is clearly $\frac{1}{\phi_1}$ (in case of $\phi_1 = 0$ we get stationary $X_t \sim \mathcal{N}(c, \sigma^2)$). Hence, the roots lie outside of the unit circle iff $|\phi_1| < 1$.

b) Consider the following AR processes:

- $X_t = c + .8 \times X_{t-1} + .1 \times X_{t-2} + \epsilon$ with $\epsilon \sim \mathcal{N}(0, \sigma^2)$
- $X_t = -\sum_{k=1}^{p} \binom{p}{k} X_{t-k} + \epsilon$ with $\epsilon \sim \mathcal{N}(0, \sigma^2)$

Are these processes stationary ?

- We have $\phi_1 = .8$ and $\phi_2 = .1$. Since $.8 + .1 < 1$, $.1 - .8 < 1$, $|.1| < 1$, the process is stationnary.
- The characteristic polynomial is $p(x) = 1 + \sum_{k=1}^{p} \binom{p}{k} x^k = \sum_{k=0}^{p} \binom{p}{k} x^k = (1+x)^p$. All the roots are equal to -1 which do not stricly lie outside the unit circle. The process is not stationnary.

Problem 3: Let $\mathbf{X}_t$ be a 2-D random vector:

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

Consider the following Markov chain.

$$\boldsymbol{X}_1 \to \boldsymbol{X}_2 \to \boldsymbol{X}_3 \to \cdots\cdots \to \boldsymbol{X}_T$$

Model parameters are as follows:

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

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

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

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

The joint probability can be factorized as:

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

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

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

The variable u_t depends on u_{t-1} and v_{t-1}. Moreover, v_{t-1} depends on u_{t-2}. So, u_t and u_{t-2} are not conditionally independent given u_{t-1} only. As a consequence, u_t is not a markov chain.

b) Let $[Y_1, ..., Y_T]$ be a 1-D first-order Markov chain with the following initial and transition probabilities ($Y_1, ..., Y_T$ are binary-valued).

$$\boldsymbol{\pi}_y = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}, \qquad \mathbf{A}_y = \begin{bmatrix} 0.2 & 0.8 \\ 0.5 & 0.5 \end{bmatrix}.$$

- Briefly explain why the sequence $\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}, \begin{bmatrix} Y_2 \\ Y_3 \end{bmatrix}, ..., \begin{bmatrix} Y_{T-1} \\ Y_T \end{bmatrix}$ is a 2-D first-order Markov chain.

We compute $P(\begin{bmatrix} Y_t \\ Y_{t-1} \end{bmatrix} \mid \begin{bmatrix} Y_{t-1} \\ Y_{t-2} \end{bmatrix}, ..., \begin{bmatrix} Y_2 \\ Y_1 \end{bmatrix})$. Since $(Y_t)_{t=1}^T$ is a markov chain, Y_t is independent on $Y_{t-2}, ..., Y_1$ conditioned on Y_{t-1}. Also, Y_{t-1} is independent on $Y_{t-2}, ..., Y_1$ conditioned on Y_{t-1}. As a consequence, we have:

$$P([Y_t, Y_{t-1}] | [Y_{t-1}, Y_{t-2}], ..., [Y_2, Y_1]) = P([Y_t, Y_{t-1}] | [Y_{t-1}, Y_{t-2}])$$

- Compute initial and transition probabilities, $\boldsymbol{\pi}_x$ and $\mathbf{A}_x$ (defined in Eqs. 2 and 3) for the sequence $\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}, \begin{bmatrix} Y_2 \\ Y_3 \end{bmatrix}, ..., \begin{bmatrix} Y_{T-1} \\ Y_T \end{bmatrix}$.

We first compute $\boldsymbol{\pi}_x$:

$$\begin{aligned} \boldsymbol{\pi}_x(i,j) &= P(\begin{bmatrix} Y_2 \\ Y_1 \end{bmatrix} = \begin{bmatrix} j \\ i \end{bmatrix}) \\ &= P(Y_2 = j | Y_1 = i) P(Y_1 = i) = \mathbf{A}_y(i,j) \boldsymbol{\pi}_y(i) \end{aligned}$$

We compute now $\mathbf{A}_x$:

$$\begin{aligned} \mathbf{A}_x(i,j,i',j') &= P(\begin{bmatrix} Y_t \\ Y_{t-1} \end{bmatrix} = \begin{bmatrix} j \\ i \end{bmatrix} \mid \begin{bmatrix} Y_{t-1} \\ Y_{t-2} \end{bmatrix} = \begin{bmatrix} j' \\ i' \end{bmatrix}) \\ &= \begin{cases} 0 & \text{if } i \neq j' \\ \mathbf{A}_y(i,j) & \text{otherwise} \end{cases} \end{aligned}$$

If $i \neq j'$, it is impossible that Y_{t-1} takes two different values. Otherwise, if $i = j'$, we have indeed:

$$\begin{aligned} P(\begin{bmatrix} Y_t \\ Y_{t-1} \end{bmatrix} = \begin{bmatrix} j \\ i \end{bmatrix} \mid \begin{bmatrix} Y_{t-1} \\ Y_{t-2} \end{bmatrix} = \begin{bmatrix} i \\ i' \end{bmatrix}) &= P(Y_t = j | Y_{t-1} = i, Y_{t-2} = i') P(Y_{t-1} = i | Y_{t-1} = i, Y_{t-2} = i') \\ &= P(Y_t = j | Y_{t-1} = i) P(Y_{t-1} = i | Y_{t-1} = i) = \mathbf{A}_y(i,j) \times 1 \end{aligned}$$