


Machine Learning for Graphs and Sequential Data

Sequential Data – Autoregressive Models

Lecturer: Prof. Dr. Stephan Günnemann
cs.cit.tum.de/daml

Summer Term 2023

Data Analytics and
Machine Learning 

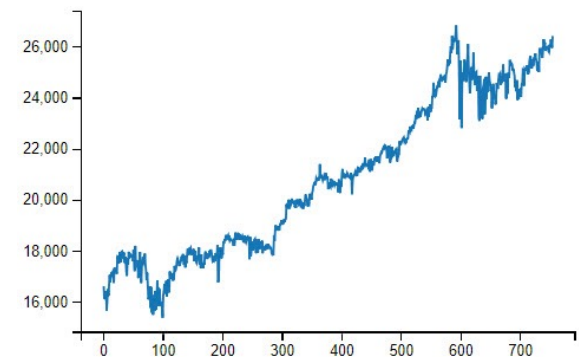
Roadmap

- Chapter: Temporal Data / Sequential Data
 1. **Autoregressive Models**
 - Motivation & Definitions
 - Parameter Learning
 - Stationarity
 2. Markov Chains
 3. Hidden Markov Models
 4. Neural Network Approaches
 5. Temporal Point Processes

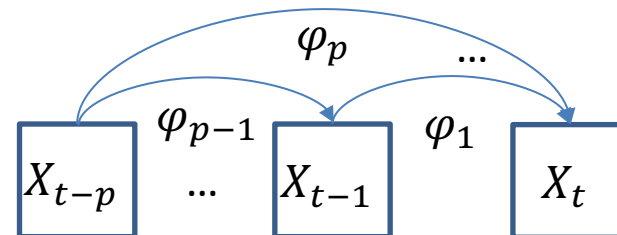
Motivation

- Autoregressive (AR) models for sequences of observations X_1, X_2, \dots, X_T .
 - The index t can correspond to time, location, etc.
 - For now, we focus on the case of **continuous observations** occurring at **discrete time-steps**
- Example: Time-series forecasting
 - X_t = measurement of a sensor at time-step t
 - Applications in weather forecasting,
e.g., X_t = temperature on t -th day
 - Applications in the field of economics,
e.g., X_t = stock market quotations on t -th day

Observations are
not independent
→ non-i.i.d. data



AR model - Definition



- Definition: An **autoregressive model** AR(p) of order p is defined as:

$$X_t = c + \sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t$$

where $\varphi_1, \dots, \varphi_p$ are the parameters, c is a constant and $\varepsilon_t \sim N(0, \sigma)$ is a **white noise**. The variable X_{t-i} is the **lagged value** at time i .

聚合 t 时间点前的 p 个 $\varphi_i X_i$

- Intuitively, we perform a regression where the lags $([X_{t-1}, \dots, X_{t-p}])_t$ are the inputs and are $(X_t)_t$ the outputs.

滞后项
 $X_t - \dots$

- Remark that a modification (or shock) on X_t will have a repercussion far into the future. The variables $(X_t)_t$ are not independent.

AR model – Graphical Model

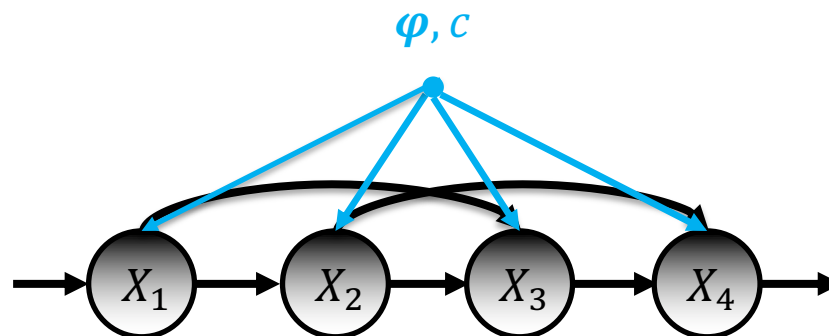
$$X_t = c + \underbrace{\sum_{i=1}^p \varphi_i X_{t-i}}_{\checkmark} + \underbrace{\varepsilon_t}_{\sim \mathcal{N}(0, \sigma)}$$

- We can rewrite the AR model:

$$P(X_t | X_{t-1}, \dots, X_{t-p}) \sim N\left(c + \sum_{i=1}^p \varphi_i X_{t-i}, \sigma\right)$$

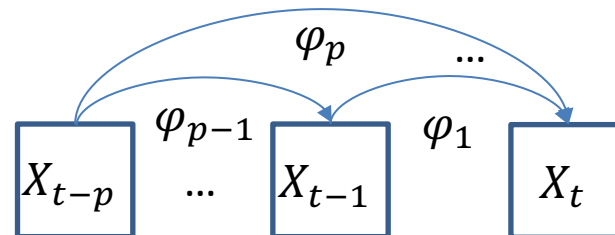
- The graphical model representation of the AR model is:

- The parameters $\boldsymbol{\varphi}, c$ are shared through time



- AR model can be viewed as a probabilistic model for continuous observations

AR model - Definition



- The mean function of an AR model is $\mu(t) = E[X_t]$.
By default, it depends on t .
- The autocovariance $\gamma(t, i) = Cov(X_t, X_{t-i})$.
By default, it depends on t and i .
- The autocovariance function can be normalized to give the Pearson autocorrelation function $\rho(t, i) = \frac{Cov(X_t, X_{t-i})}{\sqrt{Var(X_t)}\sqrt{Var(X_{t-i})}}$. It lies in $[-1, 1]$.
- The autocorrelation and autocovariance are indicators of the dependence of the variable X_t with respect to the past variables X_{t-i}

Roadmap

- Chapter: Temporal Data / Sequential Data

1. **Autoregressive Models**

- Motivation & Definitions
- **Parameter Learning**
- Stationarity

2. Markov Chains

3. Hidden Markov Models

4. Neural Network Approaches

5. Temporal Point Processes

AR model – Parameter Learning (1)

- The parameters can be learned with classic least squares regression:

$$\begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_{p-1} \\ \varphi_p \end{bmatrix} = (X^T X)^{-1} X^T y$$

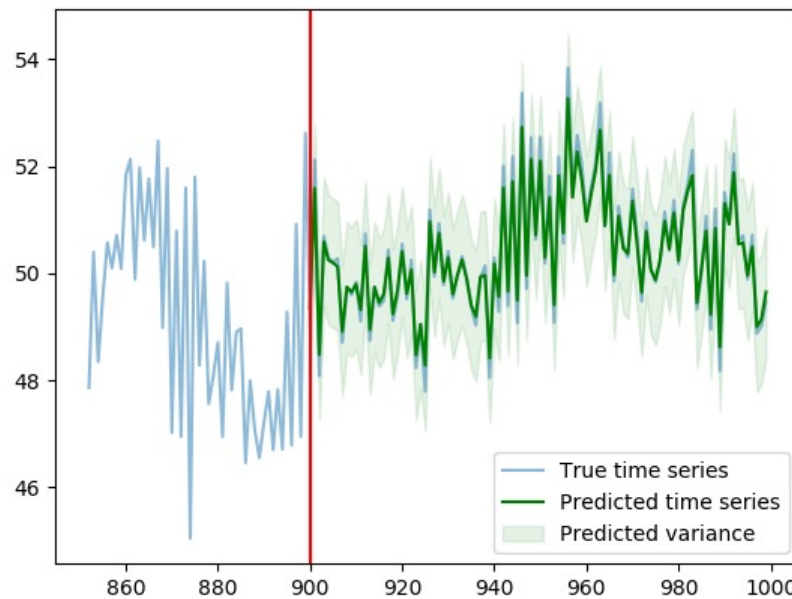
$$\text{where } X = \begin{bmatrix} X_{p-1} & \cdots & X_0 \\ X_p & \cdots & X_1 \\ \vdots & \cdots & \vdots \end{bmatrix}$$

$$\text{and } y = \begin{bmatrix} X_p \\ X_{p+1} \\ \vdots \end{bmatrix}$$

$$y = X \cdot \Phi$$

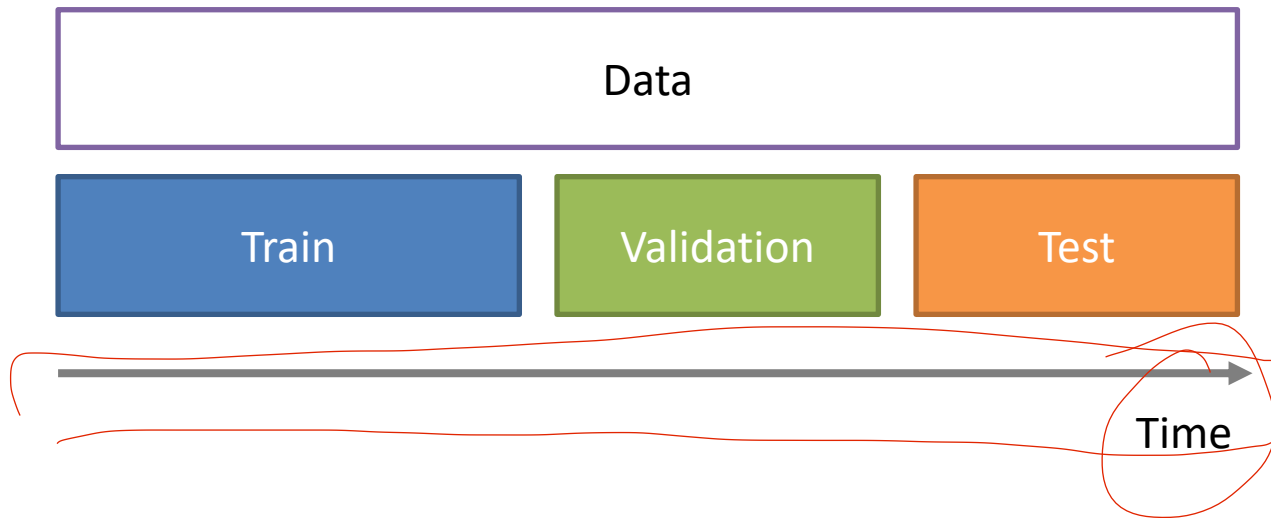
AR model – Parameter Learning

- Example: $X_t = 5 + 0.8 * X_{t-1} + 0.1 * X_{t-2} + N(0,1)$
 1. We learn parameters and the variance on the 900 first samples
 - Estimated parameters: $\varphi_1 = 0.82$, $\varphi_2 = 0.07$, $\sigma = 1.21$
 2. We predict on the last 100 samples



General Remark: Data split

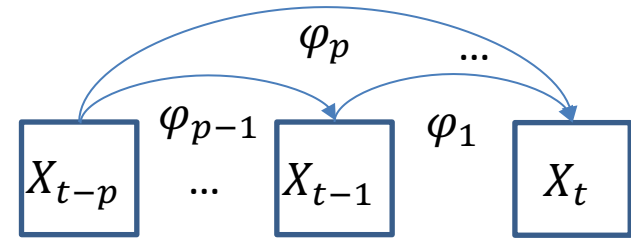
- An important part of training is model selection
 - Usually we split data into train, validation and test set
- With time series and sequential data these sets should be split in such a way to keep the temporal ordering
- A model should be tested only on the data from the future



Roadmap

- Chapter: Temporal Data / Sequential Data
 1. **Autoregressive Models**
 - Motivation & Definitions
 - Parameter Learning
 - **Stationarity**
 2. Markov Chains
 3. Hidden Markov Models
 4. Neural Network Approaches
 5. Temporal Point Processes

AR model - Stationarity



- Definition: A process is said to be stationary if

1. $E[X_t] = E[X_{t-i}] = \mu, \forall t, \forall i$

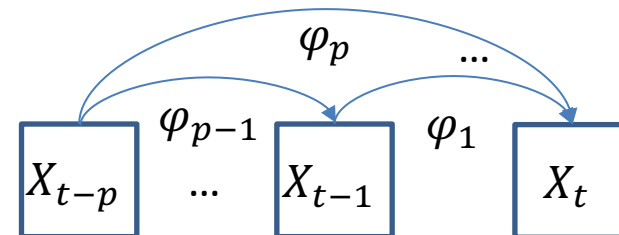
2. $Cov(X_t, X_{t-i}) = \gamma_i, \forall t, \forall i$

3. $E[|X_t|^2] < \infty, \forall t$

- The mean function $E[X_t]$ is constant.
- The autocovariance $Cov(X_t, X_{t-i})$ only depends on the lagged value at time i . It does not depend on t .

Remark: we have $\gamma_i = Cov(X_t, X_{t-i}) = Cov(X_{t-i}, X_t) = \gamma_{-i}$

AR model - Stationarity



- Definition: A process is said to be stationary if

1. $E[X_t] = E[X_{t-i}] = \mu, \forall t, \forall i$
2. $Cov(X_t, X_{t-i}) = \gamma_i, \forall t, \forall i$
3. $E[|X_t|^2] < \infty, \forall t$

- Benefits:

- Stationarity is often a good modeling assumption (like i.i.d. assumption).
- For stationary processes, it is possible to estimate mean and autocovariance by averaging measures over time.
- In other words: Parameters learned from past observations will generalize to future observations.

AR model – Parameter Learning (2)

- Assuming stationarity, the parameters can also be learned by using the Yule Walker equations:

$$\gamma_0 = \sum_{j=1}^p \varphi_j \gamma_{-j} + \sigma^2$$

$$\gamma_1 = \sum_{j=1}^p \varphi_j \gamma_{1-j}$$

$$\gamma_2 = \sum_{j=1}^p \varphi_j \gamma_{2-j}$$

...

$$\gamma_p = \sum_{j=1}^p \varphi_j \gamma_{p-j}$$

1. Estimate the moments $\gamma_0, \gamma_1, \dots, \gamma_p$

2. Inverse Yule-Walker matrix to estimate $\varphi_1, \dots, \varphi_p$

3. Use γ_0 equation to estimate σ

$$\begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_{p-1} \\ \gamma_p \end{bmatrix} = \begin{bmatrix} \gamma_0 & \gamma_{-1} & \cdots & \gamma_{2-p} & \gamma_{1-p} \\ \gamma_1 & & & & \gamma_{2-p} \\ \vdots & & \ddots & & \vdots \\ \gamma_{p-2} & & & & \gamma_{-1} \\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_1 & \gamma_0 \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_{p-1} \\ \varphi_p \end{bmatrix}$$

Yule-Walker
matrix

Questions – AR

1. What is the mean $E[X_t]$ of the following processes:
 - a) $X_t = \sin(t/10) + \varepsilon_t$ where $\varepsilon_t \sim N(0, \sigma)$
 - b) $X_t = 4 - 0.8 * X_{t-1} - 0.1 * X_{t-2} + \varepsilon_t$ where $\varepsilon_t \sim N(0, \sigma)$
 - c) $X_t = 4 + X_{t-1} + \varepsilon_t$ where $\varepsilon_t \sim N(0, \sigma)$ and $X_0 \sim N(0, \sigma)$

2. Why does Yule Walker parameter learning require stationarity?

Reading Material

- [1] Stationary Models lecture, Matthieu Stigler:
<http://matthieustigler.github.io/Lectures/Lect2ARMA.pdf>
- [2] Time Series lecture, Rauli Susmel:
<https://www.bauer.uh.edu/rsusmel/phd/ec2-3.pdf>
- [3] Introduction on AR Model lecture, Rob Reider:
https://s3.amazonaws.com/assets.datacamp.com/production/course_4267/slides/chapter3.pdf