Friedrich-Alexander-Universität
Technische Fakultät

Machine Learning Data Analytics

Machine Learning for Time Series (MLTS)

Lecture 1: Time Series Fundamentals and Definitions (Part 1)

Dr. Dario Zanca, Dr. Emmanuelle Salin
Machine Learning and Data Analytics (MaD) Lab
Friedrich-Alexander-Universität Erlangen-Nürnberg
17.10.2024

Organisational Information

Machine Learning for Time Series (MLTS)

- 5 ECTS
- Lectures + Exercises

Topics overview

1. Time series fundamentals and definitions (Part 1)
2. Time series fundamentals and definitions (Part 2)
3. Bayesian Inference and Gaussian Processes
4. State space models (Kalman Filters)
5. State space models (Particle Filters)
6. Autoregressive models
7. Data mining on time series
8. Deep Learning (DL) for Time Series (Introduction to DL)
9. DL – Convolutional models (CNNs)
10. DL – Recurrent models (RNNs and LSTMs)
11. DL – Attention-based models (Transformers)
12. DL – From BERT to ChatGPT
13. DL – New Trends in Time Series processing
14. Time series in the real world

Topics overview

1. Time series fundamentals and definitions (Part 1)
2. Time series fundamentals and definitions (Part 2)
3. Bayesian Inference and Gaussian Processes
4. State space models (Kalman Filters)
5. State space models (Particle Filters)
6. Autoregressive models
7. Data mining on time series
8. Deep Learning (DL) for Time Series (Introduction to DL)
9. DL – Convolutional models (CNNs)
10. DL – Recurrent models (RNNs and LSTMs)
11. DL – Attention-based models (Transformers)
12. DL – From BERT to ChatGPT
13. DL – New Trends in Time Series processing
14. Time series in the real world

Course times

Lectures (on campus) - *Dr. Dario Zanca and Dr. Emmanuelle Salin*

- Lectures on Thursdays, h. 14.15 - 15.45 (90 mins)
- Consultation hours by appointment
- Recordings from past years available (only **partial** topics overlap)

Exercises (on campus) – *Richard Dirauf, M.Sc.*

- Exercises on Thursdays, h. 12.15 – 13.45
  - starting on October 31st
- Recordings from past years available (only **partial** topics overlap)

Course StudON

StudOn 2024-2025: https://www.studon.fau.de/crs5911979.html

Content

MLTS Exam

Use this forum to discuss issues related to the exam. Instructors also use this channel to share information about upcoming examinations.

Articles (Unread): 0 (0)

MLTS Exercises

All resources of the MLTS exercises

MLTS Lectures and Exercises

Use this forum to discuss content from lectures and exercises.

Articles (Unread): 0 (0)

Practice Questions

Test your understanding of the MLTS lecture and exercise topics with our Practice Questions. Designed to assess your knowledge across the multiple subtopics, these questions offer insights into your grasp of key concepts. Note: The number of points you...

Status: Offline

Exams and evaluation

Written Exam (5 ECTS)

- Written examination
- 60% content from lectures, 40% content form exercises
- The exam will be in person and it will be a closed-book exam

Course organizers
Lecturers

Machine Learning and Data Analytics (MaD) Lab

- Dr. Dario Zanca, dario.zanca@fau.de *
- Dr. Emmanuelle Salin, emmanuelle.salin@fau.de *
- Prof. Dr. Björn Eskofier, bjoern.eskofier@fau.de

* Please, address all your correspondence about the course to Dr. Dario Zanca and Dr. Emmanuelle Salin

Course organizers
Teaching assistants

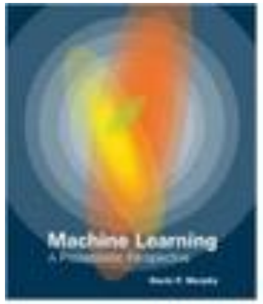
Machine Learning and Data Analytics (MaD) Lab

- Richard Dirauf (M.Sc.), richard.dirauf@fau.de

References

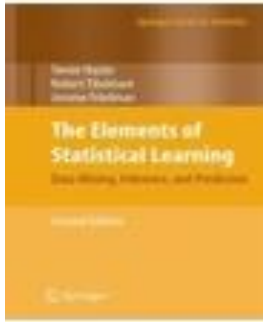
Machine learning: A Probabilistic Perspective,

by Kevin Murphy (2012)



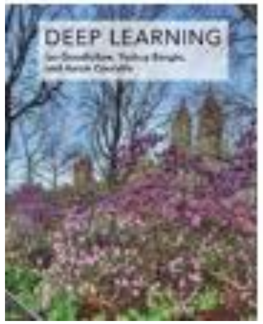
The Elements of Statistical Learning: Data Mining, Inference, and Prediction

by Trevor Hastie, Robert Tibshirani, and Jerome Friedman (2009)



Deep Learning

by Ian Goodfellow, Yoshua Bengio, and Aaron Courville (2016)



Time series fundamentals

Motivations

An old history of time series analysis: Babylonian astronomical diaries

VII century B.C.

“[…] Night of the 5th, beginning of the night, the moon was 2 ½ cubits behind Leonis […] Night of the 17th, last part of the night, the moon stood 1 ½ cubits behind Mars, Venus was below.”

- Babylonians collected the earliest evidence of periodic planetary phenomena
- Applied their mathematics for systematic astronomic predictions

An old history of time series analysis: Babylonian astronomical diaries

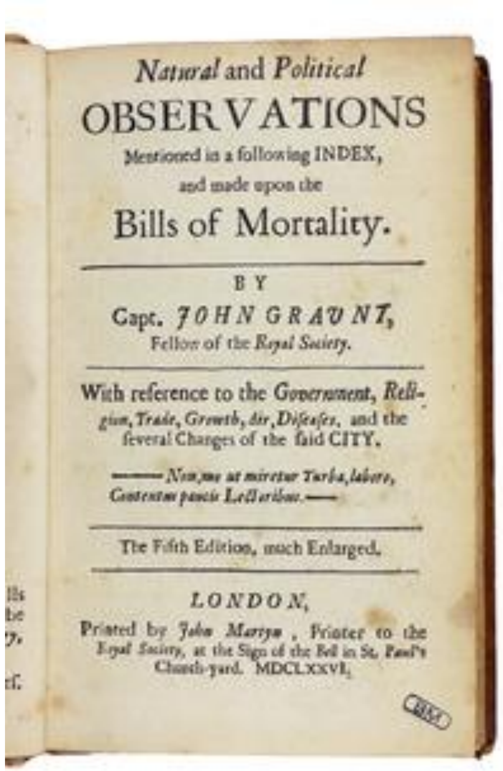
Nowadays, thousands of ground-based and space-based telescopes[a] generate new knowledge every night.

- The Vera C. Rubin Observatory in Chile is geared up to collect 20 terabytes per night from 2022[b].
- The Square Kilometre Array, the world's largest radio telescope, will generate up to 2 petabytes daily, starting in 2028.
- The Very Large Array (ngVLA) will generate hundreds of petabytes annually.

[a] https://research.arizona.edu/stories/space-versus-ground-telescopes
[b] https://www.nature.com/articles/d41586-020-02284-7

1662, John Graunt describes the data collection:

"When anyone dies, […] the same is known to the Searchers, corresponding with the said Sexton. The Searchers hereupon...examine by what Disease, or Casualty the corps died. Hereupon they make their Report to the Parish-Clerk, and he, every Tuesday night, carries in an Accompt of all the Burials, and Christnings, hapning that Week, to the Clerk of the Hall."

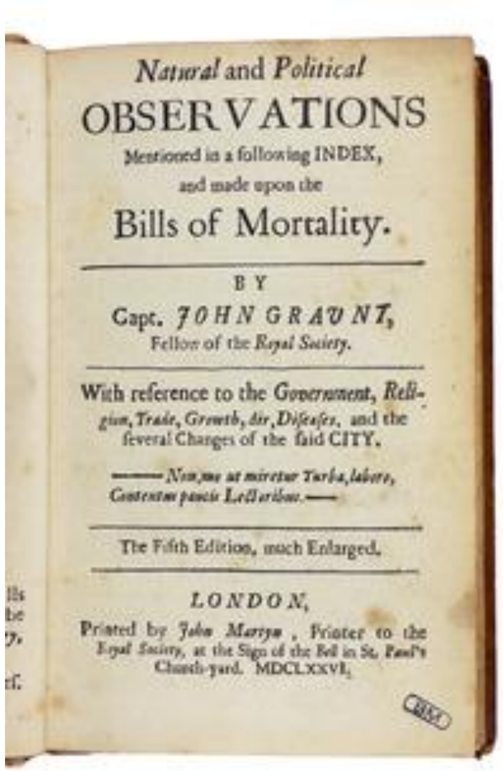


An old history of time series analysis: The Birth of Epidemiology

1662, John Graunt describes the data collection:

”When anyone dies, […] the same is known to the Searchers, corresponding with the said Sexton. The Searchers hereupon...examine by what Disease, or Casualty the corps died. Hereupon they make their Report to the Parish-Clerk, and he, every Tuesday night, carries in an Accompt of all the Burials, and Christnings, hapning that Week, to the Clerk of the Hall.”

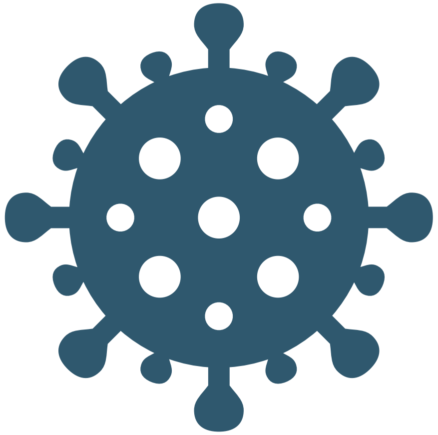
- Rudimentary conclusions about the mortality and morbidity of certain diseases
- Graunt's work is still used today to study population trends and mortality



Epidemiology nowadays

Overview of Modern Epidemiology:

- **Data-Driven:** Utilizes large health datasets.
- **Infectious Disease Tracking:** Focus on emerging infections.
- **Genetic and Global Health:** Incorporates genetics and global health issues.



Importance of Time Series Processing:

- **Trend Analysis:** Identifies patterns and seasonality.
- **Prediction & Forecasting:** Models future disease spread and resource needs.
- **Surveillance:** Early detection and intervention monitoring.

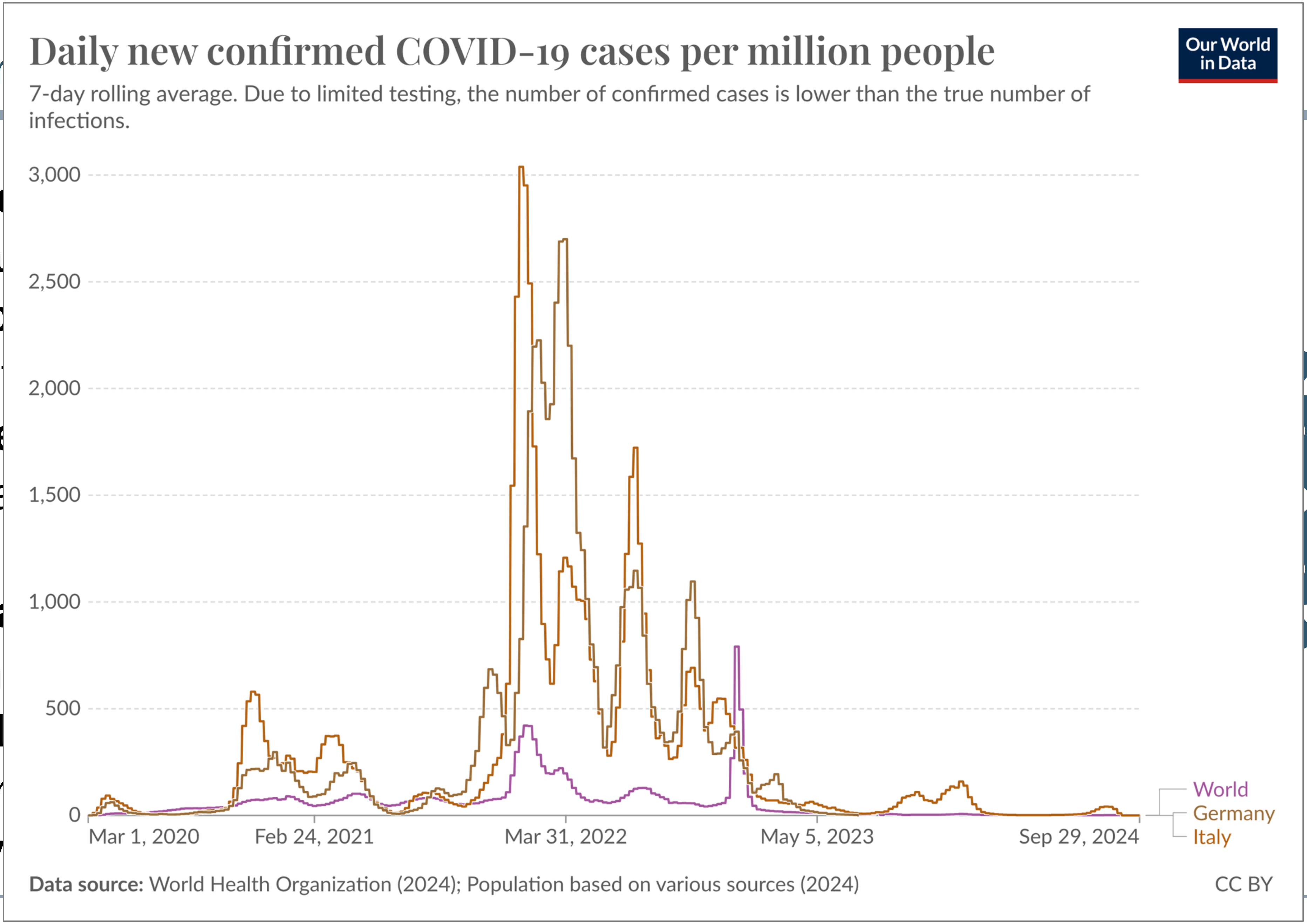
Epidem

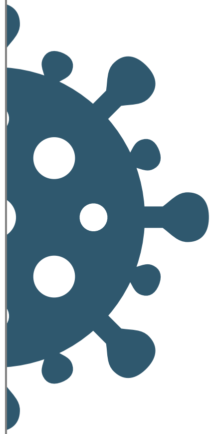
Overvi

- **Data**
- **Infec** infec
- **Gene** globa

Importa

- **Tren**
- **Pred** and r
- **Surv**



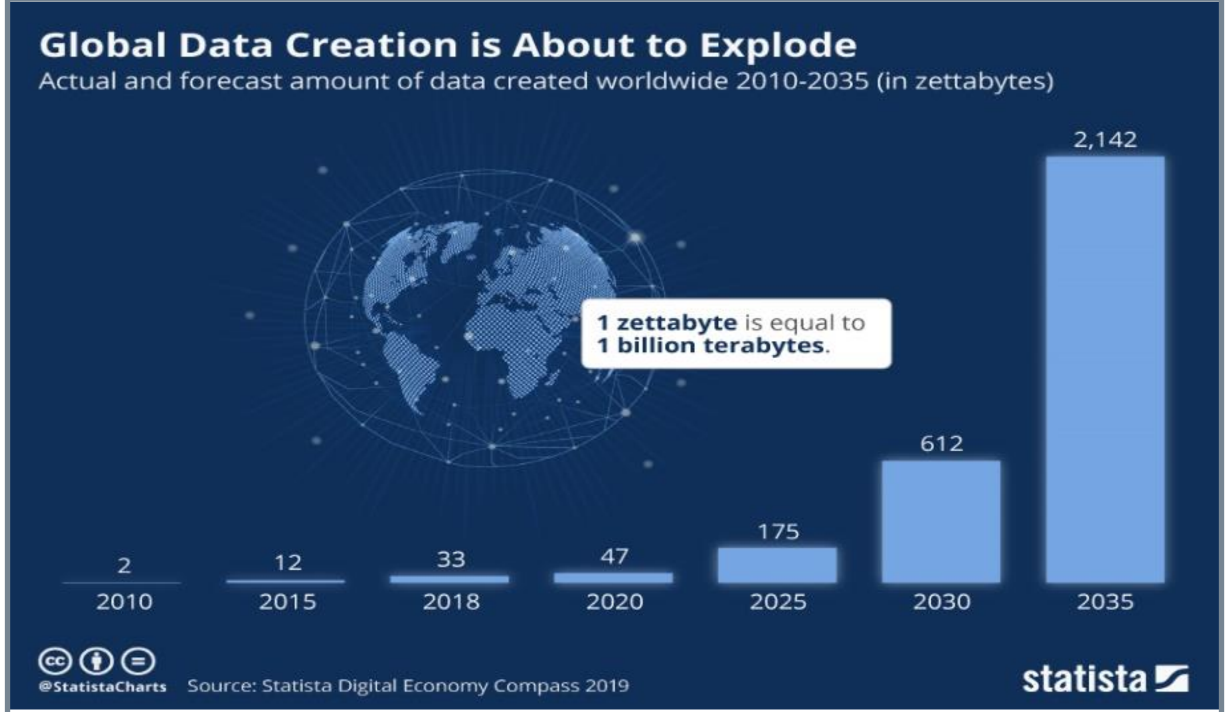


Importance of time series

Machine learning on time series is becoming increasingly important because of the massive production of time series data from diverse sources, e.g.,

- Digitalization in healthcare
- Internet of things
- Smart cities
- Process monitoring

The amount of created data increased from two zettabytes in 2010 to 47 zettabytes in 2020



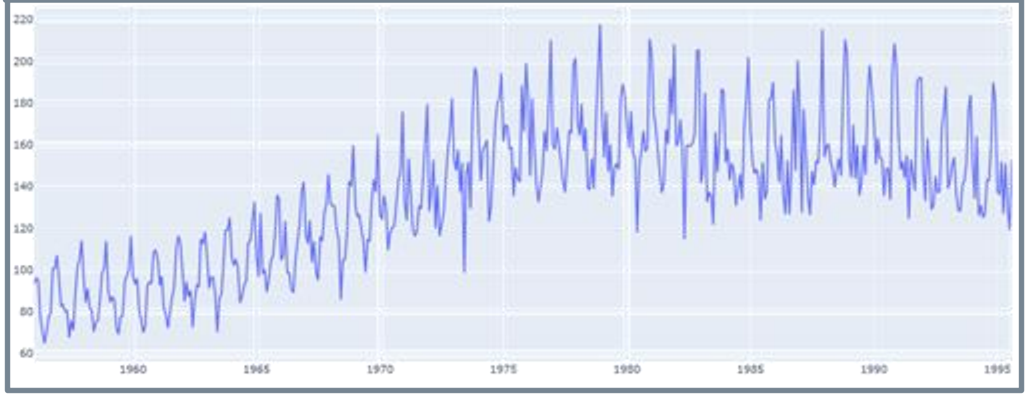
https://www.statista.com

Example: Predicting demand of amazon products

Amazon sells 400 million products in over 185 countries[a].

- Maintaining surplus inventory levels for every product is cost-prohibitive.
- Predict future demand of products

Some products are sold depending on the season



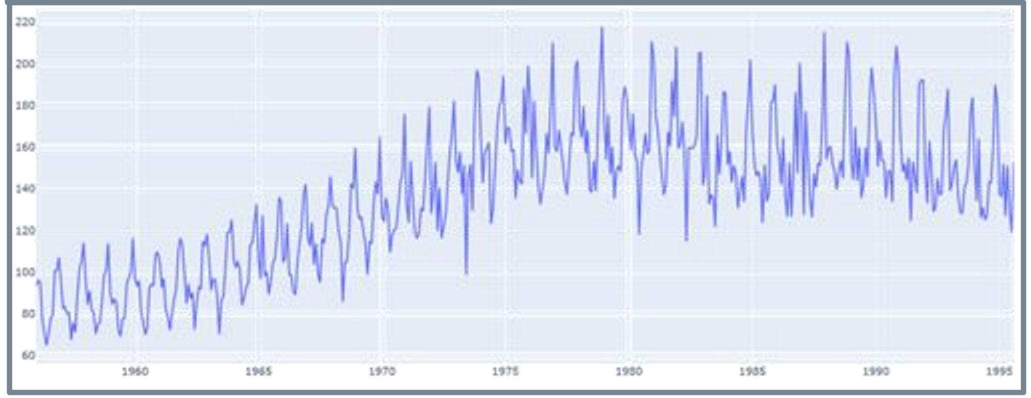
[a] https://www.amazon.science/latest-news/the-history-of-amazons-forecasting-algorithm

Example: Predicting demand of amazon products

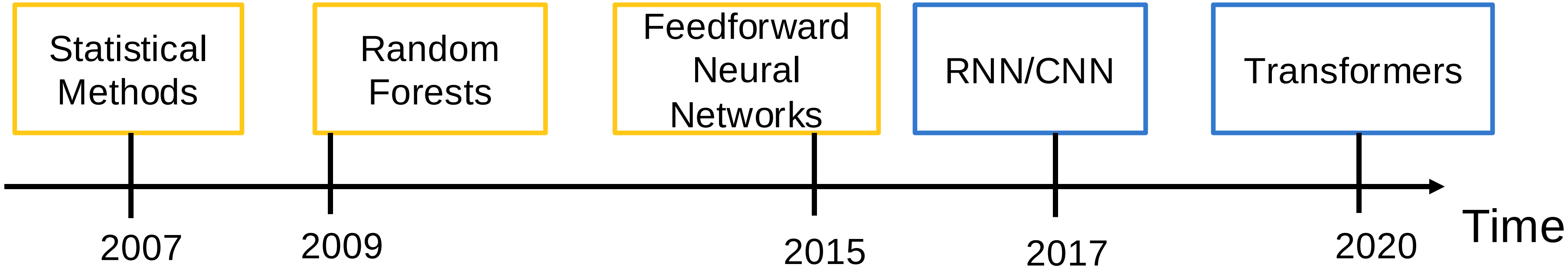
Amazon sells 400 million products in over 185 countries[a].

- ➢ Maintaining surplus inventory levels for every product is cost-prohibitive.
- ➢ Predict future demand of products

Some products are sold depending on the season



Methods:



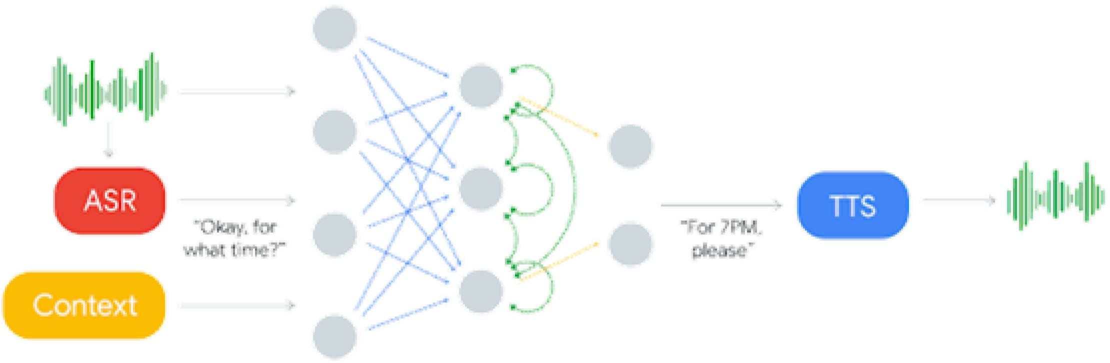
□ First models required manual feature engineering

[a] https://www.amazon.science/latest-news/the-history-of-amazons-forecasting-algorithm

Example: Google Duplex makes tedious phone calls

Long standing goal of making humans having a natural conversation with machines, as they would with each other.

- ➢ Carry out real-world tasks over the phone



- Additional audio features
- Automatic speech recognition
- Desired service, time/day

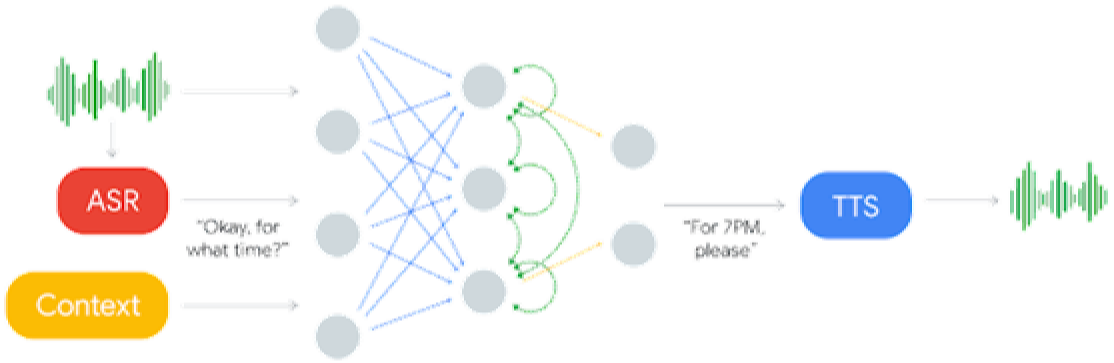
E.g., Duplex calling a restaurant.

https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html

Example: Google Duplex makes tedious phone calls

Method: An RNNs with several features. We use a combination of text to speech (TTS) engine and a synthesis TTS engine to control intonation (e.g., “hmm”s and “uh”s).

Limitations: trained on specific tasks. Cannot deal general conversations.



- Additional audio features
- Automatic speech recognition
- Desired service, time/day

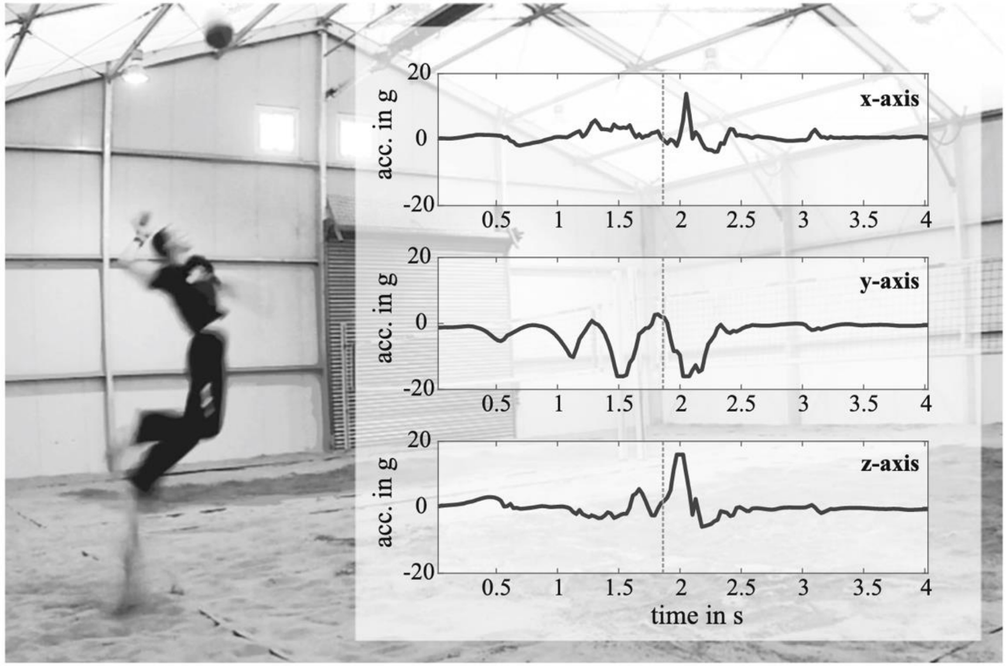
E.g., Duplex calling a restaurant.

Example: Activity recognition in sports (FAU Erlangen)

Many injuries in sports are caused by overuse.

- These injuries are a major cause for reduced performance of professional and non-professional beach volleyball players.
- Monitoring of player actions could help identifying and understanding risk factors and prevent such injuries.

Sensor attachment at the wrist of the dominant hand with a soft, thin wristband



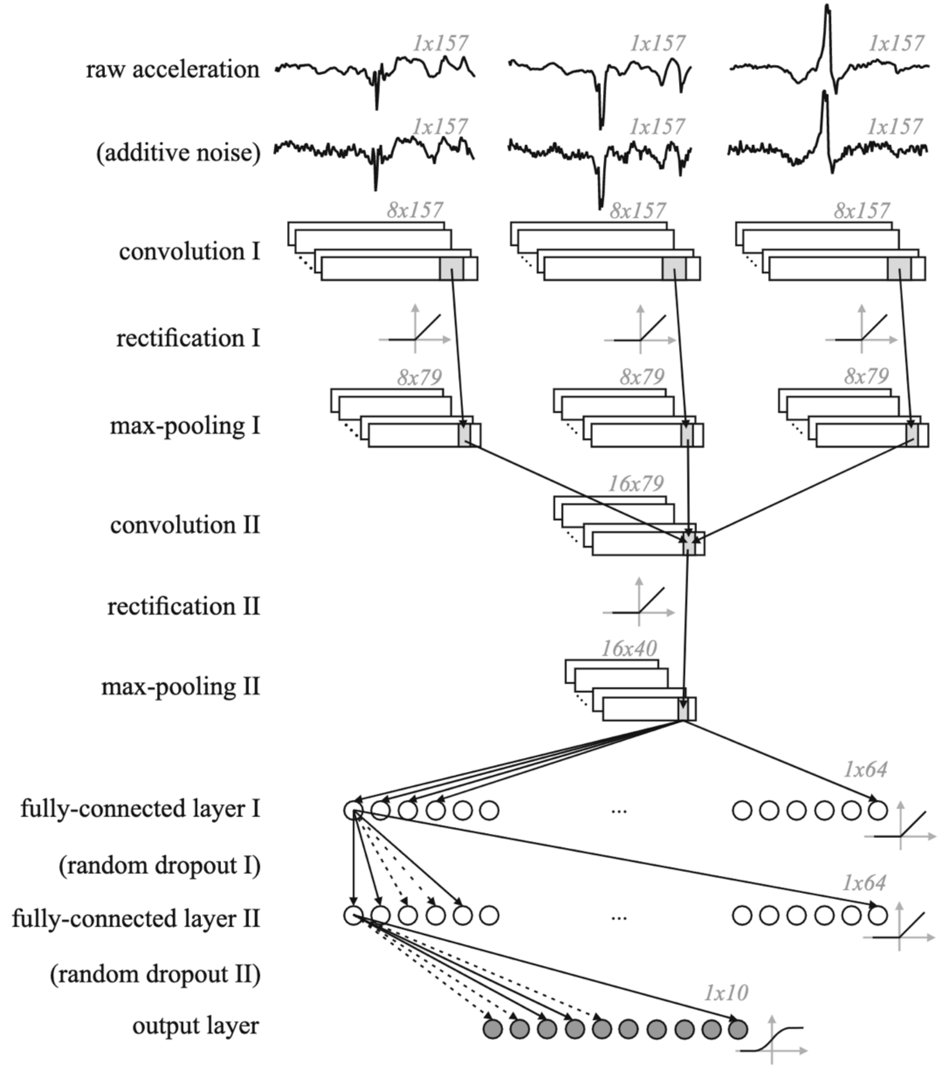
https://doi.org/10.1007/s10618-017-0495-0

Example: Activity recognition in sports (FAU Erlangen)

Method: A CNN is used to classify players' activities. Classifications allow to create players' profiles.

Actions:

- Underhand serve
- Overhand serve
- Jump serve
- Underarm set
- Overhead set
- Shot attack
- Spike
- Block
- Dig
- Null class.



https://doi.org/10.1007/s10618-017-0495-0

Time series fundamentals
Definitions and basic properties

What is a time series?

A time series can be described as a set of observations, taken sequentially in time,

$$S = \{s_1, \dots, s_T\}$$

where $s_i \in \mathbb{R}^d$ is the measured state of the observed process at time t_i.

Typically, observations are generally *dependent*

- Studying the nature of this dependency is of particular interest
- Time series analysis is concerned with techniques for the analysis of these dependencies

Examples of time series

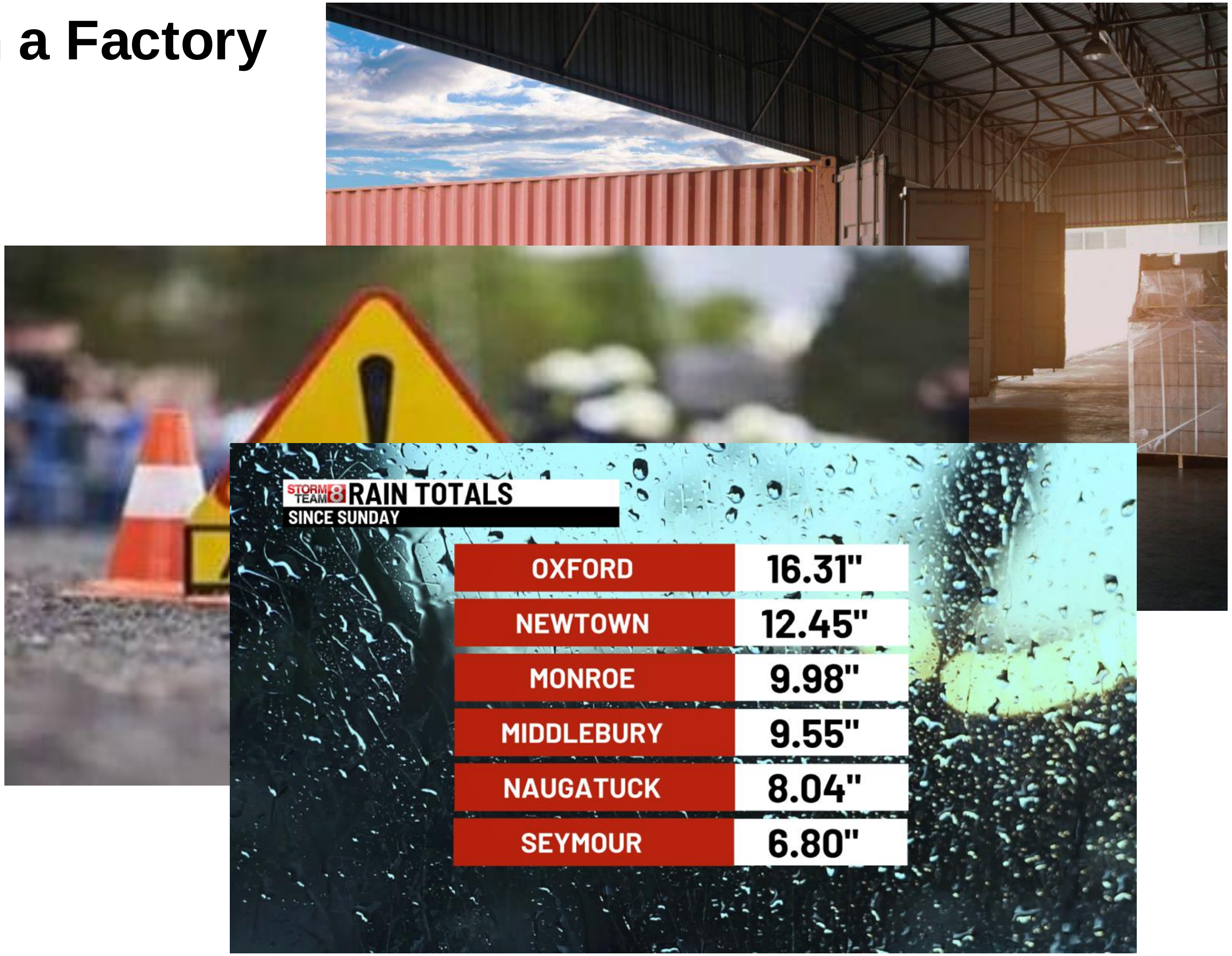
- **Monthly Goods Shipped from a Factory**

Examples of time series

- Monthly Goods Shipped from a Factory
- Weekly Road Accidents

Examples of time series

- Monthly Goods Shipped from a Factory
- Weekly Road Accidents
- Daily Rainfall Amounts
- …

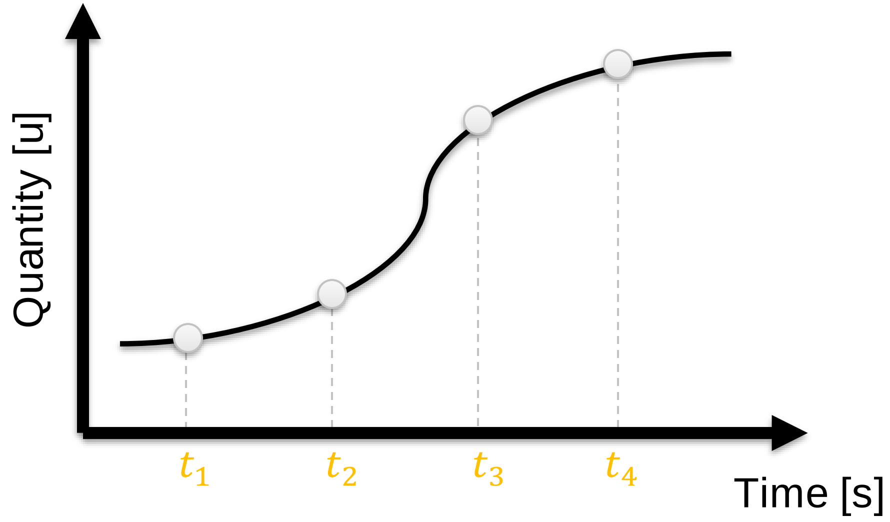


Terminology: Regularly Sampled vs Irregularly Sampled

Discrete time series are **regularly sampled** if their observations are euqally spaced in time.

$$\forall i \in \{1, \dots, T-1\},$$

$$\Delta_{t_i} = t_{i+1} - t_i = const.$$



Terminology: Regularly Sampled vs Irregularly Sampled

Discrete time series are **regularly sampled** if their observations are euqally spaced in time.

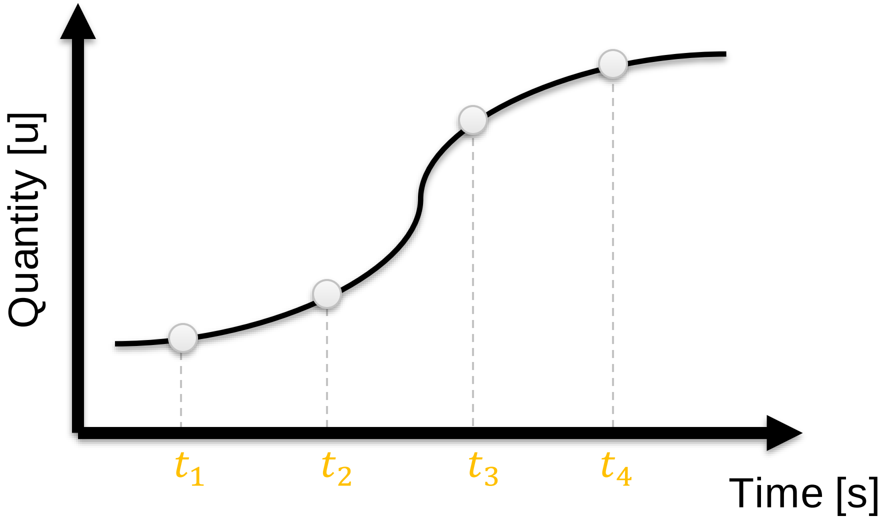
$$\forall i \in \{1, \ldots, T-1\},$$

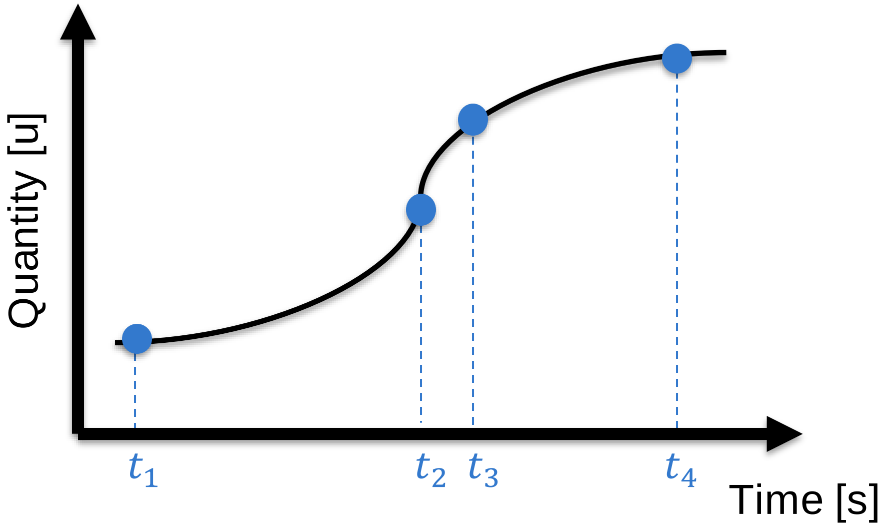
$$\Delta_{t_i} = t_{i+1} - t_i = const.$$

In contrast, for **irregularly sampled** time sequences, the observations are not equally spaced.

- They are generally defined as a collection of pairs

$$S = \{(s_1, t_1), \ldots, (s_T, t_T)\}$$



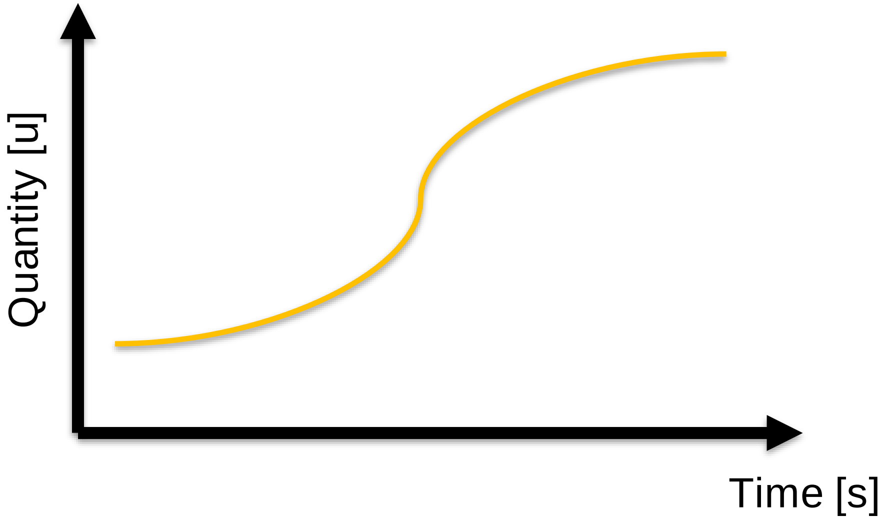


Terminology: Univariate vs Multivariate

Let $S = (s_1, \ldots, s_T)$ be a time series,
where $s_i \in \mathbb{R}^d, \forall i \in \{1, \ldots, T\}$.

If $d = 1$, S is said **univariate**.

- Only one variable is varying over time.



Time [s]

Terminology: Univariate vs Multivariate

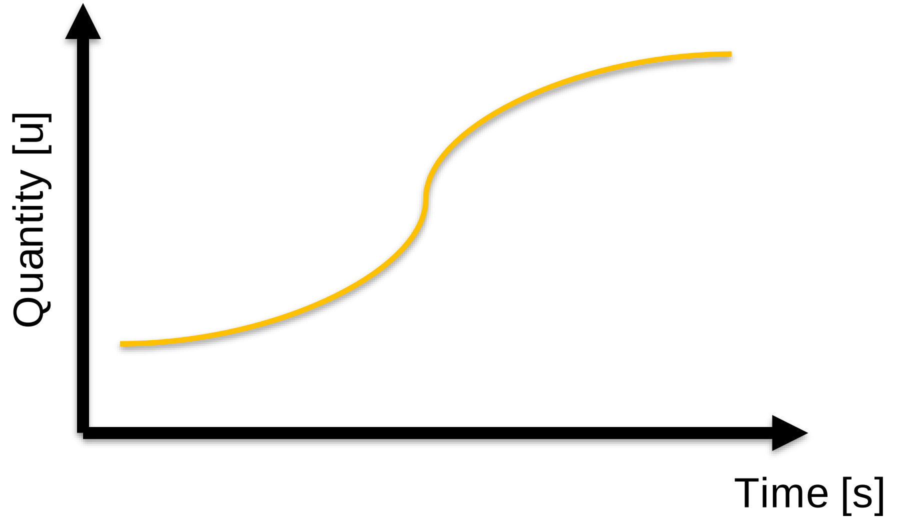
Let $S = (s_1, \ldots, s_T)$ be a time series,
where $s_i \in \mathbb{R}^d, \forall i \in \{1, \ldots, T\}$.

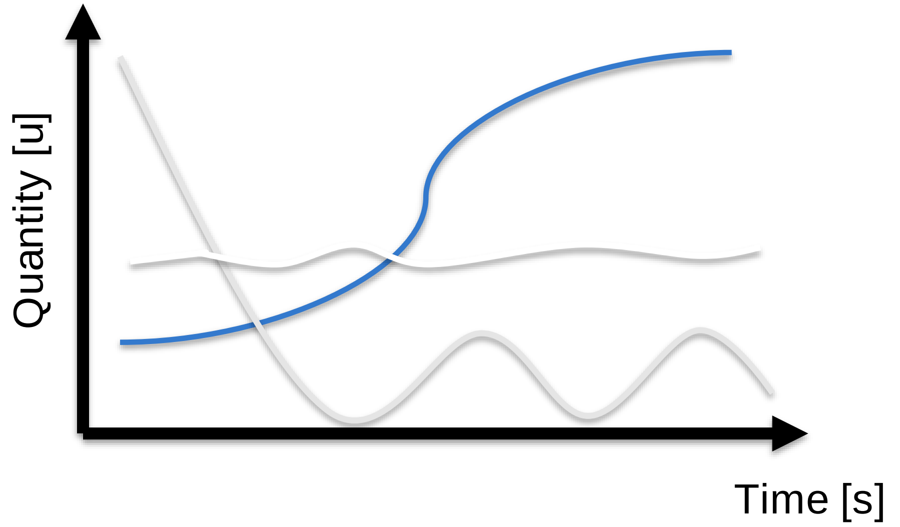
If $d = 1$, S is said **univariate**.

- Only one variable is varying over time.

If $d > 1$, S is said **multivariate**.

- Multiple variables are varying over time
  - E.g., tri-axial accelerometer measurements





Terminology: Discrete vs Continuous

A time series is **continuous** when observations are made continuously through time. The term continuous is used for series of this type even when the measured variables can take discrete set of values.

- E.g., the number of people in a room.

A time series is **discrete** when observations are taken only at specific times. The term discrete is used for series of this type even when the measured variables is a continuous variable.

- E.g, event logs.

Terminology: Discrete vs Continuous

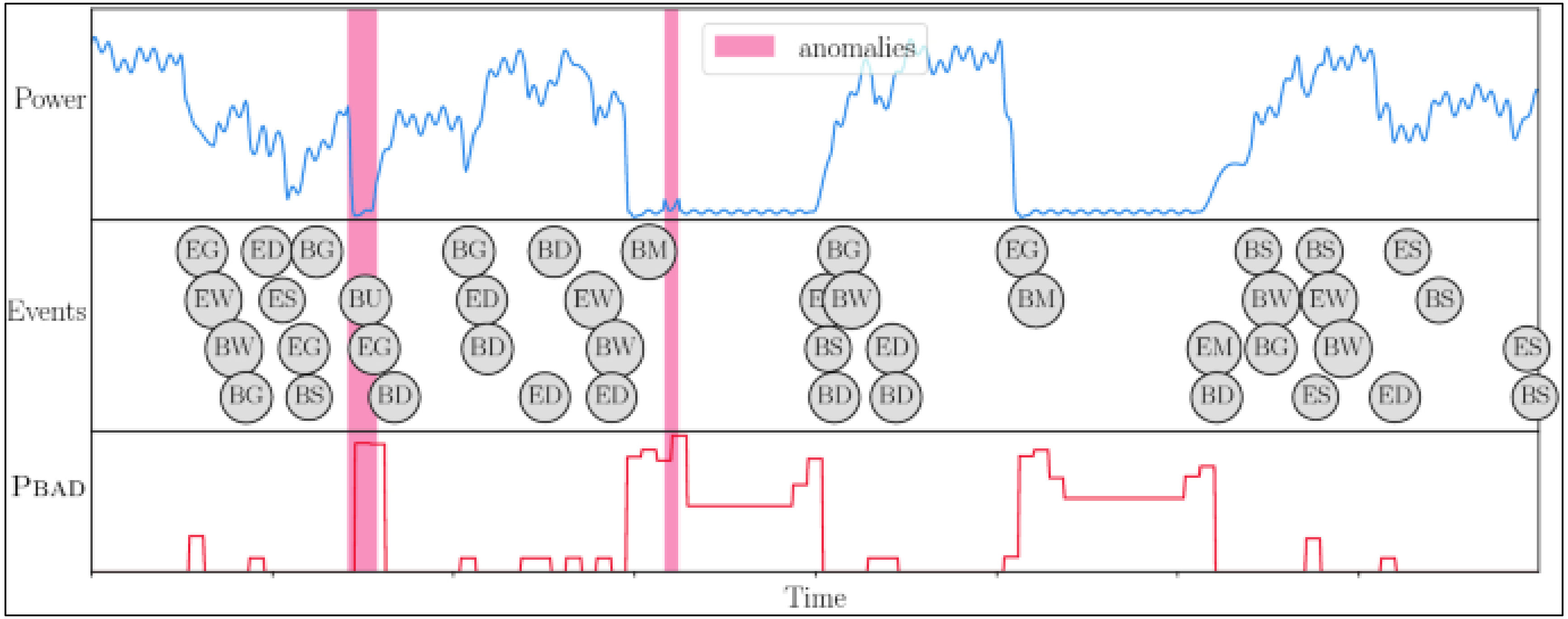
We will denote as **mixed-type** a multivariate time series consisting of both continuous and discrete observations

- E.g., a time series consisting of continuous sensor values and discrete event log for the monitoring of an industrial machine

Terminology: Discrete vs Continuous

We will denote as **mixed-type** a multivariate time series consisting of both continuous and discrete observations

- E.g., a time series consisting of continuous sensor values and discrete event log for the monitoring of an industrial machine



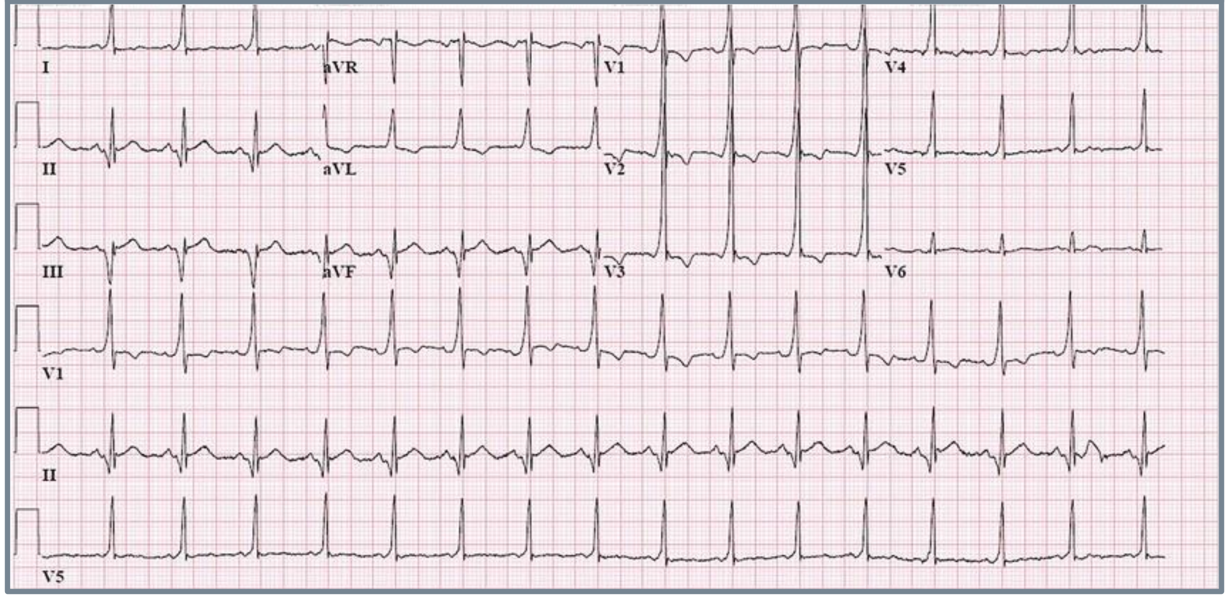
Terminology: Periodic

A time series is said **periodic** if there exists a number $\tau \in \mathbb{R}$, called *period*, such that

$$s_i = s_{i+\tau}\,, \forall i \in \{1, \ldots, T - \tau\}$$

E.g., the continuous time series defined by the trigonometric function $\mathrm{f(x)} = \sin(x)$

Is the biological signal of an heartbeat a periodic function?



Terminology: Deterministic vs Non-Deterministic

A **deterministic** time series is one that can be fully described by a known analytical expression or a set of rules. Observations are generated from a system that behaves predictably, with no element of randomness.

In contrast, a **non-deterministic** time series cannot be fully described by an analytical expression. A time series may be non-deterministic for the following reasons:

- The information necessary to describe the process is not fully observable, or
- The process generating the time series involves inherent randomness.

Stochastic Process

Non-deterministic time series can be regarded as manifestations (equiv., realization) of a **stochastic process**, which is defined as a set of random variables $\{X_t\}_{t\in\{1,\ldots,\mathrm{T}\}}$

Even if we were to imagine having observed the process for an infinite period T of time, the infinite sequence

$$S = \{\ldots, s_{t-1}, s_t, s_{t+1,\ldots}\} = \{s_t\}_{t=-\infty}^{+\infty}$$

would still be a single **realization** from that process.

Stochastic Process

Still, if we had a battery of N computers generating series $S^{(1)}$, …, $S^{(N)}$, and considering selecting the observation at time t from each series,

$$\left\{s_t^{(1)}, \ldots, s_t^{(N)}\right\}$$

this would be described as a sample of N realizations of the random variable X_t

Stochastic Process

Still, if we had a battery of N computers generating series $S^{(1)}$, …, $S^{(N)}$, and considering selecting the observation at time t from each series,

$$\left\{s_t^{(1)}, \dots, s_t^{(N)}\right\}$$

this would be described as a sample of N realizations of the random variable X_t

This random variable X_t is associated with an **unconditional density**, denoted by

$$f_{X_t}(s_t)$$

- E.g., for the Gaussian white noise process $f_{X_t}(s_t) = \frac{1}{\sqrt{2\,\pi\sigma}}\, e^{\frac{-s_t^2}{2\,\sigma^2}}$

Stochastic Process

The **unconditional mean** is the expectation, provided it exists, of the t-th observation, i.e.,

$$E(X_t) = \int_{-\infty}^{+\infty} s_t \, f_{X_t}(s_t) ds_t = \mu_t$$

Similarly, the **variance** of the random variable X_t is defined as

$$E(X_t - \mu_t)^2 = \int_{-\infty}^{+\infty} (s_t - \mu_t)^2 \, f_{X_t}(s_t) ds_t$$

Stochastic Process

Given any particular realization $S^{(i)}$ of a stochastic process (i.e., a time series), we can define the vector of the $j+1$ most recent observations

$$x_t^{\mathrm{i}} = [s_{t-j}^{(i)}, \ldots, s_t^{(i)}]$$

We want to know the probability distribution of this vector x_t^{i} across realizations. We can calculate the ***j*-th autocovariance**

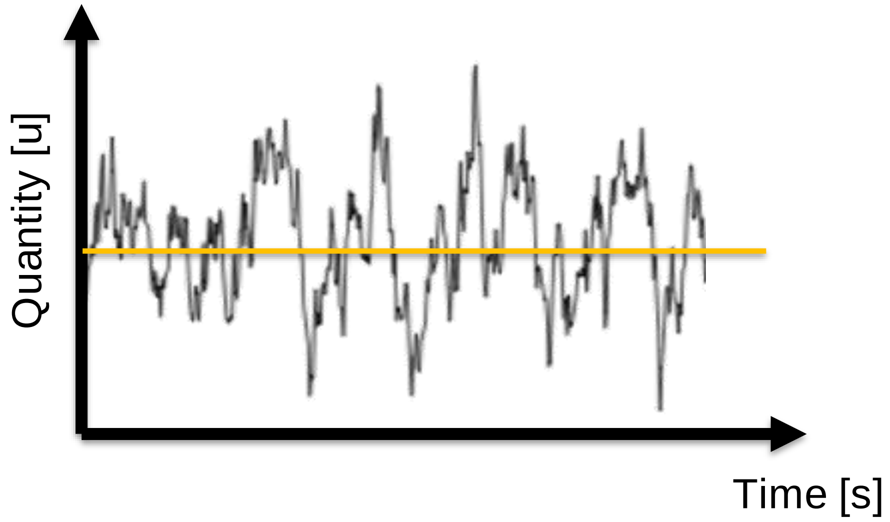
$$\gamma_{jt} = E(X_t - \mu_t)(X_{t-j} - \mu_{t-j})$$

Stationarity

If neither the mean μ_t or the autocovariance γ_{jt} depend on the temporal variable t, then the process is said to be (weakly) **stationary**.

E.g., let the stochastic process $\{X_t\}_{t=-\infty}^{+\infty}$ represent the sum of a constant μ with a Gaussian white noise process $\{\epsilon_t\}_{t=-\infty}^{+\infty}$, such that

$$X_t = \boldsymbol{\mu} + \epsilon_t$$



Then, its mean is constant: $E(X_t) = \mu + E(\epsilon_t) = \mu$

and its j-th autocovariance: $E(X_t - \mu)(X_{t-j} - \mu) = \gamma_j$

Stationarity

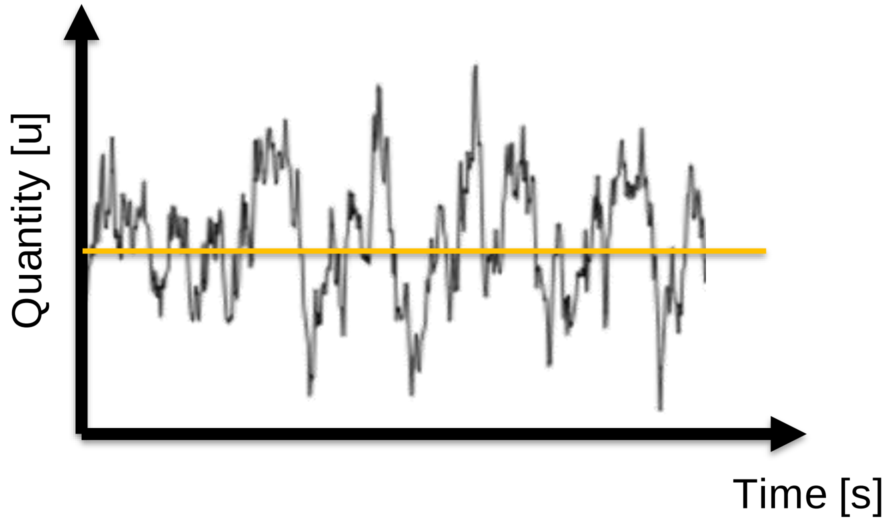
If neither the mean μ_t or the autocovariance γ_{jt} depend on the temporal variable t, then the process is said to be (weakly) **stationary**.

E.g., let the stochastic process $\{X_t\}_{t=-\infty}^{+\infty}$ represent the sum of a constant μ with a Gaussian white noise process $\{\epsilon_t\}_{t=-\infty}^{+\infty}$, such that

$$X_t = \boldsymbol{\mu} + \epsilon_t$$

Then, its mean is constant: $E(X_t) = \mu + E(\epsilon_t) = \mu$

and its j-th autocovariance: $E(X_t - \mu)(X_{t-j} - \mu) = \gamma_j$



In other words: A process is said to be stationary if the process statistics do not depend on time.

Ergodicity

Given a time series, denoted by $S^{(i)} = \left\{s_1^{(i)}, \ldots, s_T^{(i)}\right\}$, we can compute the sample temporal average as

$$\bar{s} = \frac{1}{T}\sum_{t=1}^{T} s_t^{(i)}$$

The time average of a single realization of the process converges to the ensemble average (or expected value) of the process as time goes to infinity:

- Ergodicity implies that (necessary condition) $\bar{s}$ converges to μ_t as $T \to \infty$

Ergodicity

Given a time series, denoted by $S^{(i)} = \left\{s_1^{(i)}, \dots, s_T^{(i)}\right\}$, we can compute the sample temporal average as

$$\bar{s} = \frac{1}{T}\sum_{t=1}^{T} s_t^{(i)}$$

The time average of a single realization of the process converges to the ensemble average (or expected value) of the process as time goes to infinity:

- Ergodicity implies that (necessary condition) $\bar{s}$ converges to μ_t as $T \to \infty$

In other words: A process is said to be ergodic if its time statistics equals the process statistic, provided that the process is observed long enough.

Example: Stationary but Non-Ergodic process

We give an example of stationary but not ergodic process.

Suppose the mean $\mu^{(i)}$ of the i-th realization of $\{X_t\}_{t=-\infty}^{+\infty}$ is sampled from the normal distribution $U(0, \lambda^2)$ and, similarly to the previous example, $X_t^{(i)} = \mu^{(i)} + \epsilon_t$.

We have that the process is stationary because:

$$\mu_t \ = E\left(\mu^{(i)}\right) + E(\epsilon_t) = 0$$

$$\gamma_{jt} \ = E\left(\mu^{(i)} + \epsilon_t\right)\left(\mu^{(i)} + \epsilon_{t-j}\right) = \lambda^2$$

Example: Stationary but Non-Ergodic process

However, its sample temporal mean, converges to a different value than the process mean, i.e.,

$$\bar{s} = (1/T)\sum(\mu^{(i)}+\epsilon_t) = \mu^{(i)}$$

Time series fundamentals

i.i.d. observations and central limit theorems

Time series and i.i.d. data

Observations collected in a time series $S = (s_1, \ldots, s_T)$ are **generally not i.i.d.**

- Observation s_i could be **dependent** on previous observations s_j, with $j < i$
- The distribution of the underlying data generation process could change over time, i.e. it is **not identically distributed**

Time series and i.i.d. data

Observations collected in a time series $S = (s_1, \dots, s_T)$ are **<u>generally not i.i.d.</u>**

- Observation s_i could be **dependent** on previous observations s_j, with $j < i$
- The distribution of the underlying data generation process could change over time, i.e. it is **not identically distributed**

For example:

- The price of a stock today depends on its price yesterday **(dependence)**
- and the volatility of the stock, i.e., its dispersion of returns, might change over time **(change on the underlying distribution)**

Time series and i.i.d. data

The structure of this dependence imposes challenges on the statistical data analysis of time series.

- Many tools for statistical inference are valid only for i.i.d. data

Time series and i.i.d. data

It might be useful to be able to assess the structure of the dependence between random variables. For this reason we make use of their correlation.

- Generally, we measure the correlation between two variables X_i and X_j with their **covariance** $Cov(X_i, X_j)$.
  - $Cov(X_i, X_j) = 0 \rightarrow \text{uncorrelated}$
- We measure dependence of an entire time series with a similar concept, the **long-run variance**
  - $\sigma_i^2 = \sum_{\mathbb{Z}} Cov(X_i, X_{i+h})$
  - “ the sum of all autocovariances“

The Central Limit Theorem

The **Central Limit Theorem (CLT)** suggests that the sum of random variables converges to a normal distribution, under precise conditions.

More precisely, for a sequence of i.i.d. random variables $\{X_t\}_{t\in\{1,\dots,\mathrm{T}\}}$ with $\mu = E(X_t)$ and $\sigma^2 = E(X_t - \mu)^2$, by the CLT it holds:

$$\sqrt{T}\left(\frac{1}{T}\sum_{1}^{T} X_i - \mu\right) \rightarrow \mathcal{N}(0, \sigma^2)$$

The Central Limit Theorem

The **Central Limit Theorem (CLT)** suggests that the sum of random variables converges to a normal distribution, under precise conditions.

More precisely, for a sequence of i.i.d. random variables $\{X_t\}_{t\in\{1,\dots,\mathrm{T}\}}$ with $\mu = E(X_t)$ and $\sigma^2 = E(X_t - \mu)^2$, by the CLT it holds:

$$\sqrt{T}\left(\frac{1}{T}\sum_{1}^{T} X_i - \mu\right) \rightarrow \mathcal{N}(0, \sigma^2)$$

-> For stationary time series with mean μ and finite $\text{long}-\text{run variance}\ \sigma^2$, the CLT holds as before.

Why is the CLT important?

If the CLT holds for a time series, we can draw from a larger range of methods.

- Statistical inference depends on the possibility to take a broad view of results from a sample to the population.
- E.g., the CLT legitimizes the assumption of normality of the error terms in linear regression.

However,

- Many time series we encounter in the real world satisfy CLT assumption of independence and stationarity
- Or can be transformed into stationary time series, e.g., by differentiations or other transformations

Why is the CLT important?

If the CLT holds for a time series, we can draw from a larger range of methods.

- Statistical inference depends on the possibility to take a broad view of results from a sample to the population.
- The CLT legitimizes the assumption of normality of the error terms in linear regression.

However,

- Many time series we encounter in the real world satisfy CLT assumption of independence and stationarity
- Or can be transformed into stationary time series, e.g., by differentiations or other transformations

It is a good idea to start by checking whether the data is independent or stationary.

Insight: CLT for dependent random variables

Different version of the CLT exist for dependent random variables. For example, under the assumption of a M-dependent random process[a], we have that the following limit theorem holds:

Let $\{X_t\}_{t\in\{1,\dots,T\}}$ be M-dependent stationary process with mean μ, covariance γ_j, and denoted with V_M,

$$V_M := \sum_{j=-M}^{M} \gamma_j$$

If $V_M > 0$, then,

$$\sqrt{n}(X_i - \mu) \rightarrow N(0, V_M).$$

[a] A stochastic process $\{X_t\}_{t\in\{1,\dots,\mathrm{T}\}}$ is said to be M-dependent if $\{X_t\}_{t\leq k}$ are independent of the stochastic variables $\{X_t\}_{t\geq k+M+1}$

Time series fundamentals

Recap

Recap

Time series have long been studied in history

- Recent digitalization increases the importance of time series analysis

Recap

Time series have long been studied in history

- Recent digitalization increases the importance of time series analysis

Properties of time series

- Regularly vs irregularly sampled
- Univariate vs multivariate
- Discrete vs continuous
- Periodic
- Deterministic vs non-deterministic
- Stationarity
- Ergodicity

Recap

Time series have long been studied in history

- Recent digitalization increases the importance of time series analysis

Properties of time series

- Regularly vs irregularly sampled
- Univariate vs multivariate
- Discrete vs continuous
- Periodic
- Deterministic vs non-deterministic
- Stationarity
- Ergodicity

The Central Limit Theorem only holds for stationary time series

- Less restrictive versions of the CLT exist
- Need to properly learn temporal dependencies