13.06.23

OPENCAMPUS.sh

Einführung in Data Science und maschinelles Lernen

ZEITREIHENANALYSEN

- Fehlende Werte
- Muster in Zeitreihenanalysen
- Non-Stationarity
- Baseline Modelle und Naïve Forecasting
- Projektpräsentation
- *Und jetzt?*

IMPUTATION EXAMPLES

```
---
title: "Missing Values"
output: html_notebook
---

```{r}
# Prepare Environment
library(readr)
library(VIM)
library(dplyr)
library(ggplot2)

options(datatable.use.index = TRUE)

```

```{r}
# VIM Aggregation Plot
sleep %>%
  aggr(combined=TRUE, numbers=TRUE)

```


```{r}
# Listwise deletion
sleep_deletion <- na.omit(sleep)

sleep_deletion %>%
  aggr(combined=TRUE, numbers=TRUE)

```


```{r}
# VIM Hot-Deck Imputation

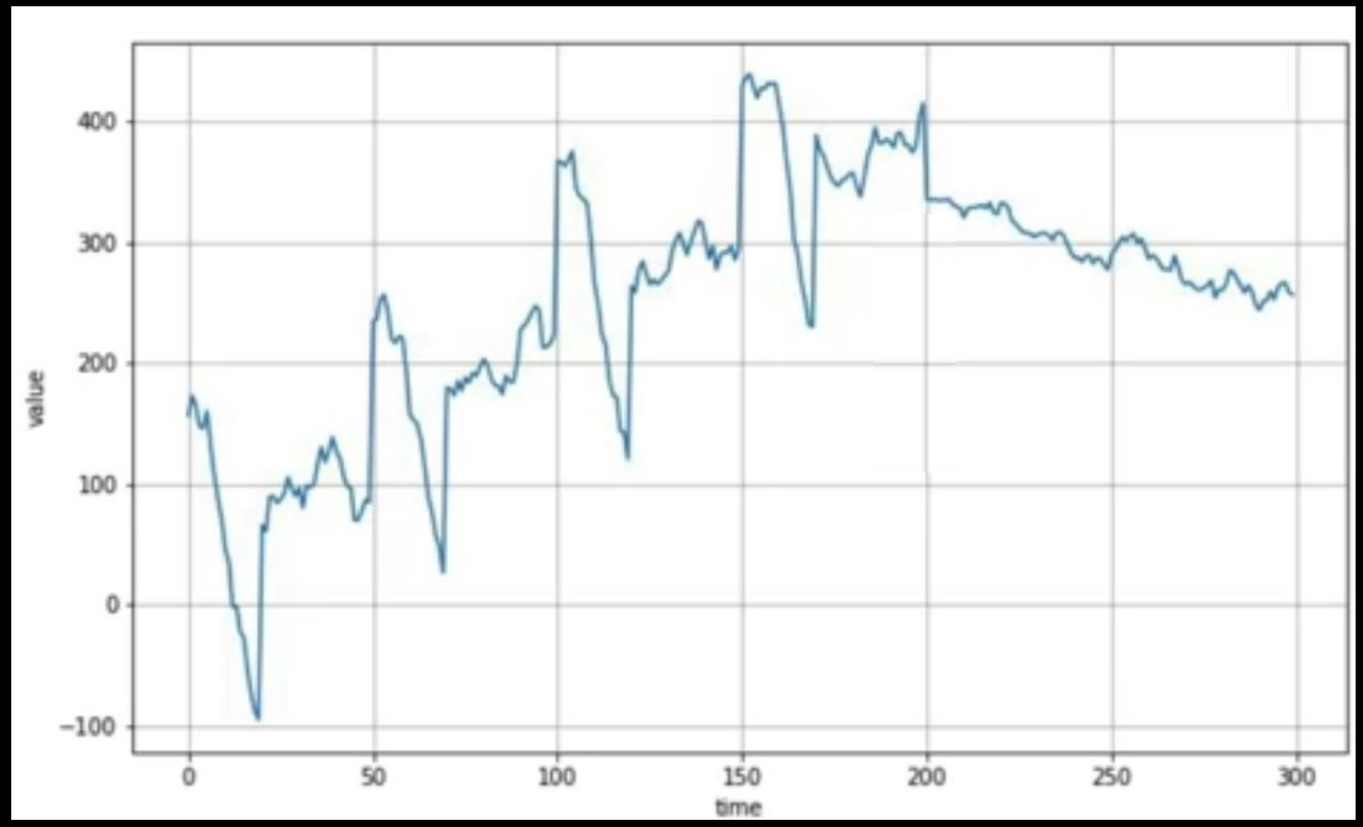
sleep_hotdeck1 <- sleep %>%
  hotdeck()
sleep_hotdeck1 %>%
  aggr(combined=TRUE, numbers=TRUE)
ggplot(sleep_hotdeck1) +
  geom_point(aes(x=Sleep, y=Dream, color=Sleep_imp))
```
```

IMPUTATION BEI VIELEN FEHLENDEN WERTEN

- **Berücksichtigung der Indikator-Variable!**
- **Ggf. Vergleich unterschiedlicher Imputationsverfahren anhand unterschiedlicher Splits in Trainings- und Validierungsdatensatz.**

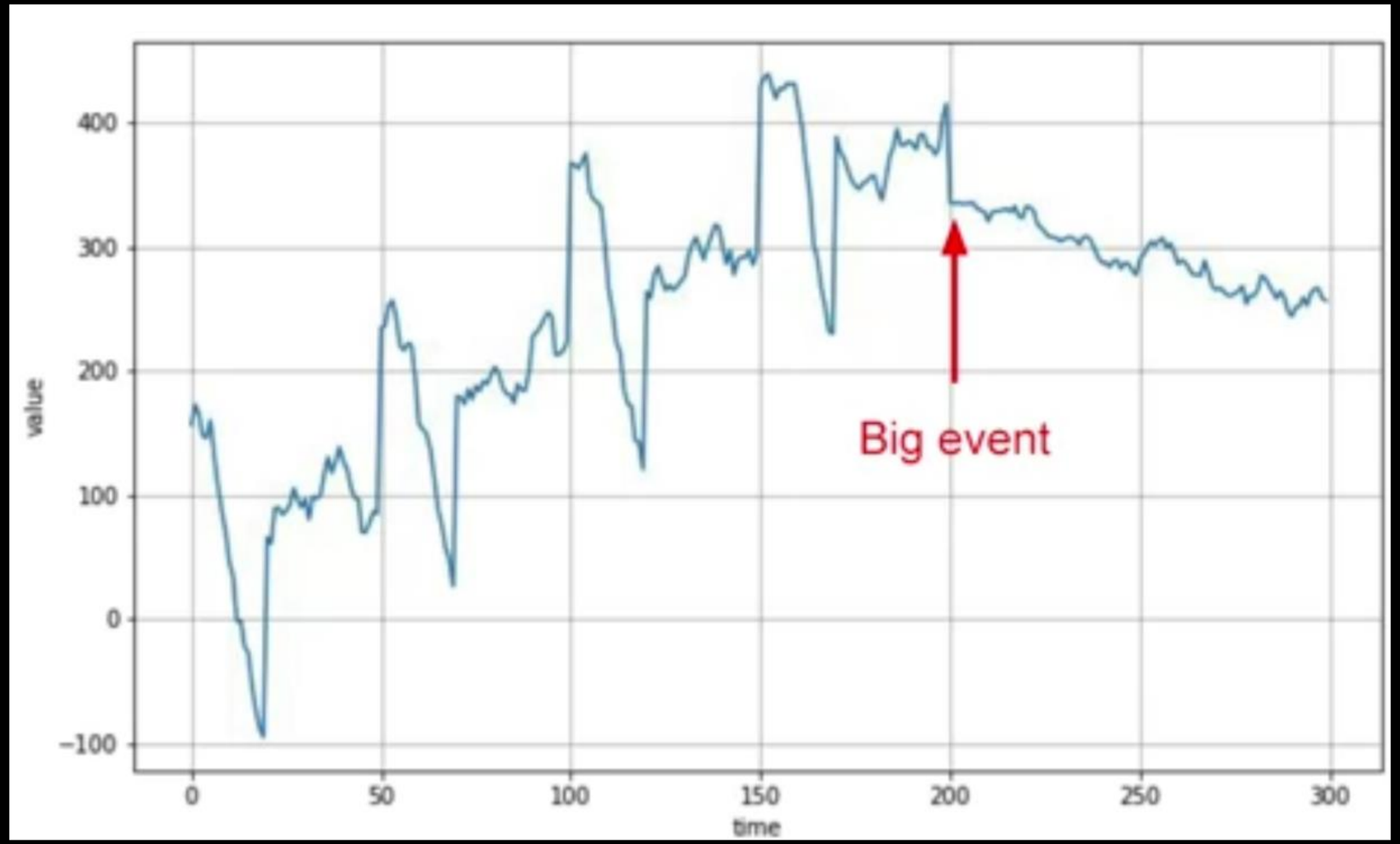
DISKUSSION

Welche verschiedenen Arten von Mustern kann man in der dargestellten Abbildung erkennen?



DeepLearning.AI. (o. J.). Common patterns in time series—Sequences and Prediction. Abgerufen 7. Januar 2021, von Coursera website: https://www.coursera.org/lecture/tensorflow-sequences-time-series-and-prediction/common-patterns-in-time-series-dWnOy

NON-STATIONARITY

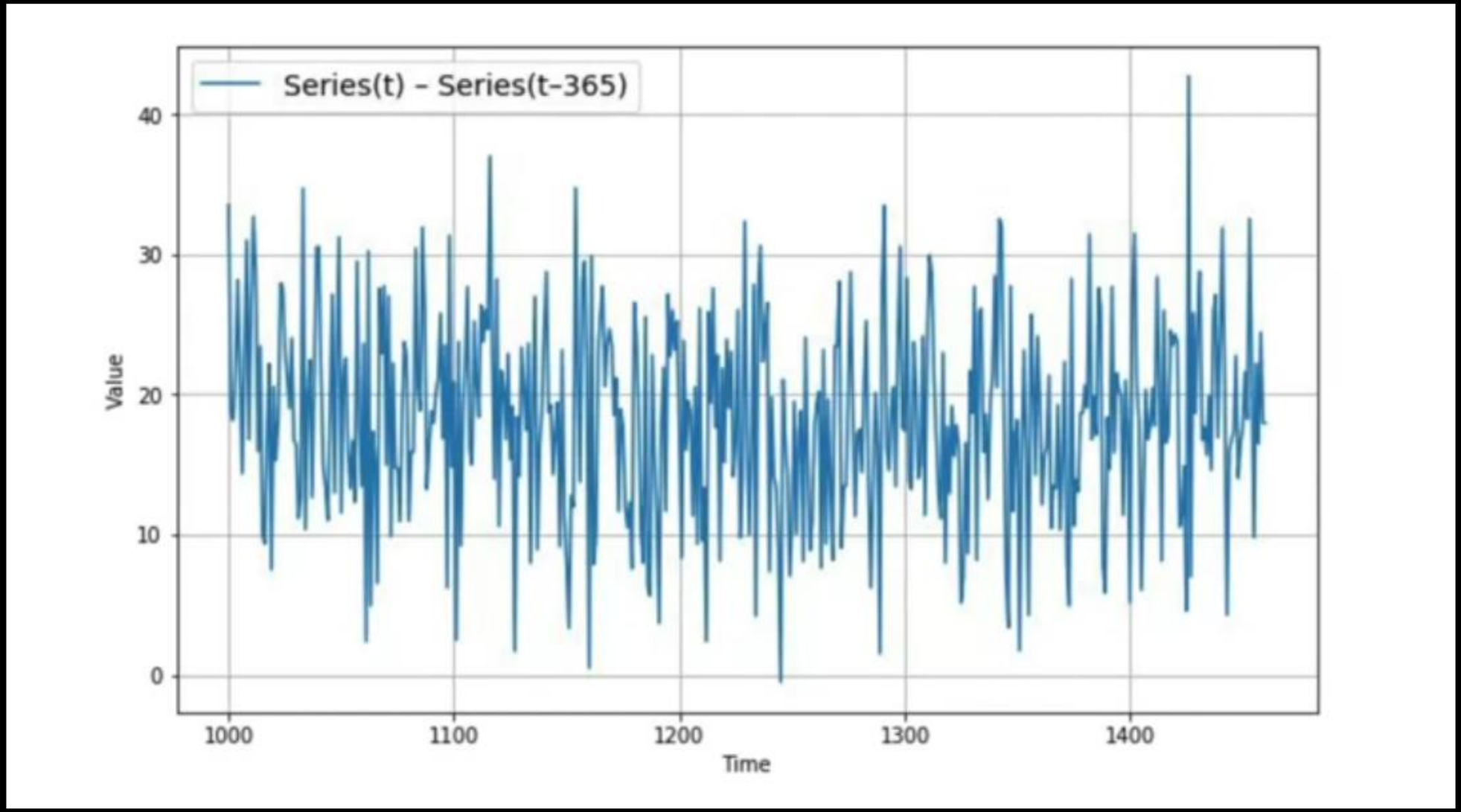


DeepLearning.AI. (o. J.). Common patterns in time series—Sequences and Prediction. Abgerufen 7. Januar 2021, von Coursera website: https://www.coursera.org/lecture/tensorflow-sequences-time-series-and-prediction/common-patterns-in-time-series-dWnOy

MUSTER IN ZEITREIHENANALYSEN

- Trends
- „Jahresgang“ (Saisonality)
- Rauschen (Noise)
- Autokorrelation (serielle Korrelation)

DIFFERENCING



DIFFERENCING

- Subtraktion der vorherigen Beobachtung von der aktuellen Beobachtung
- Methode zum Entfernen bzw. Mindern von Autokorrelation in Zeitreihen.
- Versuch „stationäre“ Zeitreihen zu erhalten

LAG-DIFFERENZ

- Die Differenz zwischen aufeinanderfolgenden Beobachtungen wird als Lag-1-Differenz bezeichnet.

- Die Lag-Differenz kann an die spezifische zeitliche Struktur angepasst werden.

- Bei Zeitreihen mit einer saisonalen Komponente kann man davon ausgehen, dass die Verzögerung der Periode (Breite) der Saisonalität entspricht.

BEISPIELAUSWERTUNGEN

```r

##################################################
### Preparation of the Environment
##################################################
### Reading the data file
##################################################
##################################################
### Prepare data
##################################################
##################################################
# Pedestrians hourly

# Basic plot
ggplot(pedestrians_hourly) +
  geom_line(aes(x=datetime, y=`pedestrians count`), color="#69b3a2") +
  xlab("") +
  theme_ipsum() +
  theme(axis.text.x=element_text(angle=60, hjust=1))

# Time frame specific plot
ggplot(pedestrians_hourly) +
  geom_line(aes(x=datetime, y=`pedestrians count`), color="#69b3a2") +
  xlab("") +
  theme_ipsum() +
  theme(axis.text.x=element_text(angle=60, hjust=1)) +
  scale_x_datetime(limit=c(as.POSIXct("2021-10-01"),as.POSIXct("2021-11-01")))


##################################################
# Pedestrians daily

# Basic plot
ggplot(pedestrians_daily) +
  geom_line(aes(x=date, y=`pedestrians count`), color="#69b3a2") +
  xlab("") +
```

BASELINE MODELLE

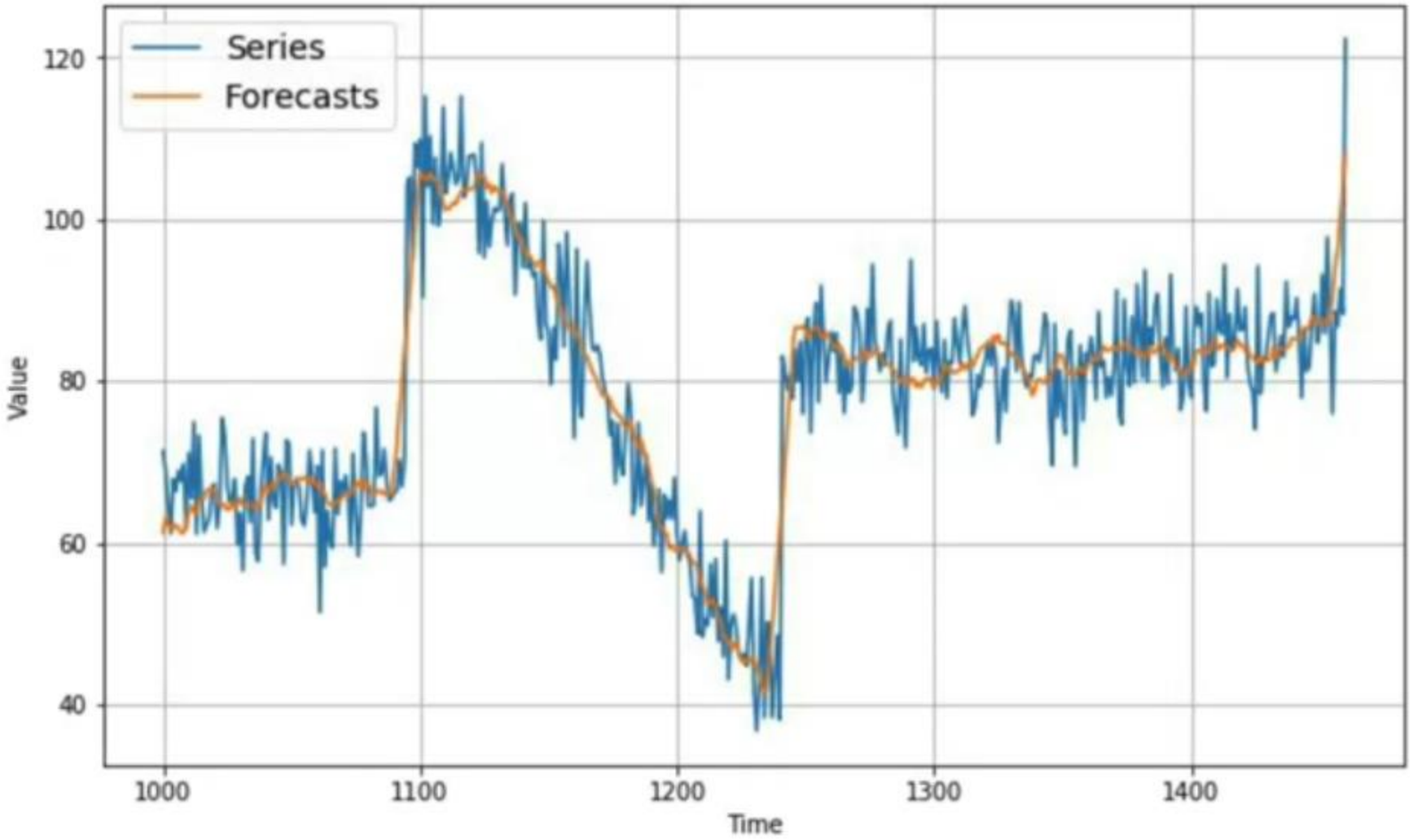
- Allein an Metriken wie MAPE oder RMSE kann man häufig schlecht abschätzen, wie viel das eigene Modell gelernt hat.
- Es ist daher wichtig, die Ergebnisse anderer Modelle als „Baseline“ bzw. Referenz zu nutzen.

Mögliche Baselines:

- Ergebnisse bisher genutzter Modelle für den gleichen Datensatz
- Ergebnisse von Modellen auf artverwandten Datensätzen
- Speziell bei Zeitreihen: Ergebnisse basierend auf Naïve Forecasting

NAÏVE FORECASTING

- **Mögliches Baseline Modell für Zeitreihenanalysen**
- **Vorhersage entspricht dem jeweils letzten beobachteten Wert**
- **Saisonal Naïve Forecasting:**
  Vorhersage entspricht dem letzten Wert mit der gleichen Saisonalität.



Forecasts = trailing moving average of differenced series + centered moving average of past series (t – 365)

BEISPIELBERECHNUNG VON LAG-DATEN

```r
library(ggplot2)
library(dplyr)

# Create some example data
ts_data <- data.frame(date = seq(from = as.Date("2022-01-01"), to = as.Date("2022-12-31"), by = "day"),
                      product = 1,
                      value = rnorm(365, mean = 100, sd = 10))

# Plot the data using ggplot
ggplot(ts_data, aes(x = date, y = value)) +
  geom_line() +
  ggtitle("Time Series Data") +
  xlab("Date") +
  ylab("Value")

# Add variable including the value of the day before
ts_data_with_lag <- ts_data %>%
  arrange(date) %>%
  mutate(value_prev_day = lag(value, default = NA))



# Example data, in which several values (labels) are given for each day
multiple_ts_data <- ts_data %>%
  rbind(data.frame(date = seq(from = as.Date("2022-01-01"), to = as.Date("2022-12-31"), by = "day"),
                      product=2,
                      value = rnorm(365, mean = 100, sd = 10)))

multiple_ts_data_with_lag <- multiple_ts_data %>%
  arrange(date, product) %>%
  group_by(product) %>%
  mutate(value_prev_within_day = lag(value, default = NA)) %>%
  ungroup()

```

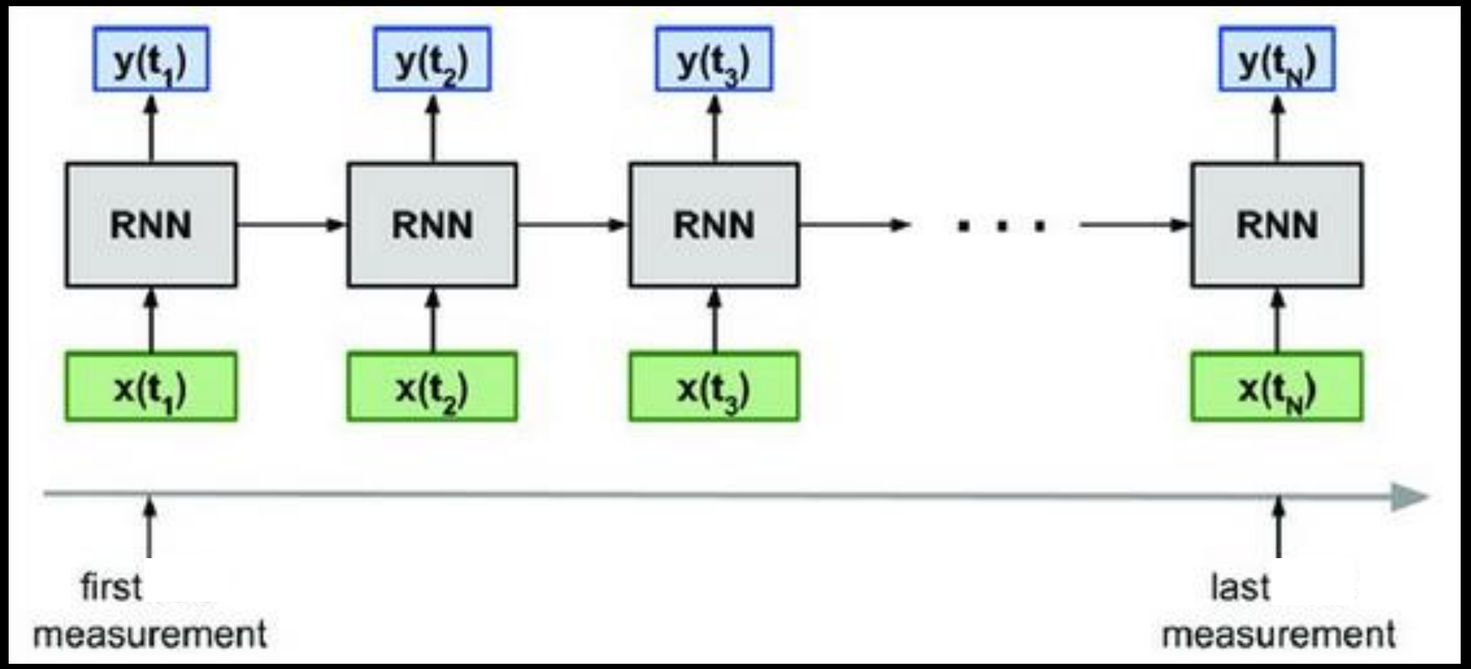
DISKUSSION

In der Praxis will man häufig nicht ein Jahr, sondern z.B. nur den nächsten Tag vorhersagen.

- **Welche Informationen aus der Zeitreihe wären zusätzlich besonders hilfreich?**

RECURRENT NEURAL NET (RNN)

- In einem RNN Layer sind die Knoten einer Schicht untereinander verknüpft.



INHALT DER PROJEKTPRÄSENTATION

- Eure Namen auf der Titelseite
- Auflistung und kurze Beschreibung der selbst erstellten Variablen
- Balkendiagramme mit Konfidenzintervallen für zwei selbst erstellte Variablen
- Optimierung des linearen Modells: Modellgleichung und adjusted r^2
- Art der Missing Value Imputation
- Optimierung des neuronalen Netzes:
  - Source Code zur Definition des neuronalen Netzes
  - Darstellung der Loss-Funktionen für Trainings- und Validierungsdatensatz
  - MAPEs für den Validierungsdatensatz insgesamt und für jede Warengruppe einzeln
- „Worst Fail“

HINWEISE

- **Dauer der Präsentation: ca. 8 Minuten pro Team**
- **Powerpoint, Keynote oder ähnliches (ggf. aus R-Markdown generiert)**
- **Zur Präsentation anhand Eures besten Modells die Vorhersagen für den Testdatensatz der Kaggle Competition berechnen und dort hochladen.**

Bis spätestens zum 1. März:

- **Die Präsentation zu Eurem Repo hinzufügen**
- **Das Repo wie in den READMEs beschrieben vervollständigen und das main README wie hier beschrieben in der EduHub-Plattform hochladen**

WIE KANN ICH WEITER MACHEN?

the decoder

EN

KI in der Praxis 6. Juni 2023

Apple spricht nicht über KI, packt sie aber trotzdem überall rein

KI-Forschung | KI und Gesellschaft | KI in der Praxis

DEEP MINDS

DEEP MINDS Podcast
Podcast über Künstliche Intelligenz und Wissenschaft

Künstliche Intelligenz und Robotik | DEEP MINDS #15

Verfügbar bei Youtube, Soundcloud, Spotify, Apple, Google und Amazon

R and OOP anti-patterns

June 9, 2023 | Bob Carpenter

Thomas Lumley just dropped a blog post, Blank cheque inheritance and statistical objects, which begins as follows. One of the problems with object-oriented programming for statistical methods is that inheritance is backwards. Everything is fine for data structures, and Bioconductor ... Continue reading → [Read more...]

'Advanced Shiny Development' the hands-on workshop

June 9, 2023 | Mirai Solutions

Best practices for a robust and maintainable shiny app, a hand on workshop on 21/06. Do you know how to build a basic Shiny web application, but would you like to bring your Shiny development to the next level? Learn from professional experts how to...
[Read more...]

Building a basic Shiny app with Golem – Part I (Video)

June 7, 2023 | pacha.dev/blog

R and Shiny Training: If you find this blog to be interesting, please note that I offer personalized and group-based training sessions that may be reserved through Buy me a Coffee. Additionally, I provide training services in the Spanish language ... [Read more...]

Understanding the file.info() Function in R: Listing Files by

Most viewed posts (weekly)

How to improve your storytelling with R
PCA vs Autoencoders for Dimensionality Reduction
How to install (and update!) R and RStudio
Update to Data Science Software Popularity
Why GLMs should be a priority when teaching statistics

Data Elixir

Stay up to date in Data Science.

Get the Data Elixir newsletter for a weekly dose of the top data science picks from around the web. Covering machine learning, data visualization, analytics, and strategy.

Sign up for Free

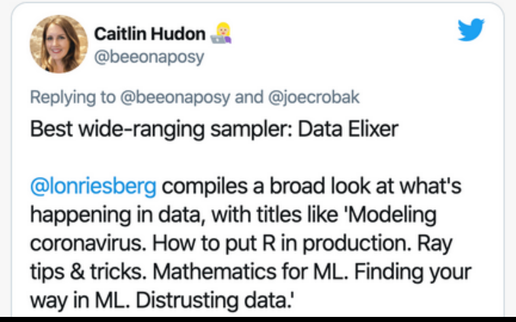
and join more than **55,000** data lovers today.

Email

SIGN UP

No spam, ever.

What readers say...



Machine Learning

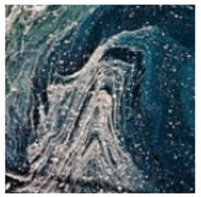
Follow Start writing

Trending Latest Best

Merve Noyan · 17 hours ago

Complete Guide on Deep Learning Architectures Part 2: Autoencoders

Autoencoder: Basic Ideas Autoencoder is the type of a neural network that reconstructs an input from the output. The basic idea here is tha...



Machine Learning 5 min read

Bex T. in Towards Data Science · 18 hours ago Member-only

10 Confusing XGBoost Hyperparameters and How to Tune Them Like a Pro in 2023

XGBoost hyperparameters done with style and visuals — Intro Today, I

251K Stories

92K Writers

Related Topics

Data Science | Artificial Intelligence | Deep Learning | Python | AI | Technology | Programming | NLP | Neural Networks

See more topics

Top Writers

The PyCoach
10M+ Views on Medium ||
Make money by writing abou...
Follow

Machine Learning

Beigetreten

r/MachineLearning

Beiträge

Beitrag erstellen

Heiß Neu Top

VON EINEM MOD ANGEHEFTET

6

Gepostet von u/AutoModerator vor 1 Tag

Discussion [D] Simple Questions Thread

20 Kommentare Auszeichnen Teilen Merken

4

Gepostet von u/ML_WAYR_bot vor 22 Stunden

Discussion [D] Machine Learning - WAYR (What Are You Reading) - Week 128

1 Kommentar Auszeichnen Teilen Merken

120

Gepostet von u/jayalammar vor 5 Stunden

Research [R] The Illustrated Retrieval Transformer (GPT3 performance at 4% the size)

Hi r/MachineLearning,

Über diese Community

Willkommen in MachineLearning

2.2m Mitglieder

488 Online

Am 29. Juli 2009 erstellt

Beitrag erstellen

COMMUNITY-EINSTELLUNGEN

Nach Flair filtern

Discussion Research Project

Twitter

LINKEDIN

- Philipp Schmid
- Lior Sinclair
- Hugging Face
- ...


edX


coursera


UDACITY


Udemy


OPEN
HPI
Hasso
Plattner
Institut


YouTube


KI-Campus
Die Lernplattform
für Künstliche Intelligenz

{ MACHINE # LEARNING > DEGREE } BY OPENCAMPUS.SH

GET A SOLID UNDERSTANDING OF MACHINE LEARNING AND LEARN HOW TO IMPLEMENT YOUR OWN STATE OF THE ART MACHINE LEARNING PROJECTS.

→ JOIN OUR COURSES

{ }

GET READY TO REALIZE YOUR OWN STATE OF THE ART MACHINE LEARNING PROJECTS

#

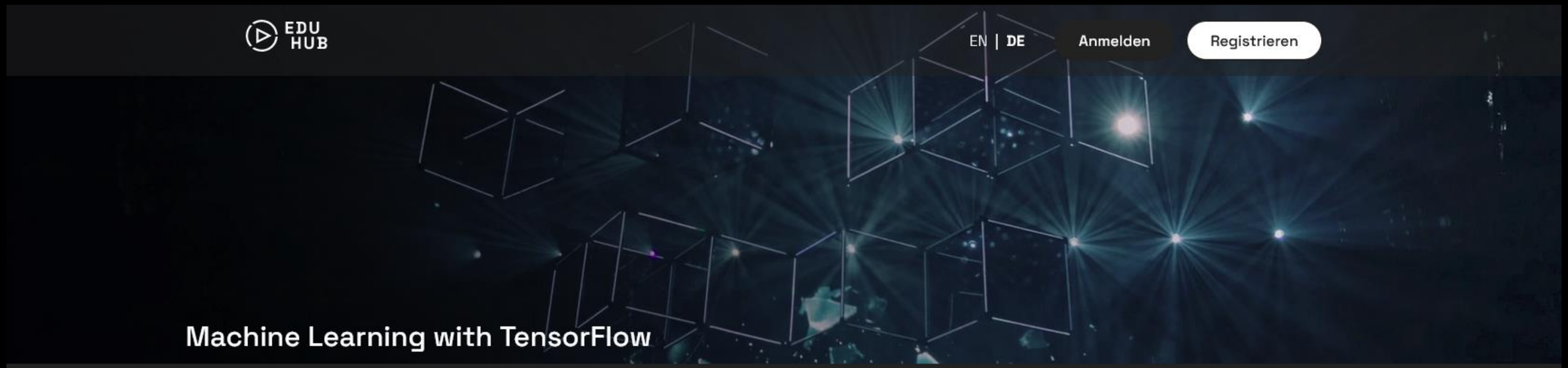
BECOME PART OF NETWORK OF MACHINE LEARNING ENTHUSIASTS

>

COMBINE WORLD CLASS LEARNING CONTENT WITH IMPLEMENTING PROJECTS

EDU HUB

EN | DE Anmelden Registrieren



Machine Learning with TensorFlow

DIENSTAG 16:00 - 17:45

Get hands-on experience in applying machine learning techniques with TensorFlow.

Die Bewerbungsfrist ist leider abgelaufen.

Du wirst lernen

- Best Practices für TensorFlow, ein populäres Open-Source-Framework für maschinelles Lernen, um neuronale Netzwerke zu trainieren
- Umgang mit Bilddaten aus der realen Welt und Erkundung von Strategien zur Vermeidung von Overfit, einschließlich Augmentation und Drop-Out
- Erstellung eines Systems zur Verabeitung natürlicher Sprache
- Anwendung von RNNs, GRUs und LSTMs zum Training dieser Lernmodelle unter Verwendung von Text- und Zeitreihendaten

DIENSTAG	ECTS
16:00 - 17:45	5
ONLINE + KIEL	ENGLISCH