25.05.21

Einführung in Data Science und maschinelles Lernen mit R

Neuronale Netze

OPENCAMPUS.sh

- Coding.Waterkant
- Zwischen-Feedback
- Wiederholung
- Aufbau Neuronaler Netze (NN)
- Hyperparameter in NN
- Frameworks zur Implementierung von NN
- Implementierung eines NN mit TensorFlow und Python

Registration: https://coding.waterkant.sh

Friday, June 4:

09:00 Welcome to Coding.Waterkant and Welcome Address by the Head of the State Chancellery, Dirk Schrödter

09:15 Keynote 1:
QuestionAId - How to use GPT Models for Task Automation (Malte Hecht, Co-Founder of SIGGI - Learn Smart)

09:45 Explore challenges, ask questions to the challenge patrons, and find a team and project

11:00 Workshop 1: *Prompt Design for GPT Models* (Vladimir Alexeev, OpenAI Ambassador)

11:00 Workshop 2: *Model Training on AWS* (Matthias Nannt, Co-Founder of Stack Ocean)

18:00 Update and Welcome to Participants from San Francisco

18:15 Keynote 2:
Neuroimaging and A.I. - What do we need, what is out there, how can we do better? (Julien Cohen-Adad, Associate Director of the Neuroimaging Functional Unit at University of Montreal)

Saturday, June 5:

09:00 Breakfast Stream

16:00 Final Stream: Project Presentations

OPENCAMPUS.sh

Title	Patron	Description	What we think	Data
Prediction of Surfing Conditions	Kay Sörnsen & Jonas Kaufmann (Windfinder)	The local weather influences on a surf spot (thermals, wind direction, cloudiness etc.) are very unique and best known to the local …	This is a classic supervised learning task and gives you the opportunity to test …	Windfinder will provide weather station data from sevaral surfspots as …
Time Series Prediction for Bakery Turnovers	Thies Schönfeldt (Meteolytics)	Bakery sales largely depend on to the day of the week, on holidays or vacation, local festivals that tale place and many more. An important …	Whether you want to gain first experience with time series predictions, train your first …	Meteolytics is providing daily turnover data for 13 different sales groups and three …
Diagnosis of Vertebral Body Fractures	Claus Glueer (University of Kiel) & Valentina Pedoia (University of San Francisco)	Using radiological image data, vertebral bodies are to be automatically examined for the presence of fractures. The …	Take the opportunity to get in contact with researchers from Kiel and San Francisco …	The University of Kiel will provide labelled X-ray scans of vertebral bodies to conduct the trainings.
Text Generation Using GPT	Doris Weßels (Kiel University of Applied Sciences)	Today's AI-based language models allow to generate text that is basically not to distinguish from text …	Instructing AI models via natural language is a very recent …	Training data is not needed but you will get access to the GPT-Neo model hosted …
Automated Essay Scoring	Sabrina Ludwig (Universität Mannheim) & Thorben Jansen (IPN Kiel)	In any learning context it is crucial to provide fast feedback to the learner. Being able to quickly analyze and categorize open text answers produced by the learner is …	By fine-tuning a state-of-the-art language model like GPT-Neo and using it for text classification, this …	The IPN will provide 900 scored texts and the University of Mannheim another set of about …
Your Own Challenge	You	Besides the challenges provided by us, you are of cause also very welcome to work on your own challenge and maybe find fellow …	The access to GPT-Neo provided by Kiel.AI might be the start for your …	

ZWISCHEN-FEEDBACK

https://forms.office.com/r/krA1ymsq0c

Wahl eines Prognosemodells

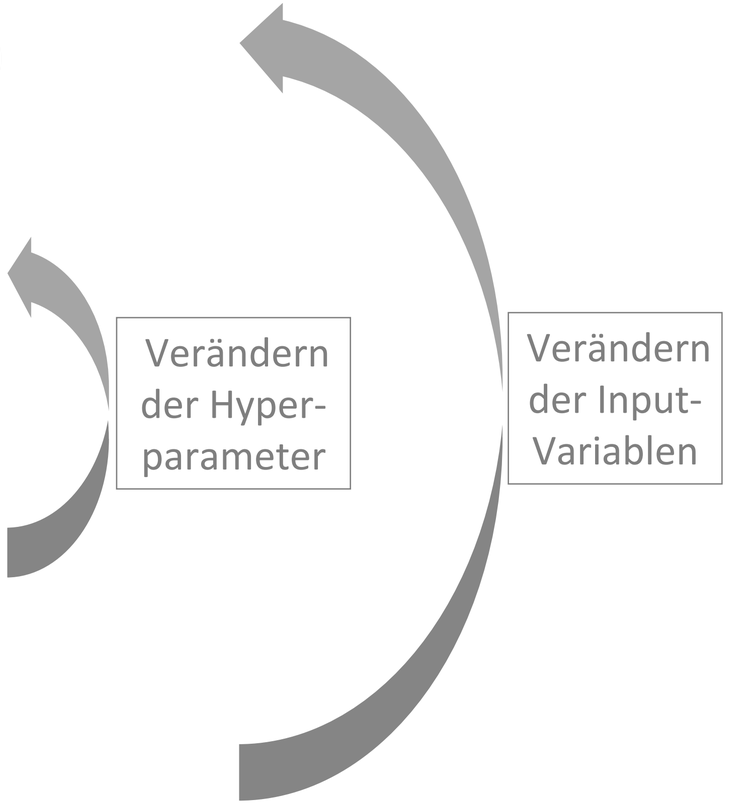
Teilung der Daten in Trainings- (70%), Validierungs- (20%) und Testdatensatz (10%)

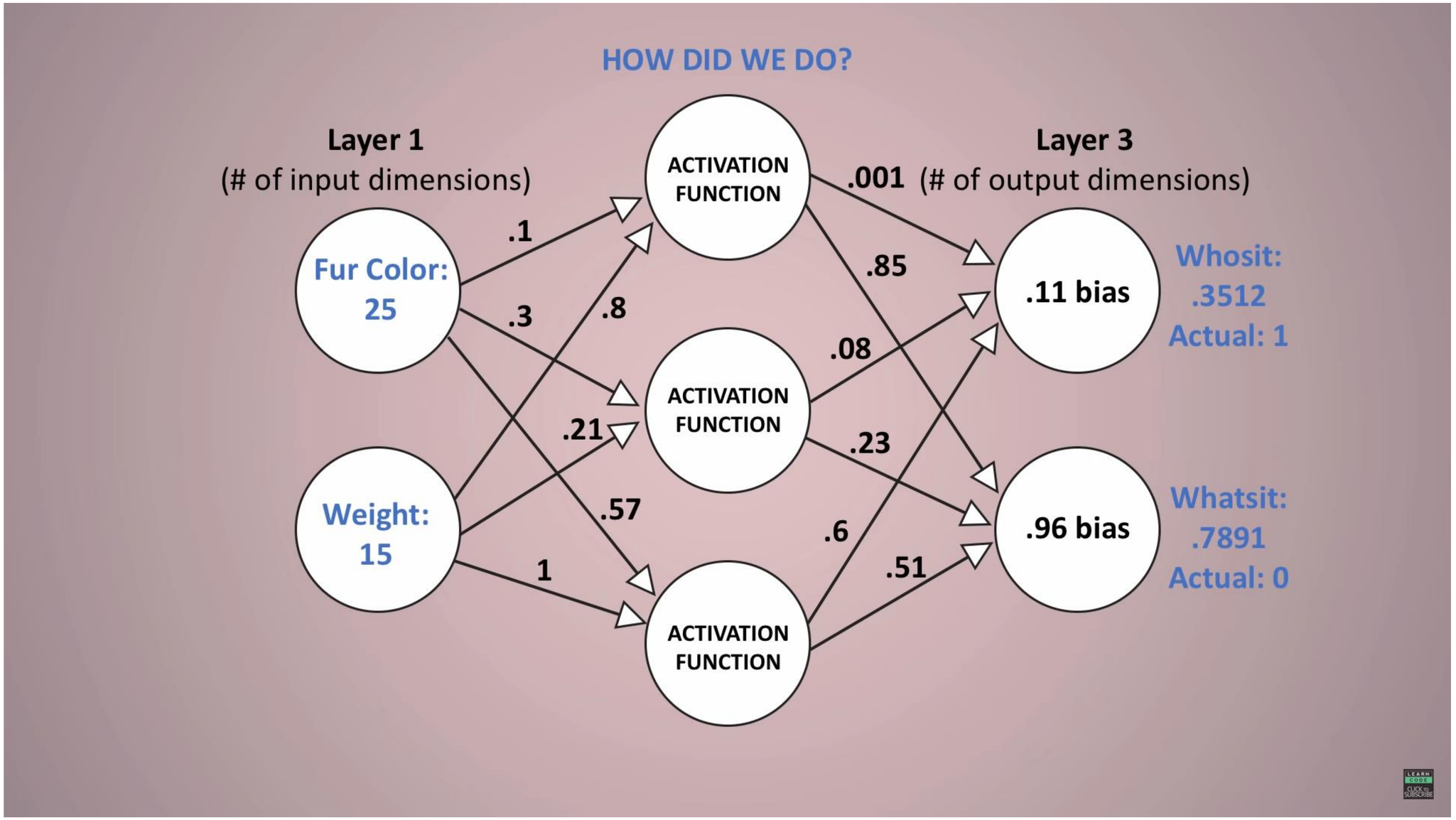
Optimierung der Modellparameter anhand des Trainingsdatensatzes

Optimierung der Hyperparameter anhand des Validierungsdatensatzes

Erweiterung/Verbesserung des Datensatzes

Überprüfung der Modellqualität anhand des Testdatensatzes




HOW DID WE DO?
Layer 1
(# of input dimensions)
Layer 3
(# of output dimensions)
Fur Color:
25
Weight:
15
ACTIVATION FUNCTION
ACTIVATION FUNCTION
ACTIVATION FUNCTION
.1
.3
.8
.21
.57
1
.001
.85
.08
.23
.6
.51
.11 bias
.96 bias
Whosit:
.3512
Actual: 1
Whatsit:
.7891
Actual: 0

WICHTIGE KONZEPTE

- **Aktivierungsfunktion („Vorhersagefunktion“)**
- **Kostenfunktion**
  - **Regularisierung**
    (Bestrafung der Verwendung von Variablen/ großen Parametern)
- **Optimierungsfunktion**
  (zur Minimierung der Kostenfunktion)
  - **Lernrate (Eigenschaft der Optimierungsfunktion)**


TensorFlow


K
Keras


scikit
learn


PYTORCH

- **Feb 2017: TensorFlow 1.0 (Estimator API)**
- **Nov 2017: TensorFlow 1.4 (Estimator API, Keras API)**
- **Jan 2019: TensorFlow 2.0 (Estimator API, Keras API)**

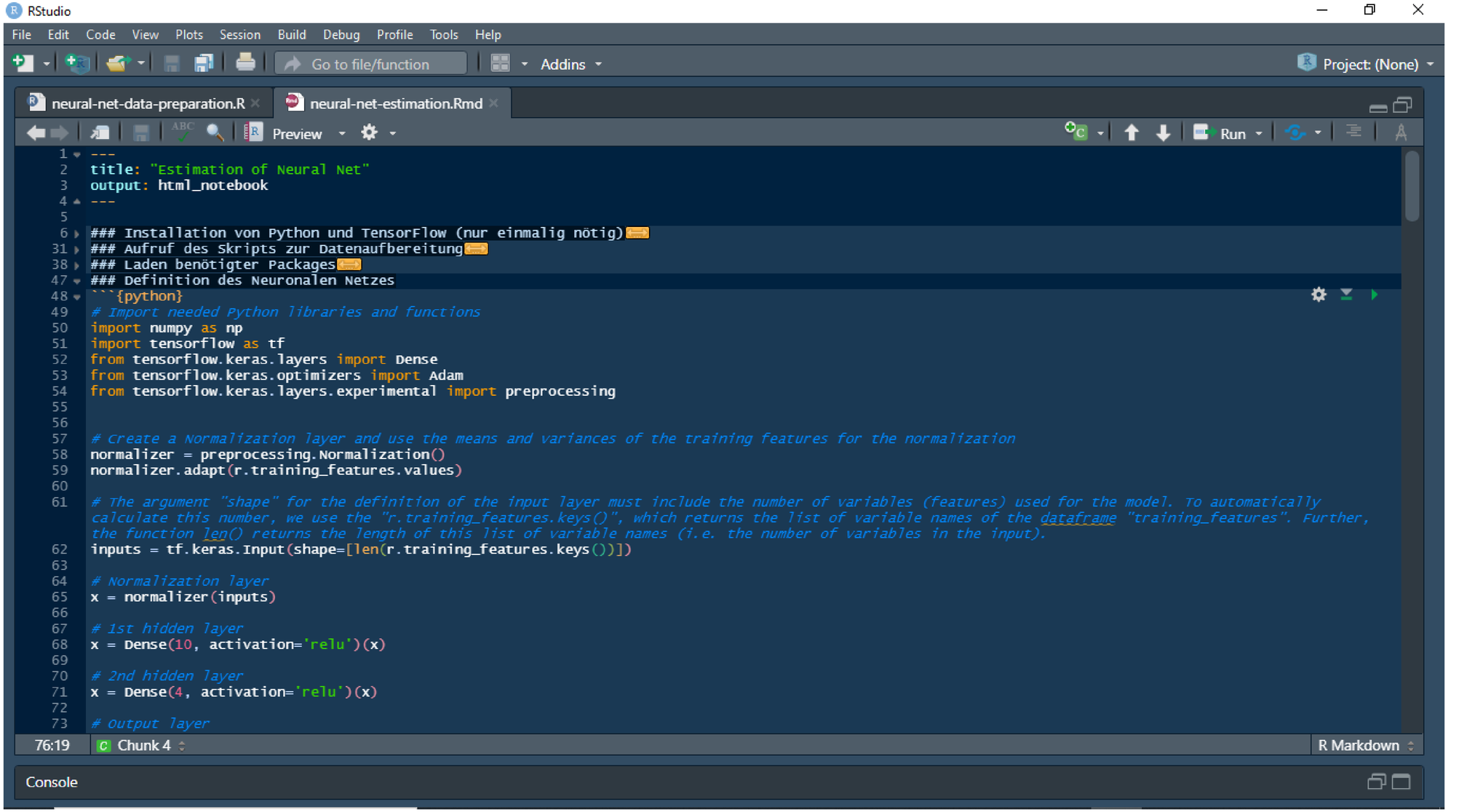
NUTZUNG VON KERAS IN R

Keras ist eine Schnittstelle (API/ ein Funktionswrapper) zur vereinfachenden Nutzung von TensorFlow.

Prinzipiell zwei Varianten :

- **Nutzung des Packages „keras“ (vgl. https://keras.rstudio.com/)**
- **Nutzung von Keras in Python und die Integration von Python über das Paket „reticulate“**

DEFINITION EINES NEURONALEN NETZES



HYPERPARAMATER IN NEURONALEN NETZEN

- Wahl der Aktivierungsfunktionen
- Wahl der Kostenfunktion
- Wahl der Optimierungsfunktion
- **Wahl der Parameter der Optimierungsfunktion**
- **Anzahl der Hidden Layer des Netzes**
- **Anzahlen der Neuronen der Hidden Layer**
- **Art der Hidden Layer**

PARAMETER DES „ADAM“ OPTIMIZERS

- **alpha (learning rate):**
  Lernparameter/Schrittweite der Optimierung

- **beta1 and beta2 (momentum):**
  Trägheit der Optimierung

AUFGABEN

- **Trainiert bitte bis zur nächsten Woche ein neuronales Netz für Euren Datensatz und erstellt eine Vorhersage für den 06.06.2019**

- **Schaut dieses Video (5 Minuten) zu Zeitreihenanalysen an.**

- **Wenn ihr etwas mehr über Python lernen wollt, könnt Ihr diese Einführung auf Kaggle nutzen.**