

Data Science & Machine Learning

Dieter Greipl

2022-01-07

Contents

| | |
|--|-----------|
| Willkommen | 5 |
| 0.1 Vorbereitungen | 5 |
| 0.2 Tests (später löschen!) | 6 |
| 0.3 Google Konto | 9 |
| 0.4 Colab | 9 |
| 0.5 Programmierkenntnisse | 9 |
| 0.6 Mathematik | 9 |
| 0.7 Teaching-Flow & Agenda | 10 |
| 1 Hello bookdown | 13 |
| 1.1 A section | 13 |
| 2 Cross-references | 15 |
| 2.1 Chapters and sub-chapters | 15 |
| 2.2 Captioned figures and tables | 15 |
| 3 Parts | 19 |
| 4 Footnotes and citations | 21 |
| 4.1 Footnotes | 21 |
| 4.2 Citations | 21 |

| | | |
|----------|--------------------------------|-----------|
| 5 | Blocks | 23 |
| 5.1 | Equations | 23 |
| 5.2 | Theorems and proofs | 23 |
| 5.3 | Callout blocks | 23 |
| 6 | Sharing your book | 25 |
| 6.1 | Publishing | 25 |
| 6.2 | 404 pages | 25 |
| 6.3 | Metadata for sharing | 25 |

Willkommen

Dieses Skript entstand (und entsteht) aus meinen Lehrveranstaltungen rund um das Thema **Data Science & Machine Learning**. Die Inhalte richten sich an Studierende, die erste Schritte auf das KI -Spielfeld wagen und das Potential von datengetriebenen Lösungsverfahren verstehen wollen.

Insofern richtet sich die Darstellung an Studierende mit vertieftem Interesse an KI, die einen für Studierende angemessenen Vorwissen im Bereich Mathematik mitbringen. Vorkenntnisse im Bereich der Programmierung sind nicht nötig, aber natürlich hilfreich.

Ich habe mich bemüht, zahlreiche Übungsbeispiele und Youtube-Videos einzubauen. Viele Themen lassen sich so besser darstellen. Sofern es Medien im Netz gibt, die die Sachverhalte gut darstellen, werde ich entsprechenden Links einbauen. Der Autor muss ja nicht der Meinung sein, alles besser zu können. Gleichwohl darf dadurch der rote Faden nicht verloren gehen.

0.1 Vorbereitungen

Dieses Skript ist als Unterlage für zahlreichen praktische Übungen mit Python angelegt. Ich werde hierzu **Colab-Notebooks** verwenden. Sie brauchen hierzu ein **Google-Konto**.

Noch einige Hinweise an Studierende meiner Module:

- Die folgende Youtube-Playlist kann zur Vertiefung einzelner Stoffteile nutzen: Youtube Playlist
- Wenige Passagen in diesem Skript sind eventuell in englischer Sprache gehalten.
- Dieses Skript
 - befindet sich in Teilen im Aufbau, leichte Fehler sind also möglich (und wahrscheinlich - für Hinweise bin ich dankbar)
 - geht nach der Prüfung off-line

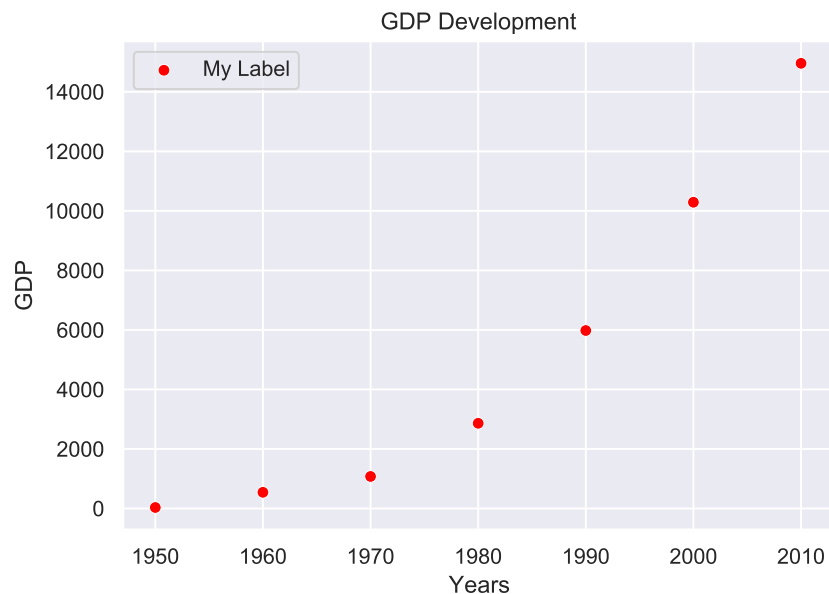
0.2 Tests (später löschen!)

```
#Scatterplot
import matplotlib.pyplot as plt
import seaborn as sns

years = [1950, 1960, 1970, 1980, 1990, 2000, 2010]
gdp = [33.2, 543.3, 1075.9, 2862.5, 5979.6, 10289.7, 14958.3]

sns.set()
fig, ax = plt.subplots()
ax.set_title("GDP Development")
ax.set_xlabel("Years")
ax.set_ylabel("GDP")
#ax.set_aspect('equal')
#ax.set_xlim(0, 50)
#ax.set_ylim(0, 35)

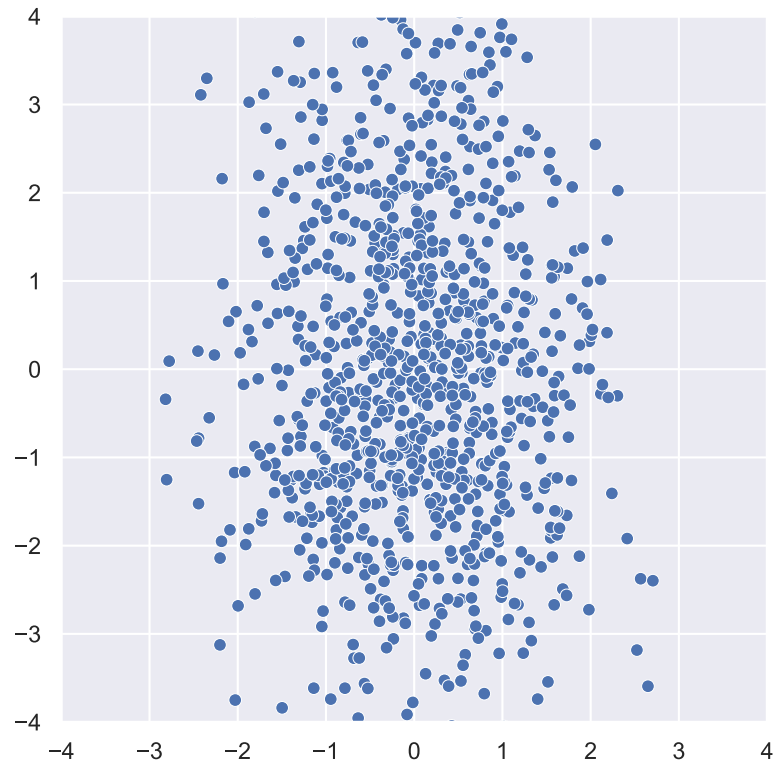
sns.scatterplot(x=years, y=gdp, color="red", label="My Label")
#sns.lineplot(x=years, y=gdp, color="red", label="My Label")
#sns.barplot(x=years, y=gdp, color="red", label="My Label")
```



```
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

N = 1000
x_werte = np.random.normal(0,1,N)
y_werte = np.random.normal(0,2,N)

sns.set()
fig,ax = plt.subplots(figsize=(6, 6))
ax.set_xlim(-4, 4)
#> (-4.0, 4.0)
ax.set_ylim(-4, 4)
#> (-4.0, 4.0)
ax.set_aspect('equal')
sns.scatterplot(x = x_werte, y=y_werte)
#print("Wert von N: ", N)
```



Code mit Ausdruck (See <https://yihui.org/knitr/options/#code-evaluation>)

```
x = 1
print(x)
print(x+1)
#> 1
#> 2
```

Video

| Start ▾ | Ende ▾ | Dauer ▾ | Content ▾ |
|--------------|--------|---------|------------|
| 08:30 | 10:00 | 01:30 | |
| 10:00 | 10:10 | 00:10 | - break - |
| 10:10 | 11:10 | 01:00 | |
| 11:10 | 11:20 | 00:10 | - break - |
| 11:20 | 12:30 | 01:10 | |
| 12:30 | 13:00 | 00:30 | - Mittag - |
| 13:00 | 14:30 | 01:30 | |
| 14:30 | 14:40 | 00:10 | - break - |
| 14:40 | 16:00 | 01:20 | |

Alternativ

0.3 Google Konto

Sie brauchen ein Google-Konto für die Nutzung von Colab und ggf. für das Laden von Dateien von Google Drive.

0.4 Colab

Für unsere praktischen Aktivitäten nutzen wir die Colab-Notebooks.

0.5 Programmierkenntnisse

Programmierkenntnisse sind nicht erforderlich. Sie werden das Nötige im Kurs lernen.

0.6 Mathematik

Wir benötigen folgende Grundlagen, die ich aber kurz wiederholen werde:

- Verteilung, Gleichverteilung, Normalverteilung, Erwartungswert, Standardabweichung, Modus, Median
- Vektor, Matrix, Skalarprodukt
- Funktionen, speziell lineare Abbildungen

0.7 Teaching-Flow & Agenda

Wir springen etwas zwischen den Kapiteln, um die inhaltlichen Kapitel (A,B) nicht mit den Python-Grundlagen (C) zu vermischen.

Link zu den in der jeweiligen Terminen erzeugten Colab-Notebooks

0.7.1 10.01.2022, 18:30 - max 20:00 Uhr: Fragen zur Prüfung

Link: <https://haw-landshut-de.zoom.us/j/86239229907?pwd=eFNTUEdCeVhyMGJaaEg2YlhUdzhXdz>

18.12.2021, 08:30 - 16:00 Uhr

| Start | Ende | Dauer | Content |
|-------|-------|-------|------------|
| 08:30 | 10:00 | 01:30 | |
| 10:00 | 10:10 | 00:10 | - break - |
| 10:10 | 11:10 | 01:00 | |
| 11:10 | 11:20 | 00:10 | - break - |
| 11:20 | 12:30 | 01:10 | |
| 12:30 | 13:00 | 00:30 | - Mittag - |
| 13:00 | 14:30 | 01:30 | |
| 14:30 | 14:40 | 00:10 | - break - |
| 14:40 | 16:00 | 01:20 | |

Figure 1: image-20211218081837145

17.12.2021, 16:00 - 20:15 Uhr

B.2 Entscheidungsbäume - DSA Master DUF

B.3 Training, Test und Validation - DSA Master DUF

B.4 Regression (und Grillen) - DSA Master DUF

| Start ▾ | Ende ▾ | Dauer ▾ | Content ▾ |
|---------|--------|---------|-----------|
| 16:00 | 17:30 | 01:30 | |
| 17:30 | 17:40 | 00:10 | - break - |
| 17:40 | 18:40 | 01:00 | |
| 18:40 | 19:00 | 00:20 | - break - |
| 19:00 | 20:15 | 01:15 | |

Figure 2: image-20211217083343363

11.12.2021, 13:00 - 17:30 Uhr

C2. Elementweise Operationen und C.2 Numpy - Matrix

C.3 Pandas und Dataframes

A.2 Plots für Iris abschließen

A.3 Preprocessing

B1. Problemstellung und Grundbegriffe

10.12.2012, 14:00 - 20:15 Uhr**A.1 Datensätze**

C.1 Python - Variablen und Datentypen

C.2 Python - Numpy (ohne Matrix)

A.2 Visualisierung (ohne "Bilder", Plots für Iris noch abschließen)**A.3 Preprocessing**

| Start ▾ | Ende ▾ | Dauer ▾ | Content ▾ |
|--------------|--------|---------|-----------|
| 13:00 | 14:00 | 01:00 | |
| 14:00 | 14:10 | 00:10 | - break - |
| 14:10 | 15:10 | 01:00 | |
| 15:10 | 15:20 | 00:10 | - break - |
| 15:20 | 16:20 | 01:00 | |
| 16:20 | 16:30 | 00:10 | - break - |
| 16:30 | 17:30 | 01:00 | |

Figure 3: image-20211211090727800

Freitag, 10.12.2021, 14:00 - 20:15

| Start ▾ | Ende ▾ | Dauer ▾ | Content ▾ |
|---------|--------|---------|-----------|
| 14:00 | 15:30 | 01:30 | |
| 15:30 | 15:40 | 00:10 | - break - |
| 15:40 | 17:10 | 01:30 | |
| 17:10 | 17:35 | 00:25 | - break - |
| 17:35 | 19:05 | 01:30 | |
| 19:05 | 19:15 | 00:10 | - break - |
| 19:15 | 20:15 | 01:00 | |

Figure 4: image-20211210130419842

Chapter 1

Hello bookdown

All chapters start with a first-level heading followed by your chapter title, like the line above. There should be only one first-level heading (#) per .Rmd file.

1.1 A section

All chapter sections start with a second-level (##) or higher heading followed by your section title, like the sections above and below here. You can have as many as you want within a chapter.

An unnumbered section

Chapters and sections are numbered by default. To un-number a heading, add a {.unnumbered} or the shorter {-} at the end of the heading, like in this section.

Chapter 2

Cross-references

Cross-references make it easier for your readers to find and link to elements in your book.

2.1 Chapters and sub-chapters

There are two steps to cross-reference any heading:

1. Label the heading: `# Hello world {#nice-label}`.
 - Leave the label off if you like the automated heading generated based on your heading title: for example, `# Hello world = # Hello world {#hello-world}`.
 - To label an un-numbered heading, use: `# Hello world {-#nice-label}` or `{# Hello world .unnumbered}`.
2. Next, reference the labeled heading anywhere in the text using `\@ref(nice-label)`; for example, please see Chapter 2.
 - If you prefer text as the link instead of a numbered reference use: any text you want can go here.

2.2 Captioned figures and tables

Figures and tables *with captions* can also be cross-referenced from elsewhere in your book using `\@ref(fig:chunk-label)` and `\@ref(tab:chunk-label)`, respectively.

See Figure 2.1.

```
par(mar = c(4, 4, .1, .1))  
plot(pressure, type = 'b', pch = 19)
```

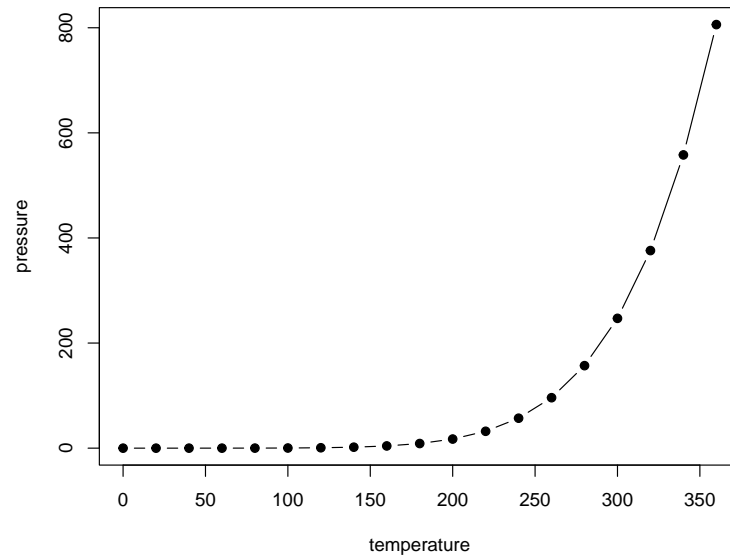


Figure 2.1: Here is a nice figure!

Don't miss Table 2.1.

```
knitr::kable(  
  head(pressure, 10), caption = 'Here is a nice table!',  
  booktabs = TRUE  
)
```


Table 2.1: Here is a nice table!

| temperature | pressure |
|-------------|----------|
| 0 | 0.0002 |
| 20 | 0.0012 |
| 40 | 0.0060 |
| 60 | 0.0300 |
| 80 | 0.0900 |
| 100 | 0.2700 |
| 120 | 0.7500 |
| 140 | 1.8500 |
| 160 | 4.2000 |
| 180 | 8.8000 |

Chapter 3

Parts

You can add parts to organize one or more book chapters together. Parts can be inserted at the top of an .Rmd file, before the first-level chapter heading in that same file.

Add a numbered part: `# (PART) Act one {-}` (followed by `# A chapter`)

Add an unnumbered part: `# (PART*) Act one {-}` (followed by `# A chapter`)

Add an appendix as a special kind of un-numbered part: `# (APPENDIX) Other stuff {-}` (followed by `# A chapter`). Chapters in an appendix are prepended with letters instead of numbers.

Chapter 4

Footnotes and citations

4.1 Footnotes

Footnotes are put inside the square brackets after a caret `^[]`. Like this one ¹.

4.2 Citations

Reference items in your bibliography file(s) using `@key`.

For example, we are using the **bookdown** package (Xie, 2021) (check out the last code chunk in `index.Rmd` to see how this citation key was added) in this sample book, which was built on top of R Markdown and **knitr** (Xie, 2015) (this citation was added manually in an external file `book.bib`). Note that the `.bib` files need to be listed in the `index.Rmd` with the YAML `bibliography` key.

The `bs4_book` theme makes footnotes appear inline when you click on them. In this example book, we added `cs1: chicago-fullnote-bibliography.cs1` to the `index.Rmd` YAML, and include the `.cs1` file. To download a new style, we recommend: <https://www.zotero.org/styles/>

The RStudio Visual Markdown Editor can also make it easier to insert citations: <https://rstudio.github.io/visual-markdown-editing/#/citations>

¹This is a footnote.

Chapter 5

Blocks

5.1 Equations

Here is an equation.

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (5.1)$$

You may refer to using `\@ref{eq:binom}`, like see Equation (5.1).

5.2 Theorems and proofs

Labeled theorems can be referenced in text using `\@ref{thm:tri}`, for example, check out this smart theorem 5.1.

Theorem 5.1. *For a right triangle, if c denotes the length of the hypotenuse and a and b denote the lengths of the **other** two sides, we have*

$$a^2 + b^2 = c^2$$

Read more here <https://bookdown.org/yihui/bookdown/markdown-extensions-by-bookdown.html>.

5.3 Callout blocks

The `bs4_book` theme also includes special callout blocks, like this `.rmdnote`.

You can use **markdown** inside a block.

```
head(beaver1, n = 5)
#>   day time  temp activ
#> 1 346  840 36.33     0
#> 2 346  850 36.34     0
#> 3 346  900 36.35     0
#> 4 346  910 36.42     0
#> 5 346  920 36.55     0
```

It is up to the user to define the appearance of these blocks for LaTeX output.

You may also use: `.rmdcaution`, `.rmdimportant`, `.rmdtip`, or `.rmdwarning` as the block name.

The R Markdown Cookbook provides more help on how to use custom blocks to design your own callouts: <https://bookdown.org/yihui/rmarkdown-cookbook/custom-blocks.html>

Chapter 6

Sharing your book

6.1 Publishing

HTML books can be published online, see: <https://bookdown.org/yihui/bookdown/publishing.html>

6.2 404 pages

By default, users will be directed to a 404 page if they try to access a webpage that cannot be found. If you'd like to customize your 404 page instead of using the default, you may add either a `_404.Rmd` or `_404.md` file to your project root and use code and/or Markdown syntax.

6.3 Metadata for sharing

Bookdown HTML books will provide HTML metadata for social sharing on platforms like Twitter, Facebook, and LinkedIn, using information you provide in the `index.Rmd` YAML. To setup, set the `url` for your book and the path to your `cover-image` file. Your book's `title` and `description` are also used.

This `bs4_book` provides enhanced metadata for social sharing, so that each chapter shared will have a unique description, auto-generated based on the content.

Specify your book's source repository on GitHub as the `repo` in the `_output.yml` file, which allows users to view each chapter's source file or suggest an edit. Read more about the features of this output format here:

https://pkgs.rstudio.com/bookdown/reference/bs4_book.html

Or use:

```
?bookdown::bs4_book
```

Bibliography

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.

Xie, Y. (2021). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.24.