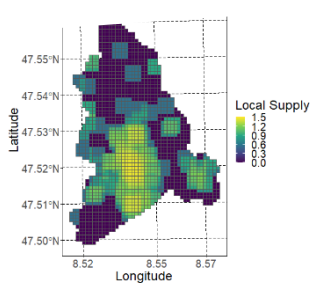
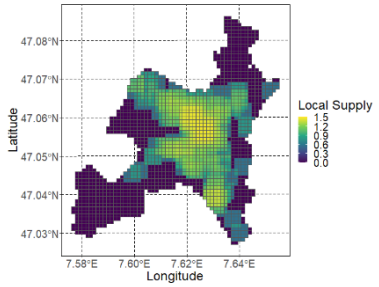
Daten Formate – Tensors and Batches

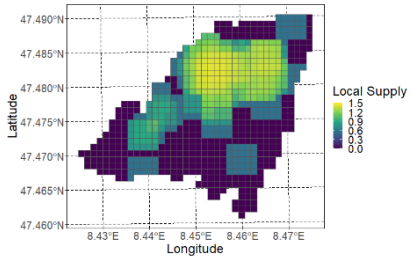
Dr. Yves Staudt



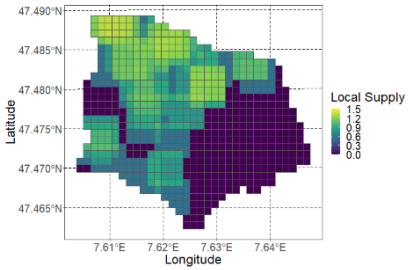
(a)
Grenchen



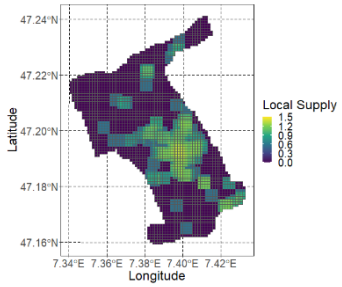
(b)
Oberglatt



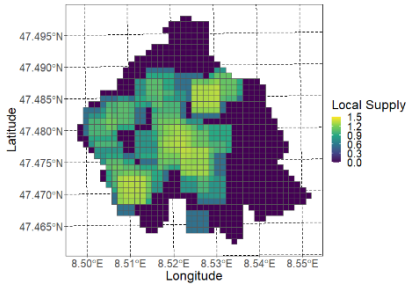
(c)
Reinach



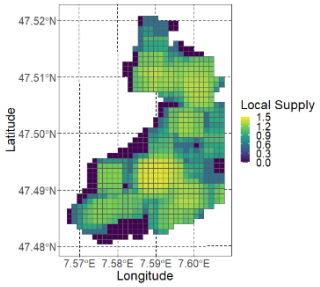
(d)
Rubigen



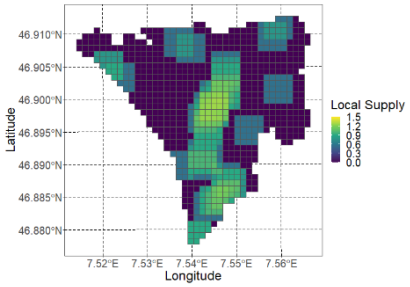
(e)
Buelach



(f)
Burgdorf



(g)
Dielsdorf



(h)
Dornach

Lernziel

Die Studierende sind in der Lage

- die Datenformate für ein Deep Learning zu bestimmen.
- das Prinzip von Batches zu bestimmen.

FH GR

Goal von Deep Learning

- The network will learn to associate images and labels
- Core building block of neural networks is the layer
- The layer can be understood as a data-preprocessing module, as for example a filter for data
- Data comes into a layer and comes out in a more useful form – layers extract representations out of the data
- Deep learning mostly consists of chaining together simple layers that will implement a form of progressive data distillation
- “Deep learning model is like a sieve for data processing, made of a succession of increasingly refined data filters – the layers.” – said by Chollet (2019)

(Chollet, 2019)

FH GR

Classes and Labels

In this slide we represents the nomenclature for classes and labels in machine learning

- **Class** is a category, called class in a classfiication problem
- **Samples** are data points
- **Label** is the class associated with a specific sample

FH GR

Data Representations

- Data structure used in deep learning models are described by tensors
- Tensor is a container of data – almost always nuermical data
- Tensor is generalization of matrics

- Mostly the following dimensions of tenosrs are encountered in praxis
  - ❖ 0D Tensors – Scalars
  - ❖ 1D Tensors – Vectors
  - ❖ 2D Tensors – Marices
  - ❖ 3D Tensors and higher dimensions

- In context of tensors a dimension is caleld an axis

FH GR

Tensors

0D Tensors – Scalars

- Tensor containing only on number is called scalar

1D Tensors – Vectors

- An array of numbers is called a vector

2D Tensors – Matrics

- An array of vectors is a matrix
- Matrix has two axes referred as rows and columns
- Matrix can be interpreted as a rectangular grid of numbers

3D Tensors and higher dimensional matrics

- 3D tensors are obtained by packing matrices in a new array
- 3D tensors can be interpreted a cube of numbers

FH GR

Key attributes

A tensor is defined by three key attributes

- **Number of axes (rank)** – 3D tensor has three axes, matrix has two axes
- **Shape** – is a tuple of integers that descirbes how many dimensions the tensor has along each axis
- **Data type** – is the the type of the ata contained in the tensor. Tensor's type could be float 32, unit8, float64 and so on. Rarely the format is char.

FH GR

Python Code and Manipulating Tensors

- Zum trainieren sollen sie die Fragestellungen auf Moodle durchführen.
- Wir werden dies dann gemeinsam verbessern.

FH GR

Notion of Data Batches

- The first axis (in Pyhton axis 0) in all data tensors in deep learning will be the sample axis
- In MNIST example, samples are images of digits
- Deep learning models do not process an entire dataset at once, rather they beak the data into small batches
- An often used batch size is 32

FH GR

Real World Examples

Most used data in real life examples fall in one of the following categories:

- **Vector data** – 2D tensors of shape (samples, features)
- **Timeseries data or sequence data** – 3D tensors of shape (samples, timesteps, features)
- **Images** – 4D tensors of shape (samples, height, width, channels – for example colors) or (samples, channels, height, width)
- **Video** – 5D tensors of shape (samples, frames, height, width, channels) or (samples, frames, channels, height, width)

FH GR

Vector Data

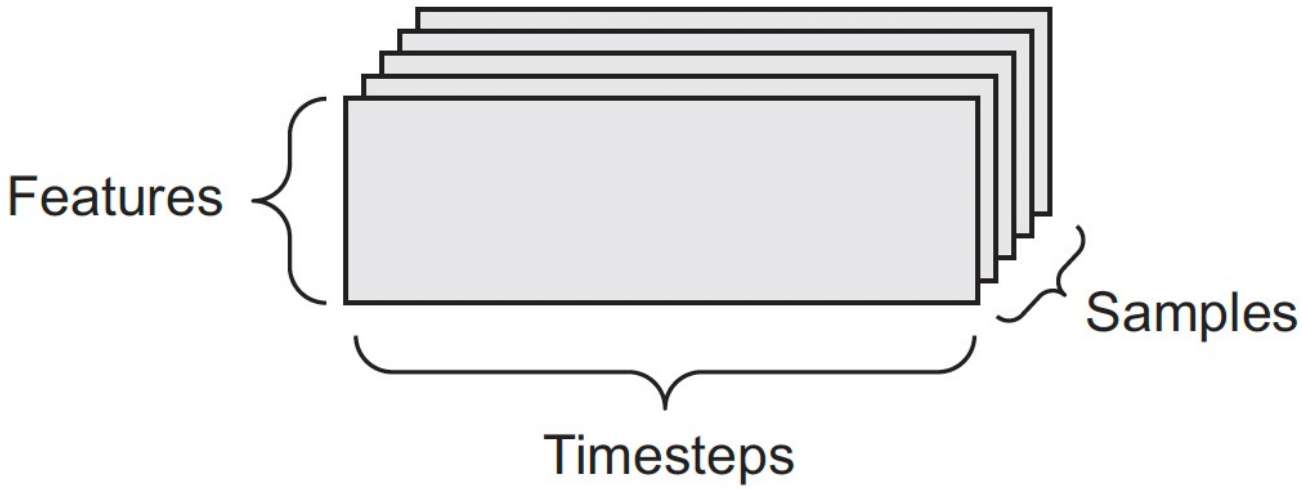
- Die meisten Daten über Personen werden in Tabellen gespeichert.
- Die Zeilen stellen die Proben/Observationen (sample) dar, die zweite die Eigenschaften.
- Beispiel

Alter	Postleitzahl	Einkommen
32	7000	100000
40	1700	120000
…	…	…

FH GR

Timeseries Data Or Sequence Data

Representation of a 3D tensor times series tensor:

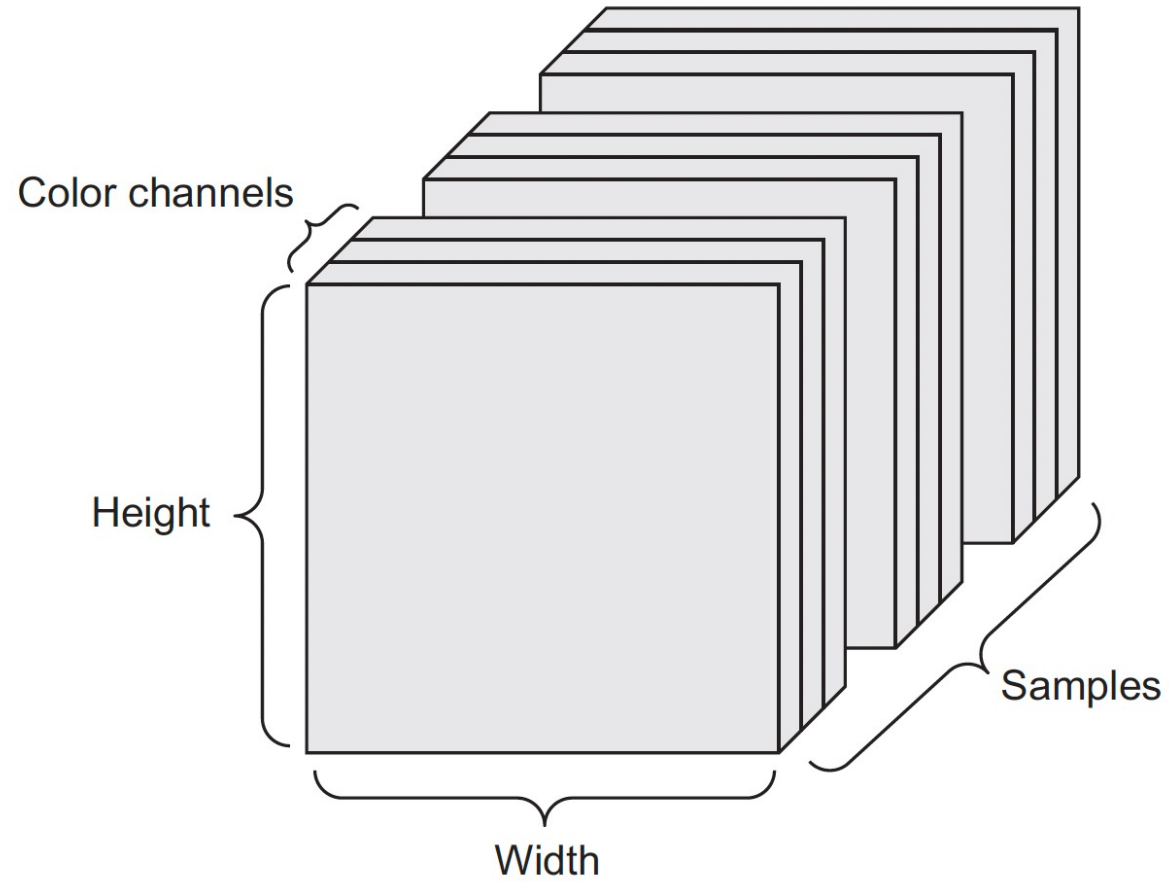


Beispiel: Speicherung von Aktien Preisen jede Minute für die 250 Handelstage: Wenn ich jeden Tag als Beobachtung sehe mit den Eigenschaften vom Niedrigsten, Höchsten und Aktuellen Preis speichere, enthalte ich für jeden Tag wleche 390 Minuten Handel beinhaltet eine Tabelle der Dimension (390, 3). Somit erhalte ich für die 250 Tage einen 3D Tensor mit folgenden Dimensionen entlang der Achsen (250, 390, 3)

FH GR

Image Data

Abbildung eines Tensors für ein Bild



Referenzen

- Chollet, F. (2019), Deep Learning with Keras, Manning.

FH GR

Fragen

Darstellung eines Fragesymbol aufgerufen von der Webseite https://www.qnigge.de/news/detail/modul-v/#images am 12.07.2021.

Fachhochschule Graubünden
Pulvermühlestrasse 57
7000 Chur
T +41 81 286 24 24
info@fhgr.ch

Vielen Dank für Ihre Aufmerksamkeit.
Grazia fitg per l'attenziun.
Grazie per l'attenzione.

Fachhochschule Graubünden
Scola auta spezialisada dal Grischun
Scuola universitaria professionale dei Grigioni
University of Applied Sciences of the Grisons

swissuniversities

SCHWEIZERISCHER AKKREDITIERUNGSRAT
CONSEIL SUISSE D'ACCRÉDITATION
CONSIGLIO SVIZZERO DI ACCREDITAMENTO
SWISS ACCREDITATION COUNCIL
Institutionell akkreditiert nach HFKG 2018-2025