Introduction to Deep Learning (I2DL)

Exercise 3: Data

Today's Outline

- Organizational stuff
- Exercises outline
- Contents of
  - Example Datasets & -loader
  - Exercise 3 (Submission #2)

Organization

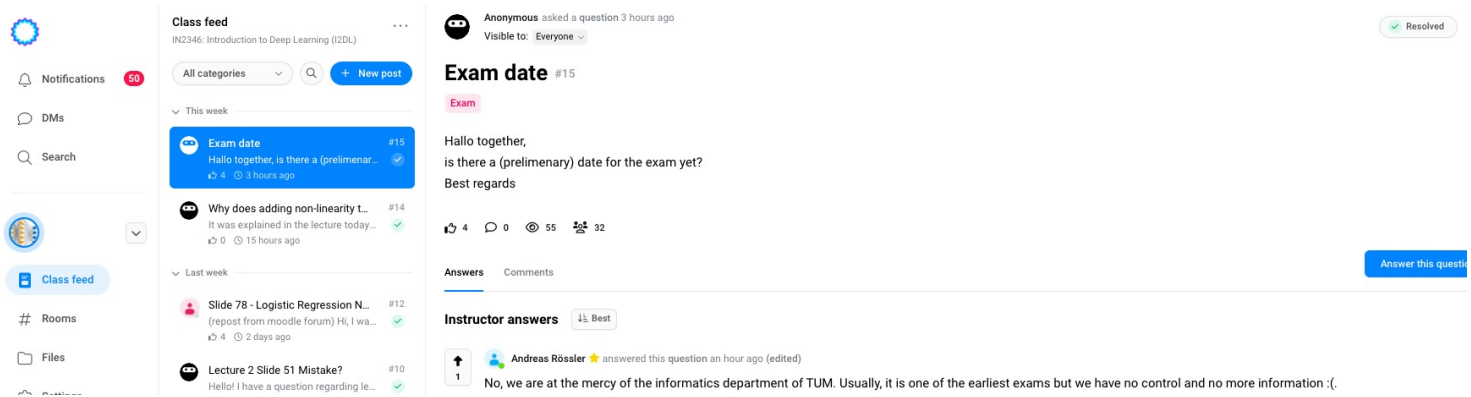
Organization

- Weekly announcements on moodle
  Will replace most of the video annoucement content. Let us know if you also want it repeated in the videos

- Office hours have started
  Have mercy with our TAs
  - Office hour schedule is on moodle
  - We will try to allocate more TA hours to office hours later on
  - For now no personal contact with lecturers

Organization

- New forum: campuswire
  Rational behind campuswire is on moodle and a thread on campuswire with more behind the scenes info
  - Right now no more forum administration on moodle



Organization

- Random stats
  - ~2000 registered students on moodle
  - ~1100 submissions for exercise 1

1140	u1779	60.00
1141	u1987	60.00
1142	u1004	60.00
1143	u1055	60.00
1144	u0063	60.00

- Reminder: Use the resources!
  And give us feedback where we can do better

General Exercise Overview

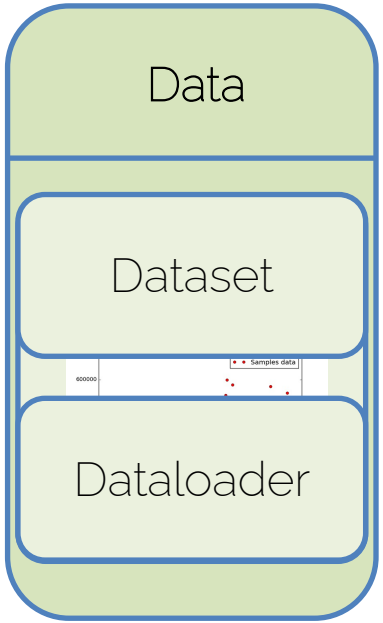
Your task for exercise 3-5

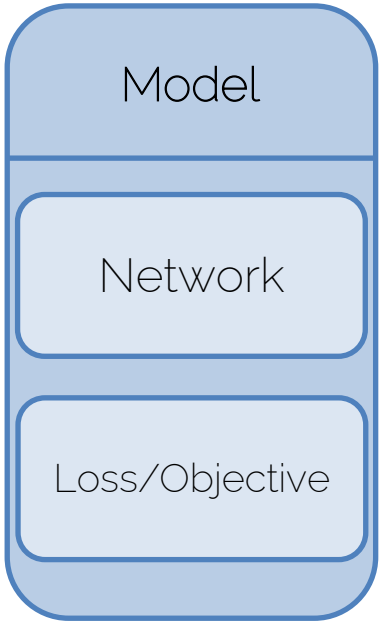
Exercise 03: Dataset and Dataloader
Exercise 04: Solver and Linear Regression
Exercise 05: Neural Networks
Exercise 06: Hyperparameter Tuning

Numpy
(Reinvent the wheel)

- Implementation of
  - A simple dataset and data loading
  - Regression/classification pipeline using Neural Networks

The Pillars of Deep Learning

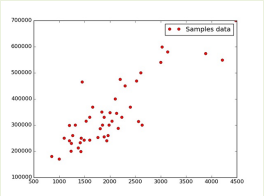




The Pillars of Deep Learning

Data

$$\{x_{1:n}, y_{1:n}\}$$



Model

$$\hat{y}_i = \theta_0 + \sum_{j=1}^{d} x_{ij}\theta_j$$

bias

$$J(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$$

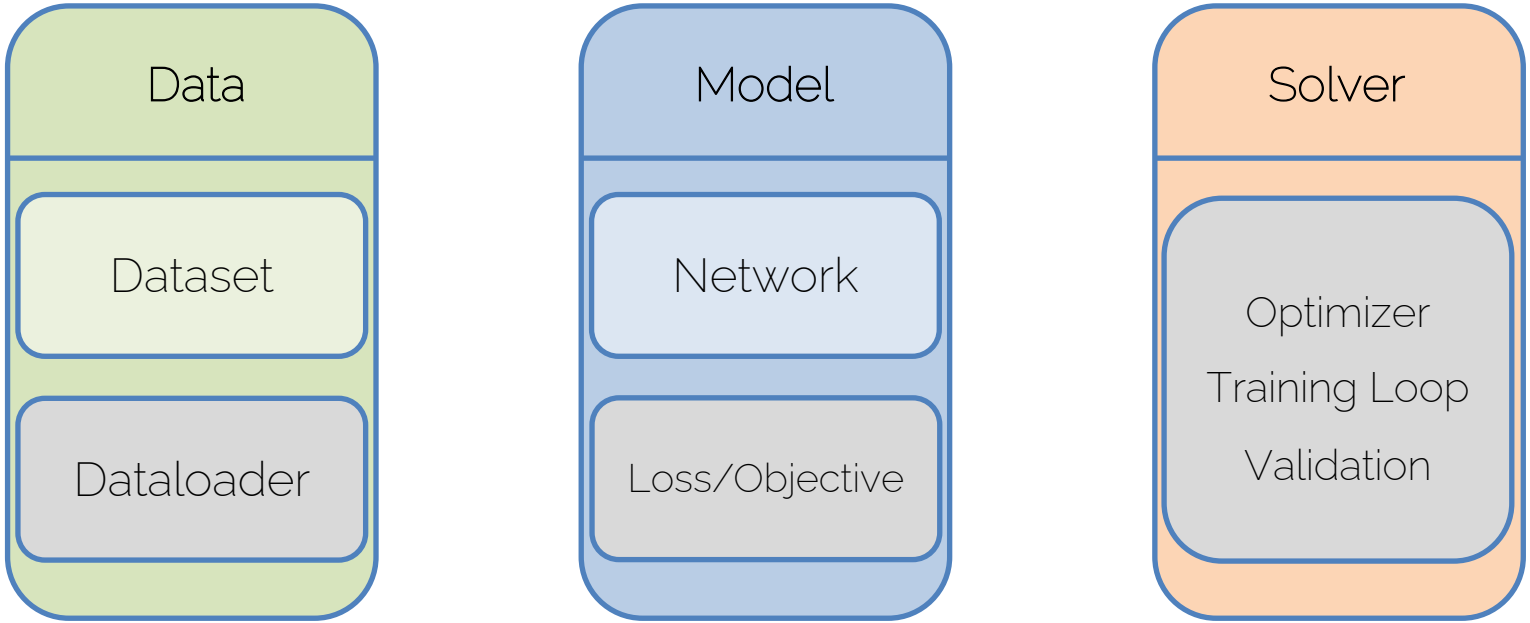
Solver

Optimizer

Training Loop

Validation

The Pillars of Deep Learning

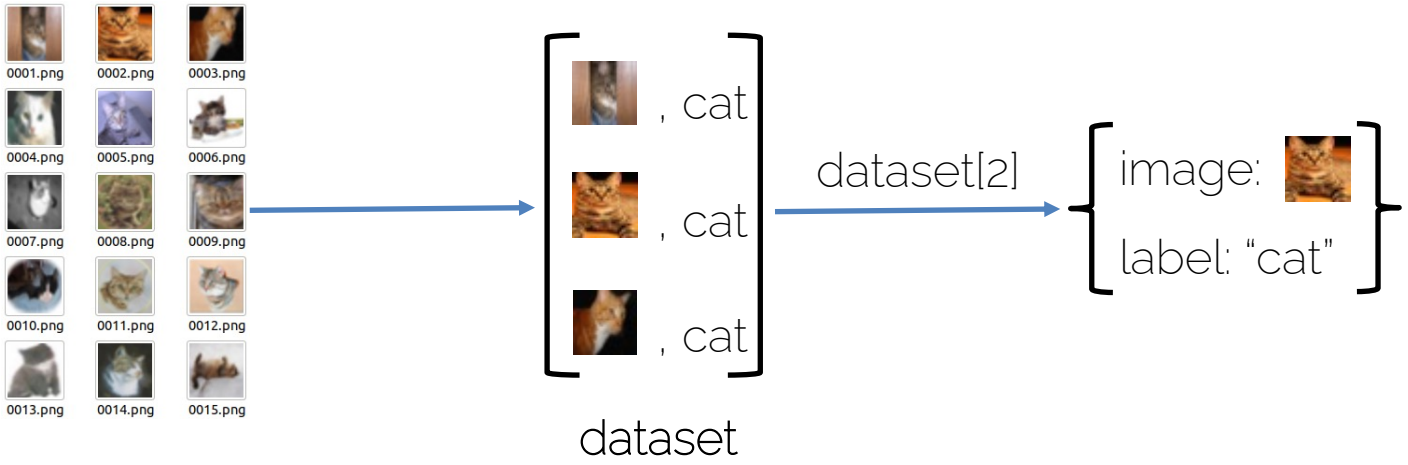


Can be implemented once and used in multiple projects

Exercise 3

Exercise 3: Dataset

- Reads data and provides a simple way to access it
- Performs data preprocessing steps using Transforms
- Example: Image Folder Dataset



Exercise 3: Dataset

- What we excluded
  - Low level "scripting" details using operating system calls

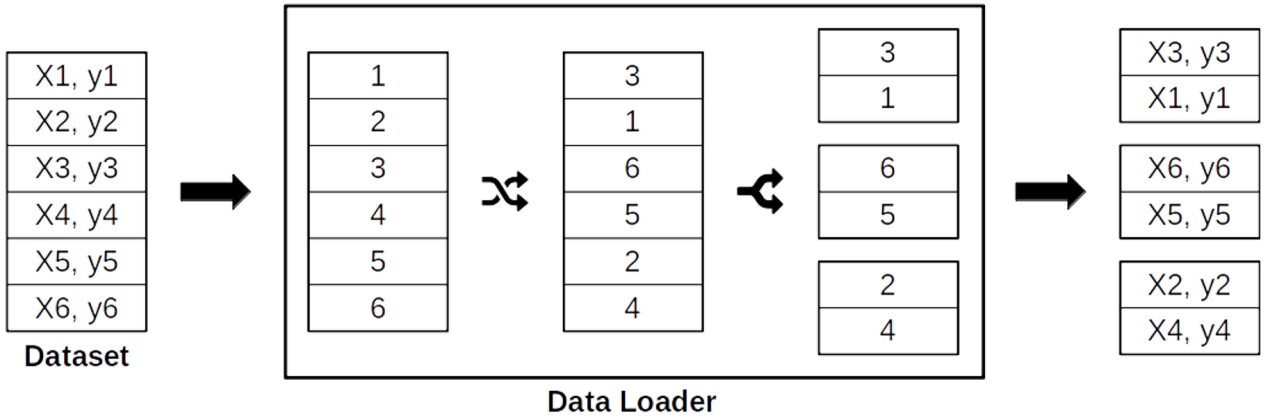
Actually important!
(For later, not the class)

- Newly included this semester
  - Memory only datasets vs actually reading every file from disk

    Usually you can't do the former but it provides exceptional performance boosts when applicable

Exercise 3: Dataloader

- Defines how to load the dataset for model training
- Shuffles the dataset
- Splits the dataset into small subsets



Overview Exercise 3

- Two notebooks
  - Dataset: CIFAR10
  - Dataloader

Fixed Deadline!
May 23, 2022 23:59

- First "real" Submission
  - Have to implement parts of both objects
  - Single submission file creation in Dataloder notebook

See you next week