PROGRAMMING IN PYTHON II

Project Design and Outline

Michael Widrich
Institute for Machine Learning

JOHANNES KEPLER
UNIVERSITY LINZ

Institute for
Machine Learning

Outline

1. Project Design

2. Python II Project

PROJECT DESIGN

Project Design

- When designing an ML project, you have to consider and constantly re-evaluate multiple aspects

Project Design

- When designing an ML project, you have to consider and constantly re-evaluate multiple aspects
- Common important aspects (in my experience) as checklist:
  1. What is the project goal?
  2. What data do you have? What data do you need? What does the data look like?
  3. What hardware do you have? What hardware could you have?
  4. What ML method(s) should you use?
  5. How to evaluate the methods/models?

Project Design

- When designing an ML project, you have to consider and constantly re-evaluate multiple aspects
- Common important aspects (in my experience) as checklist:
  1. What is the project goal?
  2. What data do you have? What data do you need? What does the data look like?
  3. What hardware do you have? What hardware could you have?
  4. What ML method(s) should you use?
  5. How to evaluate the methods/models?
- There is no one-fits-all solution! Specific tasks require specific considerations!

Goal

- What is the project goal?

Goal

- What is the project goal?
- Very important aspect and often overlooked
- Requires communication with people from different fields, including management
- DO NOT make simplifications here! Make sure you are aware of the real goal and communicate this!

Data

- What data do you have? What data do you need? What does the data look like?

Data

- What data do you have? What data do you need? What does the data look like?
- Sometimes the goals will follow from existing data
- Perform analysis of the data (e.g. clustering) and look for possible issues (e.g. biases, batch-effects)
- Talk to experts in the field/read up on the topic
- Check if there is auxilliary data available (pre-training on similar data, unused sorted out data, . . .)
- Perform preprocessing of the data (normalization, oversampling, cross-validation splits)

Hardware/Software

- What hardware/software do you have? What hardware/software could you have?

Hardware/Software

- What hardware/software do you have? What hardware/software could you have?
- CPU, GPU, or TPU based?
- Size of RAM and disk storage?
- Hardware compatible with ML software? Software restrictions from company/collaborations?
- Short term or long project?

Methods

- What ML method(s) should you use?

Methods

- What ML method(s) should you use?
- Depends on goal, data, and hardware
- You will need a theoretical understanding of the methods to judge which ones to consider
  - Literature research
  - Later semesters of AI study
- Start with baselines/less complex methods and models
  - Statistics, logistic regression, SVM,
  - Check Supervised Learning before Reinforcement Learning and Unsupervised Learning

Evaluation

- How to evaluate the methods/models?

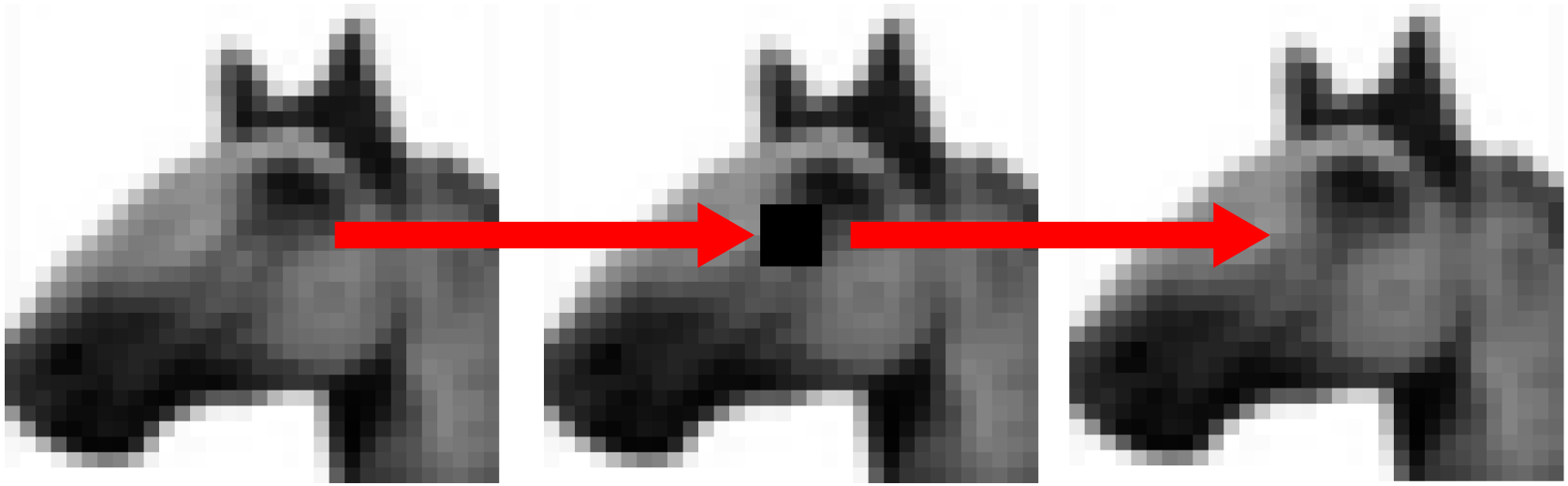
Evaluation

- How to evaluate the methods/models?
- Which score/performance measure?
- Do you need to correct for biases?
- Which aspects of the goal are more important?
- What do you want to generalize to?

PYTHON II PROJECT

Python II Project: Goal

- Restore cropped-out parts of images
- Data imputation task



Python II Project: Data

- We will create our own dataset
- JPG images up to 850kB
- 100 images per student
  - → roughly 35k images
- We will crop out small areas of the images
  - → we do not need to collect labels!
- Evaluation on testset with different images
- We will perform analysis and preprocessing of the data

Python II Project: Hardware/Software and Methods

- Hardware/Software
  - Harware is up to you (see introduction slides)
  - Python 3.6 or higher
  - PyTorch
- Methods
  - Simple Convolutional Neural Network (CNN)
  - You may also use other NN types/more complex settings
  - Design and fine-tuning is up to you

Python II Project: Evaluation

- We will have a challenge server where you can evaluate your model on testsets
- Public testset and private (final) testset
- Mean squared error over predicted pixels as performance measure

Python II Project

- More information during semester