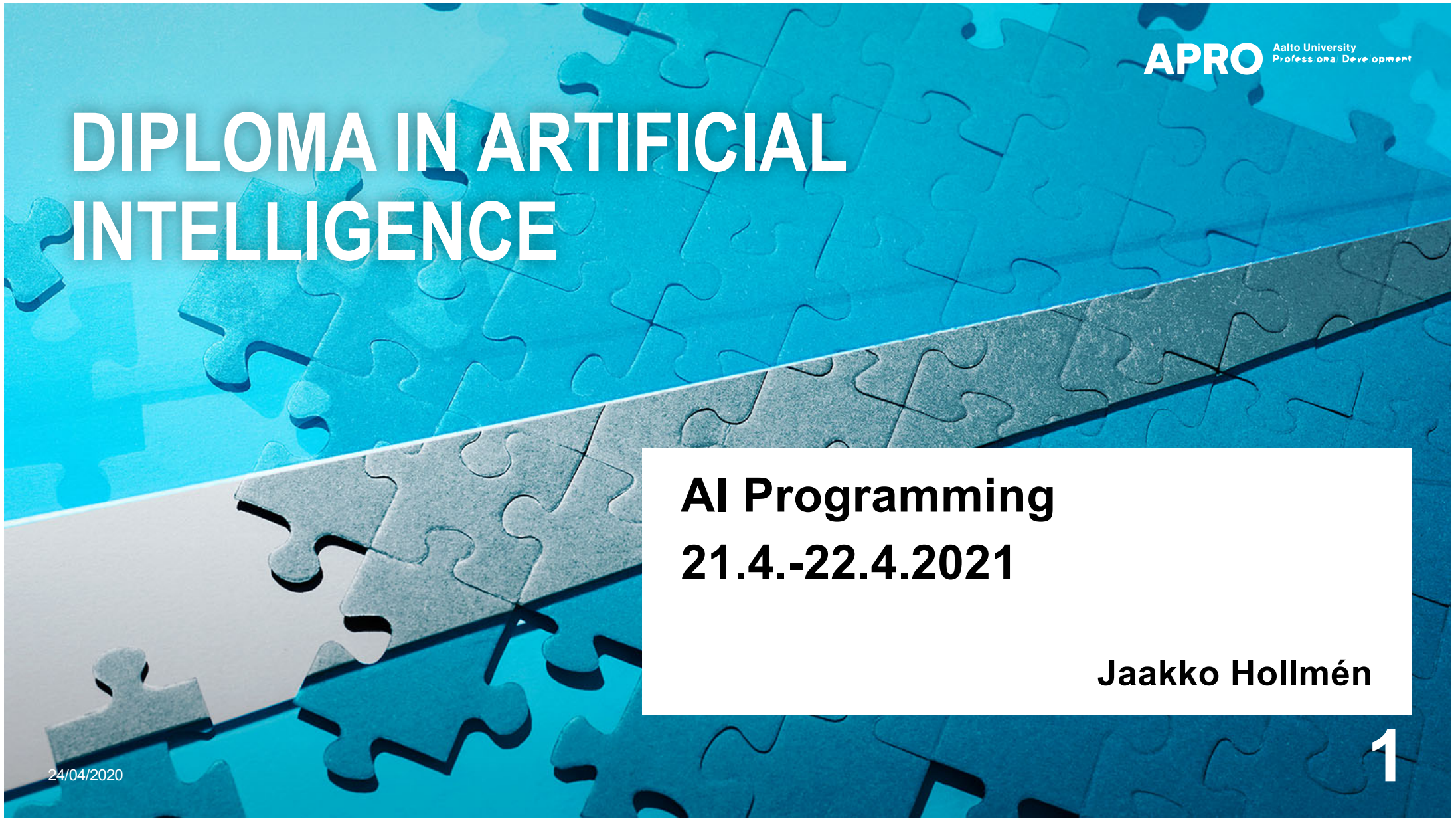
APRO
Aalto University
Professional Development
DIPLOMA IN ARTIFICIAL INTELLIGENCE
AI Programming
21.4.-22.4.2021
Jaakko Hollmén
24/04/2020
1

AI PROGRAMMING

The two-day training is divided into thematic sessions, where problems are presented to the students and when students create the solutions to the problems during the session.

Each student should have a computer and preparedness to run Python programs in Jupyter notebooks.

Each session contains a brief introductory lecture to the topic, and description of the programming exercise. Then, student proceed by programming, either alone or in pairs. Towards the end of the session, solutions will be reviewed.

SESSION 6: CROSS-VALIDATION TECHNIQUES

- This session will be covered on the second day during 10.45-12.00.

- The learning objective is get familiar with the concept of generalization and why it is impossible to optimize directly to best generalization. We also get familiar with some widely used form of cross-validation

LECTURE CONTENTS

- Generalization
- Cross-validation techniques
- Description of the exercises

GENERALIZATION

- Generalization refers to the ability to generalize to unseen cases
- In time series context, future is unseen. Still, the model needs to perform well in the future
- In medical context, the generalization would mean the ability to diagnose unseen patients correctly

- You can not optimize the model with regard to an unknown quantity
- Modeling becomes a tradeoff between accuracy to *this* data set and the generality

CROSS-VALIDATION TECHNIQUES

- We can try to simulate the unseen future by leaving out some parts of the data sample
- Train on the training set, test on the validation set
- Training error can always be reduced by choosing a more complex model
- See both performances and select an appropriate trade-off between complexity/accuracy and generality

CROSS-VALIDATION TECHNIQUES

- Leave-one-out crossvalidation technique
- K-fold cross-validation
- Cross-validation repeated
- 10-fold cross-validation repeated 20 times
- Bootstrapping

EXERCISES

- The exercises are listed in the Jupyter notebook Session-6-cross-validation.ipynb
- Work one exercise at the time
- Not all exercises need to be completed

REVIEW OF THE SOLUTIONS

- How do the solutions look like?