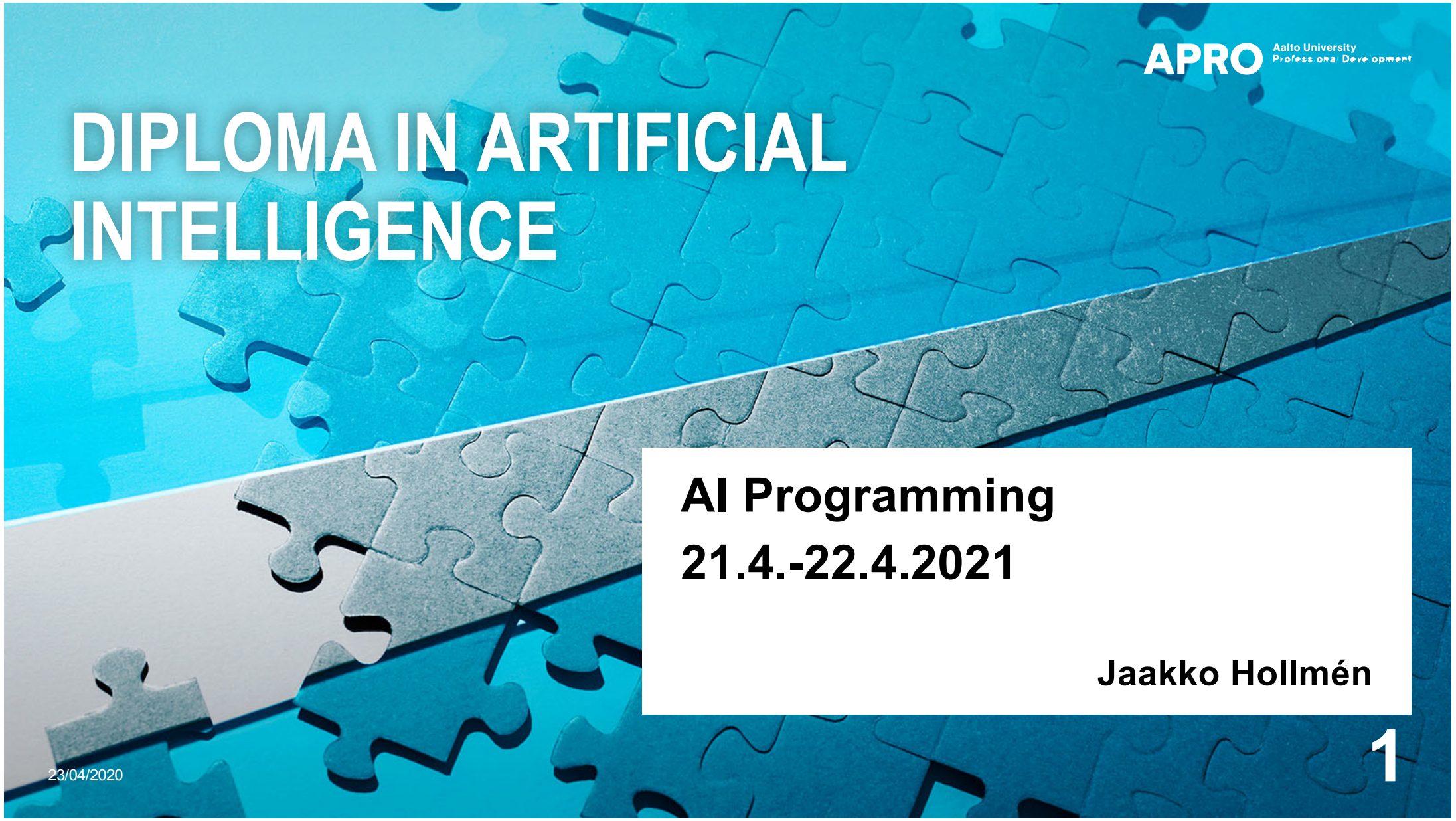
APRO
Aalto University
Professional Development
DIPLOMA IN ARTIFICIAL INTELLIGENCE
AI Programming
21.4.-22.4.2021
Jaakko Hollmén
23/04/2020
1

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 2: SELECTING APPROPRIATE METHODS

- This session will be covered on the first day during 10.40-12.00

- The learning objective is get familiar with some typical data analysis scenarios and to think of some practical constraints in selecting the analysis approach

LECTURE CONTENTS

- A quick run-through of typical data analysis scenarios
- Constraints from the data and implications for defining a task
- How to find a Python library or implementation for a task?
- Description of the exercises

DATA ANALYSIS SCENARIOS

- Clustering: grouping of entities, result is the emergent grouping
- Classification: Based on entities and pre-existing categorization, learn to assign entities to classes
- Prediction: estimation of unknown quantities, for instance future values of time series. Learning to predict is based on known pairs of inputs and outputs.

- What is the ”business problem”?
- What data is available? Does the lack of data affect my definition of ”business problem”?

COST OF ACQUIRING DATA

- Acquiring labels for data needed for tasks like classification, or prediction often requires expertise and is costly
- Diagnoses made by a medical doctor, for instance
- Often accumulated in the course of an operation, like the health care system is storing the decisions made by clinical doctors

- Crowdsourcing as one solution of acquiring labeled data

HOW TO FIND A PYTHON LIBRARY

- Assume a given form of a problem, such as classification or clustering.
- Now you are faced with the problem of either making an implementation, or finding a ready implementation of it or a library support this approach
- Use widely used implementation
- You are still responsible for the correctness of results!

- How to find a Python library?
- Use official documentation: https://www.python.org/doc/
- Use commonly known libraries: Numpy, Scipy, Pandas, Scikit-learn

EXERCISES

- The exercises are listed in the Jupyter notebook Session-2-and-3-Selecting-tools-languages-and-methods.ipynb
- Work one exercise at the time
- Not all exercises need to be completed

REVIEW OF THE SOLUTIONS

- How do the solutions look like?