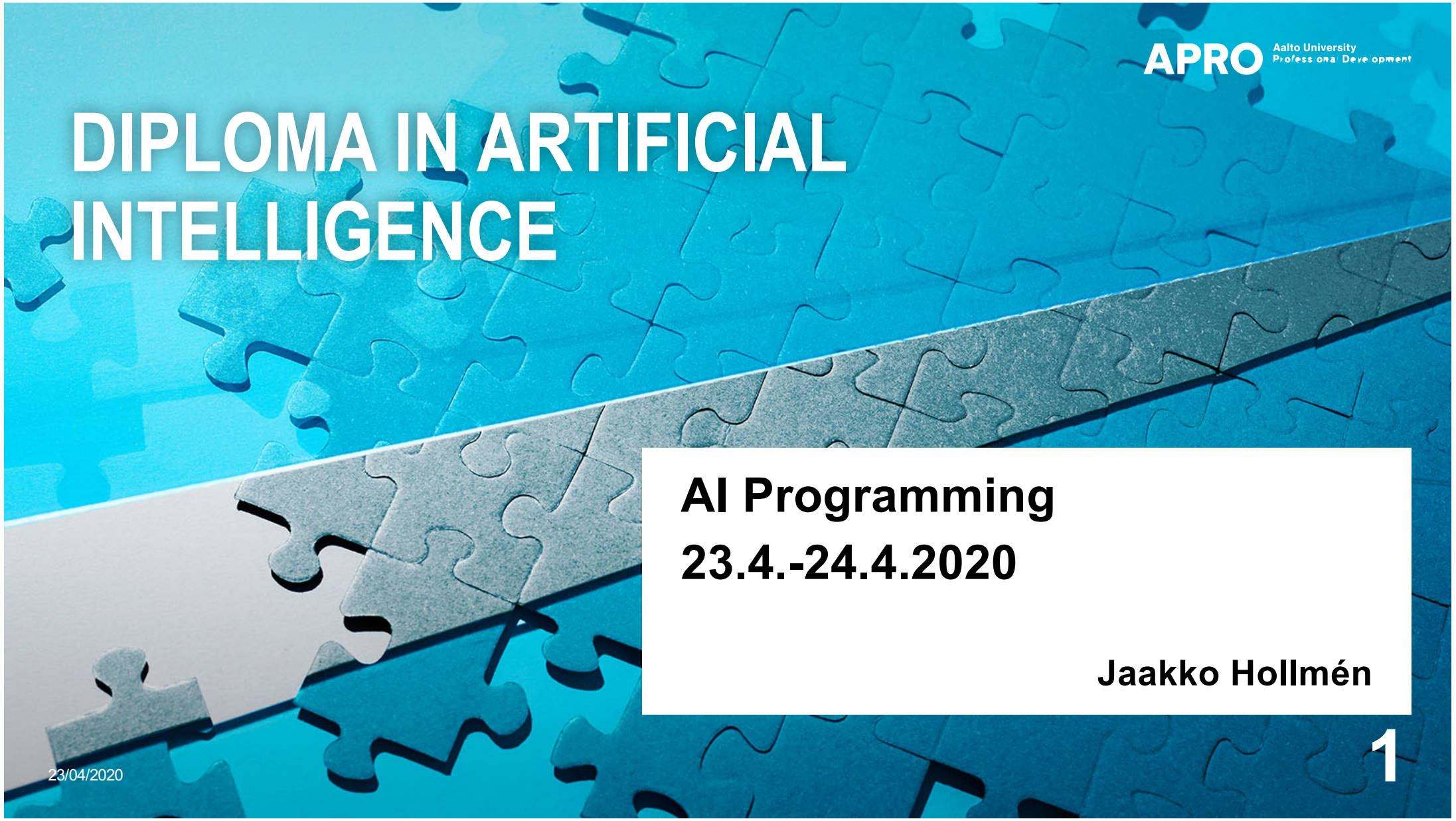
APRO
Aalto University
Professional Development
DIPLOMA IN ARTIFICIAL INTELLIGENCE
AI Programming
23.4.-24.4.2020
Jaakko Hollmén
23/04/2020
1

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.

APRO Aalto University Professional Development

SESSION 1: TEXT ANALYSIS

- This session will be covered on the first day during 10.40-12.15.
- The learning objective is get familiar with reading text from data files and learn how to represent text as vectors for further analysis

LECTURE CONTENTS

- Written text as data: special characteristics of text
- Vector space representations of text
- Description of the exercises

LANGUAGE AND TEXT

- Language is relatively free, but computer representations of data are strictly defined and constrained
- Think of the idea of constraining a language to ”Sentences with exactly seven words” ?!??

- Allow for the free characteristics of language and define a representation that is compatible with the computer world
- Written language consists of words: concetrate on the statistical properties of words in text, count word occurences

LANGUAGE AND TEXT

- ”Mary had a little lamb. The lamb was…”
- Count the number of occurences of words: A: 1, had: 1, lamb: 2, Mary: 1, the:1, was:1
- Just list the set of words (independent of the number of times): {A, had, lamb, Mary, the, was} which is equivalent of binarizing the number of occurences: 1, had: 1, lamb: 1, Mary: 1, the:1, was:1
- Weight the words according to commonality or importance
- Inherent problem: the number of words grows rapidly, typically tens of thousands
- Does the words ”A” or ”The” carry meaning? Not really..

EXERCISES

- The exercises are listed in the Jupyter notebook Session-2-Text-analysis.ipynb
- Work one exercise at the time
- Not all exercises need to be completed

REVIEW OF THE SOLUTIONS

- How do the solutions look like?