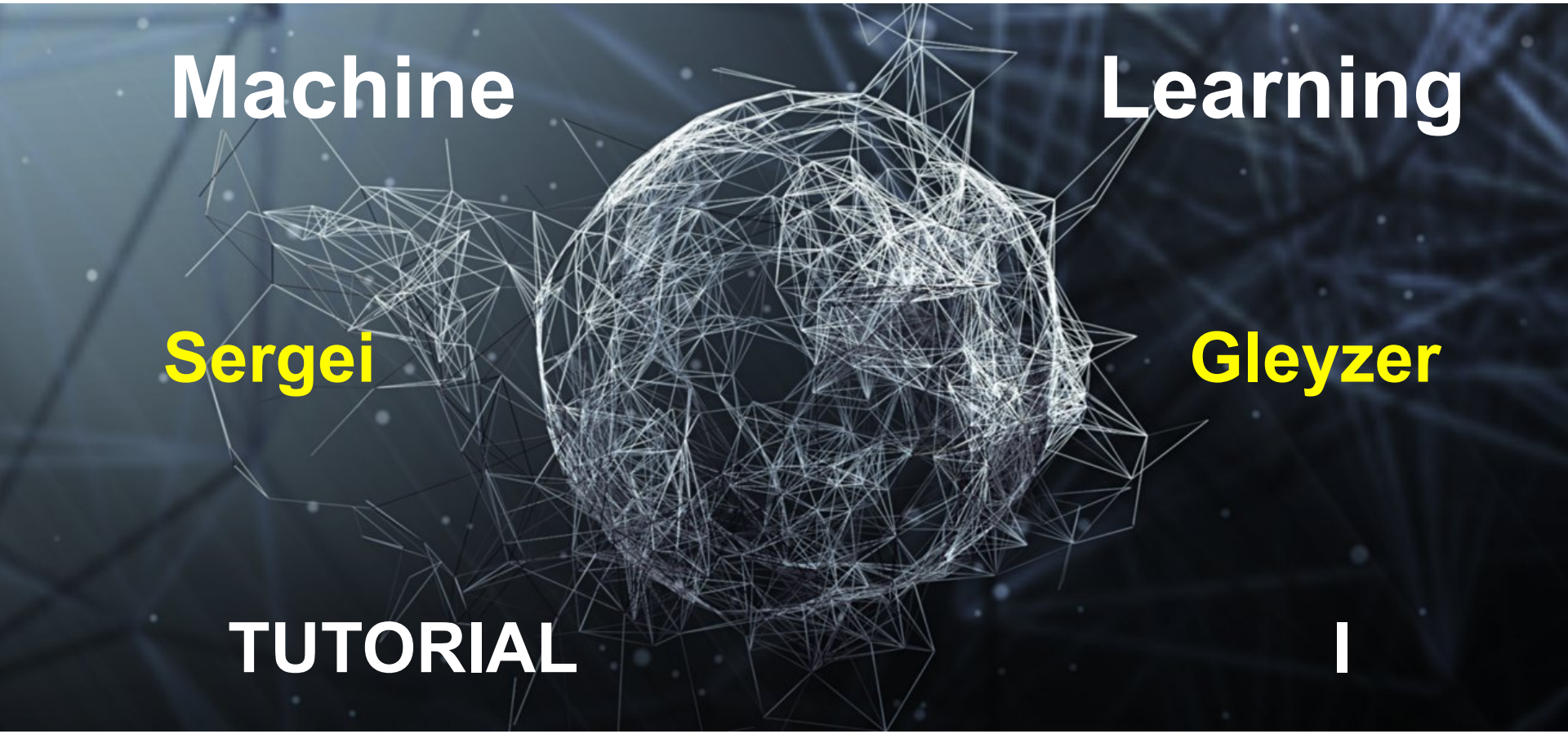


Tutorial

April 24, 2019

Today's Goal

Build our first **classifier**

Today: **tree-based**

Later: **deep neural networks**

Jupyter Notebooks

Open source **web-based** application blends code with elements such as text, figures, links

- Blend of instructions and executable code
- Can be run locally, on a server, laptop or smartphone
  - All you need is a browser

SWAN

Service for Web-based Analysis

http://swan.web.cern.ch

Jupyter notebook service

Today will use only the terminal

Swan Login

http://swan00Y.cern.ch Y= 1, 2,3,4

Username: mlearn[XX] XX=3..30

Password: distributed

Instructions

Open a terminal >_

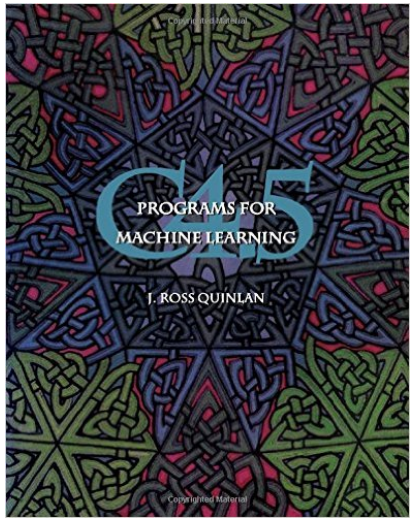
Check out the tutorial:

git clone https://github.com/iml-wg/tutorials

Go to folder C50

C5.0

- Classic **ML tool** for
  - **decision trees**
  - **rules**
  - **boosted classifiers**
- Written by **J.R. Quinlan**
  - Name: ID3 → C4.5 → C5.0
    - Use c5.0 to familiarize with decision tree based classifiers



Tutorial Part I

Examples: playing golf, breast-cancer

- Create your first classifiers
  - **Decision trees**
    - c5.0 –f golf
    - c5.0 –f breast-cancer

    Needed:

    .names file that includes the names of classes and variables, and variable types(continuous/discrete)

    .data file with variable values

Tutorial Part I

- Look at **Decision Tree** structure(s)
- Consider **accuracy** of predictions
  - Prediction errors
    - on **training** examples
    - on **testing** examples
  - Understand **confusion** matrix

```
 (a)   (b)    <-classified as
----  ----
 125     5    (a): class 2
   6    63    (b): class 4
```

Decision Rules

Decision Rules:

- Deconstruct **Decision Tree**
- Set of **if – then – else** rules
  - Example of **"weak"** learners (better than random guessing)
  - In weighted ensemble become a competitive classifier
    - RuleFit: Friedman, Popescu, 2005

Tutorial Part II

Examples: playing golf, breast-cancer

- Create your first classifiers
  - **Rules**
    - c5.0 –r –f golf
    - c5.0 –r –f breast-cancer
  - Compare Rule(s) to Decision Tree(s)
    - Note: all decision trees are rules but not all rules are trees