TU 257 – Fundamentals of Data Science

Data Analytics

Lab 7 – Tuning & AutoML

Brendan Tierney

Agenda

- Exercise 1 – Demo Notebook 1 – Grid Search
- Exercise 2 – Demo Notebook 2 – Automating the Process
- Exercise 3 – Demo Notebook 3 - AutoML
- Assignment

- Reminder
  - For all Labs you should be adding your own notes to them
    - Using Python Comments
    - Using cells as Markdown

  - Add your own notes/comments – What makes sense to you
  - This will help with your Understanding of what is happening
  - Why? What does this Show? What do the results mean?
  - …

Exercise 1 – Demo Notebook 1
Exercise 1 – Demo Notebook 2
Exercise 1 – Demo Notebook 3

Same instructions apply to each Demo Notebook

Exercise 1 – Demo Notebook 1, 2 & 3

- Examples/Demo Notebook
  - Download the Demo Notebook
  - Download the Dataset needed for Demo Notebook

  - Run all cells
  - Examine the output generated
  - Follow what is happening from cell-to-cell
  - Notebook might contain some additional Questions/Exercises for you to do
  - Add extra annotations/descriptions based on your understanding
    - Enrich the descriptions to make them more meaningful for you

AutoML install/setup

- This can be a little challenging in Anaconda
- Some of these AutoML libraries need specific versions of other libraries
  - These might not be what you have installed!

- Create a new Anaconda Virtual Environment
  - Install the AutoML into it
  - Here are some blog posts illustrating this
    - Installing PyCaret in Anaconda
    - Pycaret Installation Documentation

Similar needed for autosklearn

- Although some might work in your current Anaconda environment

tpot - It if isn't listed in available list of libraries to install, run the following

```
conda install -c conda-forge tpot
```

Some Blog Posts

- AutoML, what is it good for? It Depends!
- AutoML – using TPOT
- AutoML – using autosklearn in Python
- AutoML using Pycaret
- OML4Py – AutoML – Step-by-Step Approach

See Installation Tip on previous slide

Assignment A

Assignment

- Assignment Time
- Groups of 2 people
    - Can be similar to your group in Data Wrangling module
- Based on Demos and Exercises from this week and next week
    - Build upon your learning
    - Using a different dataset
- See Assignment Handout in BrightSpace for more details
    - Read carefully and make sure you do everything as specified
- See Template for doc/notebook for completing the assignment.
    - This is what you will submit
    - You will be graded on what it contains
    - Add all code, comments, observations, discussions, explanations, etc.
    - Demonstrate understanding, discussion, additional exploration

Assignment

- Start working/thinking about it early -> today, this week
- Review what we have covered todate
- What can you do to showcase your work and understanding of the topics
- What can you add to the data analysis
- …

Any Questions ?

What Now/Next ?

Complete **all** Lab Exercises before Next Week