184.702 Machine Learning – Exercise 3

The third exercise allows you to choose from different topics, all dealing with advanced and particular topics in machine learning. The exercise shall be done in groups of 3 students, which may be the same as in the first two exercises. Please arrange with your colleagues, and then register for a group **and** for your topic in TUWEL.

If you have an interesting idea of any other topic, please don't hesitate to contact me (mayer@ifs.tuwien.ac.at) and discuss it with us. If we see it fitting to the topics of the course, you'll get our ok to work on your idea.

There is an online submission of the report and other artefacts that you created during this exercise (code, data, ...), **and if you did NOT present exercise 2, also a presentation.**

The descriptions of the different topics are likely not a complete specifications, they just summarise the general idea and goals. If there is an unclarity, please do not hesitate to ask in the forum.

As the more specialised tasks are also less explored and thus less predictable in outcome, we will not grade you based on results alone, but also your efforts.

In general, you have three types of topics available, and you need to chose **one** of these (some types, especially 3.2, have further sub-options from which again you need to chose only one).

3.1.: Advanced topics in security / privacy of Machine Learning

3.2.: Deep Learning for Image/Text Classification

3.3.: Auto ML - implementation of search using simulated annealing

General comments for the exercise

This exercise is supposed to be solved with all the steps being programmatically, i.e. there shouldn't be a need for GUI or some other manual processing of data (rather scripted).

- A zip file with all needed files (your source code, your code compiled, data sets used (but **NOT** the ones **we provide** to you), a build script that resolves dependencies, or include any libraries you are using. Your submission needs to be self-contained!
    - If applicable, provide a means to compile your classes, preferably with a build file (e.g. Maven or ANT if you use Java)
    - If your build file allows for automatically obtaining the dependencies, then they don't need to be included, otherwise please include them. If using python, consider using a virtual environment or similar that can easily be recreated.
    - Provide a short how-to explaining the way to start your program (which is the main file, which command-line options does it expect, ..).
    - Please also state clearly what is the version of the software package(s) you use; unless for a specific reason, please work with the latest versions (you can of course e.g. work in python 2, but don't work with old versions of libraries etc.).
      Again, make sure dependencies are either packaged along, or are easily resolved (build file, virtual environment, etc., whatever applicable for your approach)
- A report, describing
    - Your task
    - Your solutions (also describe failed approaches!)
    - Your results, evaluated on different datasets, parameters, ...
    - An analysis of the results
- Where applicably, your program shall be configurable via command-line options or a configuration file, to modify parameters, evaluation types, etc… i.e. it should not be needed to modify the code to change these options. There are many framework for command-line-options available, you don't need to code this yourself (e.g. for Java, Apache Commons CLI (http://commons.apache.org/cli/, http://martiansoftware.com/jsap/).

Computing resources

If you are in the need of computing resources for topics 3.1 or 3.2, you can contact mayer@ifs.tuwien.ac.at especially for **CPU** power on a Linux server @TU.

Further, you can try infrastructure-as-a-service offers, as some Cloud operators have some free starter credit. E.g the Google Cloud Computing platform gives you **300 USD if you register** (you need to provide your credit card, but they don't charge) which you can use to create remote machines with various OS preinstalled. Also other providers have similar offers. Make sure to shut down your machines when not using them, to avoid being charged.