CHALLENGE

This year's programming assignment will be organized as a contest. Teams of two persons will "compete" in the following AI task:

"predict the composer, year, instrument, key and style for each song."

Cooperation between teams is encouraged, a chat room has been set up in Slack to allow you to communicate. Everyone should have received an invitation from Slack, drop me an email if you haven't!

https://vub-ai-lab.slack.com

Data set

The dataset is based on the "Jazzomat Research Project", which consists of 299 jazz songs and associated metadata. More information on the data, and the meaning of the fields can be found at:

http://jazzomat.hfm-weimar.de/dbformat/dbcontent.html

At this site, you can listen to the songs, and see a graphical "piano-roll" representation so you can gain insight in the data. I will provide a set of CSV file with the data. Exactly 5 songs will be retained for each artist in order to obtain a balanced dataset (and thus only artists with at least 5 songs will be included).

Test script

To be able to compare the results between the teams, every team has to commit executable code every week, in the form of a script (e.g. "yourProgram") that is executed as follows:

```
./yourProgram training-data-file test-data-file output-file
```

Obviously, the test-data-file may not be used for training. 5-fold cross-validation will be used for validation and calculating the confusion matrix. The output-file should be a CSV file containing the song id (from the dataset input) and composer, year, instrument,key and style predictions in separate fields (in this order) and in the same encoding as the input dataset-file.

Evaluation criteria

The individual team performance will not count very much in the evaluation of the programming assignment. However, the diversity of solutions and techniques used will, as well as team play, active involvement in the forum, elegance of the solution, innovation, creativity, steady improvement, scientific soundness and rigor of the literature review.

PHASE 1 – deadline April 8th, 2016

In a first phase, which takes only one week, you should come up with a computational representation for the problem. This may not yet be in executable code, but in a formal format with appropriate explanations for understanding (two pages max). Make sure the representation is adapted to the problem, and do stand on work done in the scientific community: explaining your choice is essential. Also, **the diversity of representations between teams should be maximized:** in an ideal case, no two teams should start from the same representation!

PHASE 2 – weekly commits

Starting the week of the 11th of April, the "programming" starts. We will work by short development cycles of one week, which means that **by the end of *each week*, you should send me a running prototype** of the code according to the criteria above.

In a first week, this could be just a mock up reading the data and randomly generating outputs (or, just a bit smarter, according to the output distribution). It is allowed to discuss problems, literature, solutions, issues with the other teams (I will set up a chat channel) and learn from each other. However, it is important that everyone sticks (except minor revisions) to the representation chosen in phase 1.

PHASE 3 – hand in report: June 3rd, 2016

At the end of the semester, every team has to hand in a short report (10 pages max.!!) describing their approach, role, lessons learnt, and future work.