IST Austria: Data Science and Scientific Computing
Track Core Course 2015/16 – Segment 2: "Predictive Models"

Instructor: Christoph Lampert <chl@ist.ac.at>
TAs: Anna Levina <alevina@ist.ac.at>, Srdjan Sarikas <ssarikas@ist.ac.at>

Final Project

Project Description

The goal is to learn a model that can predict in which year a song was released from its audio characteristics.

Procedure

1) form small working groups, ideally different ones than for the previous project. Each group should contain at most 2 participants who want ECTS points, but a third participants can join who just audits the course. Groups can talk to each other and share experience, but each group must produce it's own model and presentation.

2) visit the *"YearPredictionMSD" dataset* webpage `http://archive.ics.uci.edu/ml/datasets/YearPrediction` read the description and download the data.

3) delete the last 51630 examples (test data) from the data file on your computer

4) use the remaining (trainval) data to build a predictive model of the year (first column) from the remaining features. You can use any methods, software, etc. you want.

5) when you have selected a final model, send an email to the TAs declaring that your model is final

6) afterwards, load the dataset again from the given URL, extract the test part and evaluate the quality of your model

7) prepare a short team presentation to be given on April 27th that explains all steps you did to reach the model and why you made them. Report the results you achieved and put them into context (are they good? what would a trivial baseline method achieve, e.g. the data mean?).

8) for every team member separately and without copying from your team partners(!): write a short manuscript (3 to 5 pages) that presents your method as the new state-of-the-art for dating songs. The style should be of a *scientific manuscript* with *Abstract*, *Introduction*, *Method* section, *Experimental Evaluation* and *Conclusion*. It should explain to other researchers in the music dating community what they should do from now on to identify which year a song was released in. Do not describe the process ("we did then, then we did that"), but present the outcome as a scientific discovery. If you are unfamiliar with this style or just want further tips, grab a copy of *[H. Glasman-Deal, "Science Research Writing"]* from the library.
The manuscript is due on May 4th (you can send it earlier, if you like), please send it directly to `chl@ist.ac.at`. I will make corrections/suggestions and return them to you within a few days. The revised manuscript will be due a week after I gave you the comments. Beware that this process might repeat more than once.

Time Frame

- April 20: official begin of project period
- April 25: no lecture, Q&A about the project
- April 27: team presentations about the projects
- May 4: manuscripts are due