Dissecting Spotify Valence

In this assignment you will dissect Spotify's Valence metric.

> Panos Louridas, Associate Professor
> Department of Management Science and Technology
> Athens University of Economics and Business
> louridas@aueb.gr

Spotify uses a metric called *valence* to measure the happiness of a track. The metric itself, however, was not developed by Spotify. It was originally developed by Echo Nest, a company that was bought by Spotify in 2014. We don't know exactly how valence is calculated. Some details are given by a blog post, which you can find here:

https://web.archive.org/web/20170422195736/http://blog.echonest.com/post/66097438564/plotting-musics-emotional-valence-1950-2013

Your task is to untangle the mystery behind valence and propose how this is derived.

Spotify offers the following information that may be relevant to your task:

- Get Track's Audio Features and Get Tracks' Audio Features.
- Get Track's Audio Analysis.

To tackle the problem you can use the Spotify charts data from Zenodo at https://doi.org/10.5281/zenodo.2594556, but you are not limitted to that. In fact you are encouraged to use additional data that you may find, or get yourself.

The rank of your assignment will contribute to the grade. That is, if x is the unranked grade of an assignment and r is the respective model's ranking among n students according to Q2 below, then the final grade will be computed as $0.75x + 2.5[1 - (r - 1)/(n - 1)]$.

The ranking will be performed on a test dataset that will be provided in due course. The rank metrics will be Mean Average Error (MAE).

Questions

Q1: Expore which Track Features Influence Valence

You will use inferential statistic methods to study how track features influence valence. You must find the best possible model for explaining the valence based on the features that you find significant.

Q2: Predict Valence

Use Machine Learning techniques to predict valence based on track features:

- You will use at least three different methods. For each methods you should ensure that you tune your hyperparameters as best as you can.
- Once you identify the best method and hyperparameters, explain, to the extent that is possible, which features influence the valence metric.
- You will evaluate your predictions on a holdout testing dataset that will be provided to you. Your evaluation and the value of the MAE on the holdout testing dataset must be included at the end of your submission.

Submission Instructions

You will submit a Jupyter notebook that will contain all your code and analysis. Ensure that the notebook will run correctly in a computer that is not your own. That means, among other things, that it does not contain absolute paths. Remember that a notebook is not a collection of code cells thrown together; it should contain as much text as necessary for a person to understand what you are doing.

Honor Code

You understand that this is an individual assignment, and as such you must carry it out alone. You may seek help on the Internet, by Googling or searching in StackOverflow for general questions pertaining to the use of Python, libraries, and idioms. However, it is not right to ask direct questions that relate to the assignment and where people will actually solve your problem by answering them. You may discuss with your fellow students in order to better understand the questions, if they are not clear enough, but you should not ask them to share their answers with you, or to help you by giving specific advice.