Задача

Ваша задача — научиться искать похожих музыкальных исполнителей.

В терминах машинного обучения это означает, что для каждого исполнителя нужно построить векторное представление таким образом, чтобы похожие исполнители оказались близко в векторном пространстве, тогда задача поиска похожих исполнителей сведётся к задаче поиска ближайших соседей.

Оценивать качество полученных векторных представлений будем на задаче *Artist Recommendation* — рекомендация исполнителя для пользователя.

Данные

Ссылка: 30Music [http://recsys.deib.polimi.it/datasets/]

Датасет состоит из событий о разбитых на сессии прослушиваниях музыки пользователями сайта Last.fm.
Датасет также содержит много дополнительной информации о треках, исполнителях, альбомах и т. д.

Более подробное описание данных можно найти в оригинальной статье [http://ceur-ws.org/Vol-1441/recsys2015_poster13.pdf]. Формат файлов описан тут [https://github.com/crowdrec/idomaar/wiki/DATA-FORMAT].

Что требуется сделать

1. В первую очередь нужно подготовить данные: перейти от истории прослушиваний треков к истории прослушивания исполнителей (полезные файлы: sessions, tracks, persons).

2. Так как мы хотим сравнивать между собой разные алгоритмы на задаче *Artist Recommendation*, нужно определиться с протоколом оценки качества:
    - разбить данные на Train/Validation/Test,
    - выбрать метрики, которые будем использовать для оценки качества.

3. Реализовать несколько методов (простой baseline и более сложную модель) для построения векторных представлений.
   Будет плюсом, если вы придумаете, как использовать дополнительную информацию из датасета.

4. Обучить модели и подобрать гиперпараметры.

5. Сравнить качество реализованных методов. Будет здорово, если вы оцените статистическую значимость полученных результатов. Также для нескольких исполнителей нужно привести примеры топ-20 наиболее похожих.

Формат

Jupyter Notebook с воспроизводимым кодом и выводами. Важно, чтобы Notebook содержал обоснования принятых вами решений. Можно выложить на GitHub.