

## ПРАКТИЧЕСКОЕ ЗАДАНИЕ: КЛАССИФИКАЦИЯ МУЗЫКИ ПО ЖАНРАМ

сдачи **22.10.2016.**

## Мотивация для решения задачи

Музыку можно интуитивно разделить на "веселую" и "грустную" и вы можете достаточно точно узнать любимый жанр музыки. Музыкальные сервисы, такие как Spotify, Zvooq или Яндекс.Музыка, стараются понять, как треки, попадающие к ним от звукозаписывающих компаний, распределены по жанрам.

На входе у музыкальных сервисов есть файлы в разных форматах (mp3, ogg, flac, etc.), но это уже сжатые форматы, для нашего практического задания мы будем использовать формат WAV, так что вам не нужно думать о загрузке и преобразовании.

Вам, как специалистам по машинному обучению нужно построить алгоритм предсказания жанра WAV-файла.

---

## Описание данных и метрик

Каждый файл с данными, это pcl (pickle) файл с numpy.array, состоящий из:

1. WAV-файлы по строкам,
2. сэмплы по столбцам.

В качестве метрики используется micro-averaged F1:

$$F_1 = \frac{1}{N} \sum_i^N \frac{2 * P_i * R_i}{P_i + R_i}$$

где  $P_i$  и  $R_i$  - это точность и полнота предсказания для  $i$ -ого класса соответственно.

## Сдача задания и правила игры

1. Можно использовать любые реализации изученных на курсе моделей (и не только).
2. Нужно будет прислать отчет в виде ipython notebook с описанием экспериментов.
3. Задание оценивается максимальным баллом, если вы
  - (a) побили baseline + сдали задание вовремя,
  - (b) выполнили дополнительные задания про визуализацию;
  - (c) за первые места в каждой группе и в целом по курсу будут бонусные баллы.

## Методические указания

Для того, чтобы работать со звуком вам придется придумать, как обработать данные - для нейронных сетей или для любого другого классификатора. Попробуйте понять, как стоит работать с сэмплами, какие из последних ваших знаний можно применить для этого.

*Разница между списыванием и помощью товарища иногда едва различима. Мы искренне надеемся, что при любых сложностях вы можете обратиться к семинаристам и с их подсказками самостоятельно справиться с заданием. При зафиксированных случаях списывания (одинаковый код, решение задачи), баллы за задание будут обнулены всем участникам инцидента.*