

# Machine Learning, Advanced Topics, 6th Seminar

## Приложения современного машинного обучения в музыке

Valentin Malykh

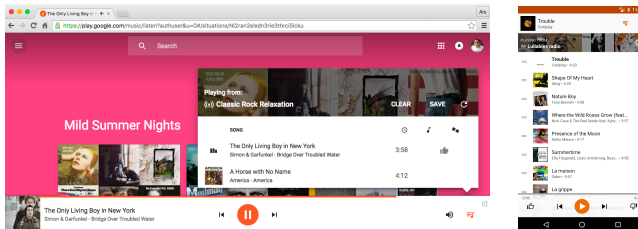
Moscow Institute of Physics and Technology

mail: [valentin.malykh@phystech.edu](mailto:valentin.malykh@phystech.edu)

12 октября 2016 г.

# How to apply ML for Music Data to get Money?

- ▶ Вы работаете data scientist в большом музыкальном сервисе



- ▶ У сервиса есть много музыкальных данных – mp3 files

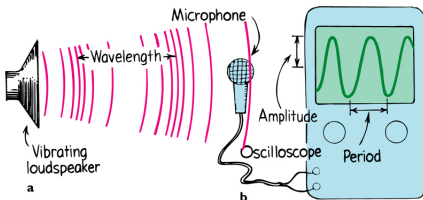
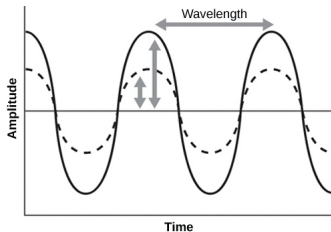
user_id	tracks_id
123	[1, 2, 3]
124	[1000, 11, 23, 23]
...	...
999999	[1]

tracks_id	file
1	1.mp3
2	2.mp3
...	...
999999	999999.mp3

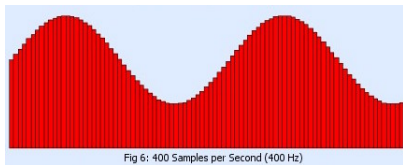
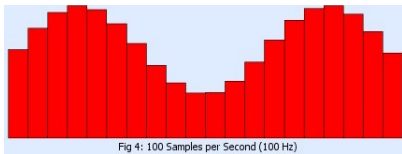
- ▶ У вас есть задача – получить деньги, используя эти данные

# Что такое звук?

## ► Waves and Recording



## ► Как сохранить звук? Как огромный массив данных



## ► [1, 2, 3, 5, 3, 2, 1, 1, 1, 1, 2, 3, 5, 3, 2], Обычно 16 000 float в секунду

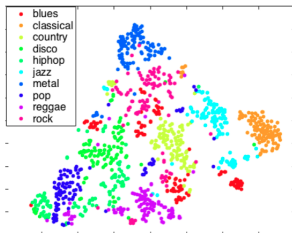
# Нахождение похожих треков

- ▶ Как найти похожие треки, используя ML-методы?

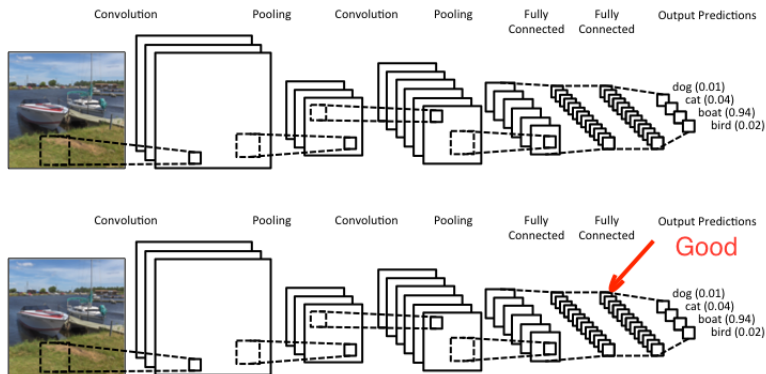
**Data:** 30 sec \* 16000 features,  $10^7$  items

**Task:** определить функцию  $\text{similarity}(\text{track}_i, \text{track}_j)$

- ▶ Почему обычные методы такие плохие?
  - ▶ смещение и шум, переобучение
- ▶ Метрический подход по-прежнему хорошая идея, если у нас высокоуровневое описание
- ▶ Хорошее представление музыкального трека
  - ▶ Human – guitar, rock, Queen, 1997, UK, 3 min., ....
  - ▶ Computer – good small vector of numbers



# Получим хорошее представление, используя Convolutional Neural Nets

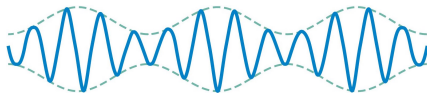


Issue

We need to get picture!

## Что такое звук? (2)

У нас есть некоторая волна



представим эту волну как сумму двух волн

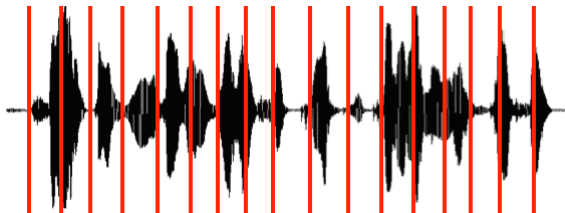


звук — комбинация различных волн



Что мы потеряли в нашем представлении?

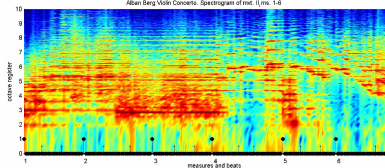
# Получение частоты



High Freq	1	2	1	1	2	1	1	2	1	1	2	1	1	2	1	1	2	1
Mid Freq	2	1	2	2	1	2	2	1	2	2	1	2	2	1	2	2	1	2
Low Freq	1	2	2	1	2	2	1	2	2	1	2	2	1	2	2	1	2	2

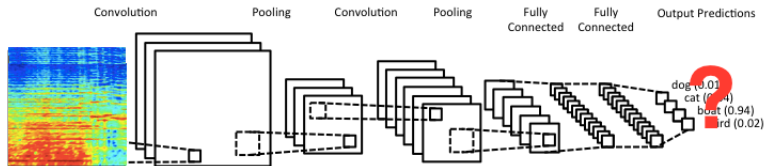
High Freq	1	2	1	1	2	1	1	2	1	1	2	1	1	2	1	1	2	1
Mid Freq	2	1	2	2	1	2	2	1	2	2	1	2	2	1	2	2	1	2
Low Freq	1	2	2	1	2	2	1	2	2	1	2	2	1	2	2	1	2	2

Alben Berg Violin Concerto, Spectrogram of mvt. II, mts. 1-8



# Мы должны обучить нейронную сеть, но как мы можем сделать это?!

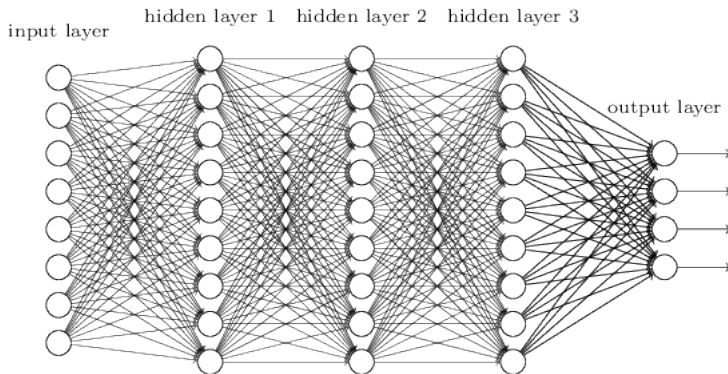
- ▶ Но как мы обучим сеть на музыке?
- ▶ Давайте подадим ей картинку



- ▶ классификация по жанру
- ▶ классификация по исполнителям
- ▶ предсказание рейтинга
- ▶ .....



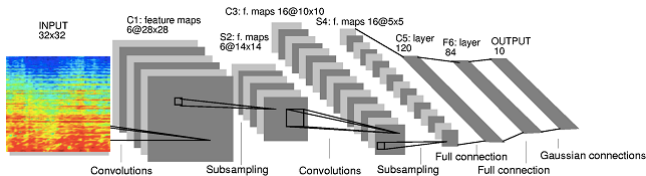
# Полносвязные NN



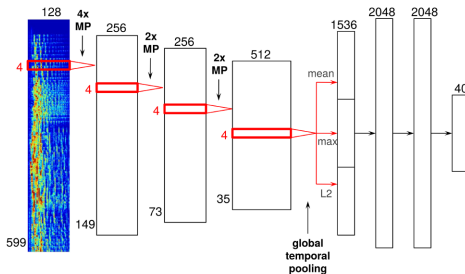
- ▶ слишком много параметров – число весов =  $16^4 * neurons + \dots$
- ▶ It doesn't work =(

# Convolution NN

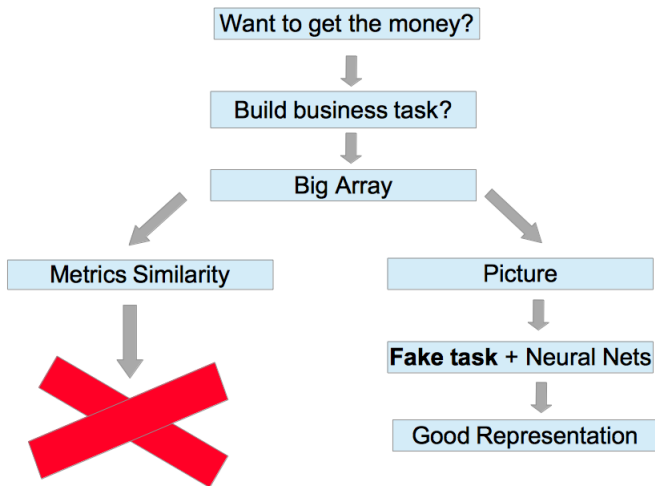
Давайте возьмем некоторую сверточную архитектуру



важная деталь – свертка только по оси времени [Spotify Deep Learning]



# General scheme, what did we do?



# Как измерять качество хорошего представления?

Что мы имеем?

- ▶ Мы имеем представление каждого трека в виде вектора
- ▶ Но м.б. наше решение слишком плохое, как мы можем это понять?
- ▶ Как тестировать "хорошее представление"?

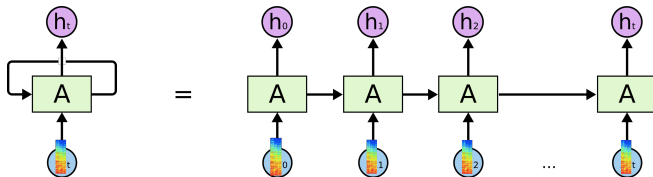
Давайте придумаем метрики:

- ▶ используя ассесоров
- ▶ качество рекомендации
- ▶ используя векторы для классификации других меток

# Давайте адаптируемся к различной длине и дополнительной информации

Как использовать различную длину?:

1. Усреднение предсказаний для многих кусков
2. Рекуррентная сеть на многих кусках



3. Что еще?

Как учитывать?:

1. Текст

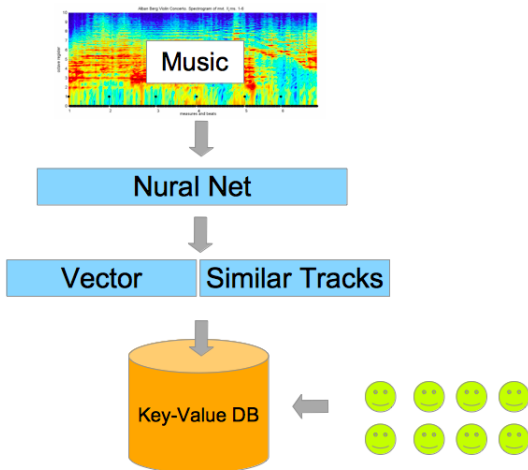
$\text{Concat}(\text{TextRNN}, \text{Conv}) \rightarrow \text{FC} \rightarrow \text{Cost}$

2. Жанр, исполнитель, год – embedding too, multi-cost task
3. ....

# Технические детали

Как построить быструю систему для миллионов пользователей?

1. предварительное вычисление векторов и симулирование треков
2. быстрое key-value хранилище



End



Current Status of your Field!  
Thanks for your attention!