

HSE Deep Learning Course

Application of Modern Machine Learning in Music

Ashuha Arseniy

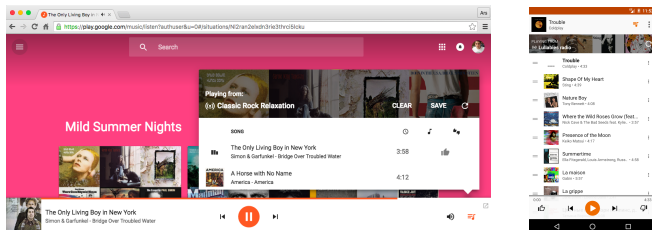
Moscow Institute of Physics and Technology

mail: ars.ashuha@gmail.com, slides: ars-ashuha.ru/slides

November 25, 2016

How to apply ML for Music Data to get Money?

- ▶ You are working in a big music service as a data scientist



- ▶ In this service there's a lot of music data – mp3 files

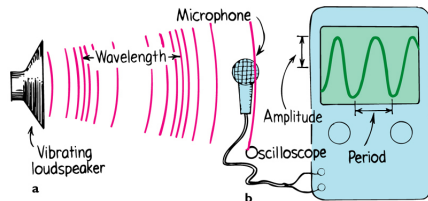
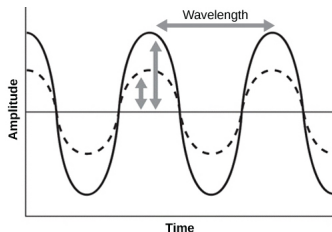
user_id	tracks_id
123	[1, 2, 3]
124	[1000, 11, 23, 23]
...	...
999999	[1]

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

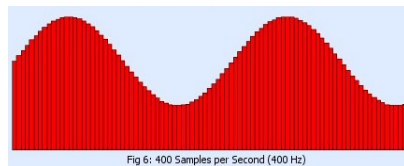
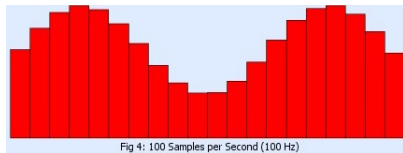
- ▶ You were given the task – make money using this data

What is the sound?

▶ Waves and Recording



▶ How to store sound? Store as big-big array with sampling frequency



▶ [1, 2, 3, 5, 3, 2, 1, 1, 1, 1, 2, 3, 5, 3, 2], Usually 16 000 float per second

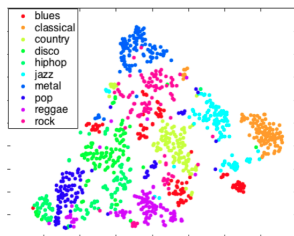
Finding similar tracks

- ▶ How to find similar tracks using ML methods?

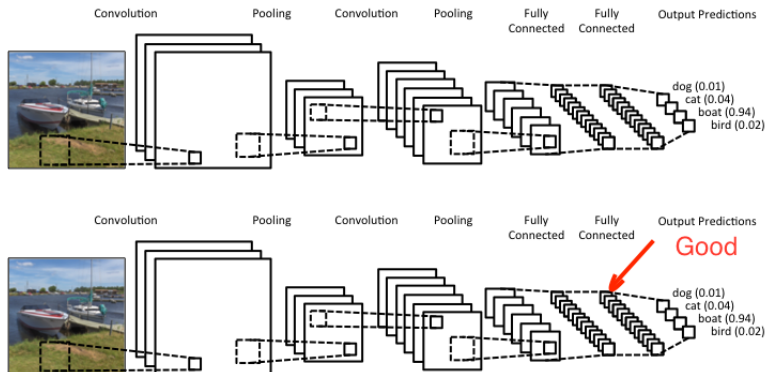
Data: 30 sec * 16000 features, 10^7 items

Task: define function of $\text{similarity}(\text{track}_i, \text{track}_j)$

- ▶ Why ordinary methods are so bad?
 - ▶ shift and noise tolerance, over-fitting
- ▶ Metric approach is still good idea, if we have a high level description
- ▶ Good representation of music track
 - ▶ Human – guitar, rock, Queen, 1997, UK, 3 min.,
 - ▶ Computer – good small vector of numbers



Get good representation using Neural Nets

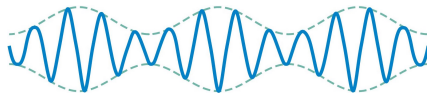


Problem

We need to get picture!

What is the sound part 2?

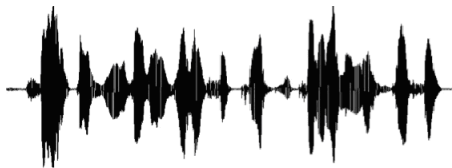
We have some wave



represent wave as a sum of two waves

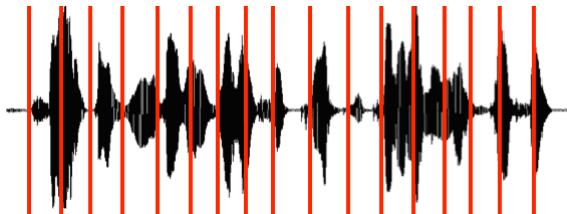


sound is a combination of big waves range



What we lost in our representation?

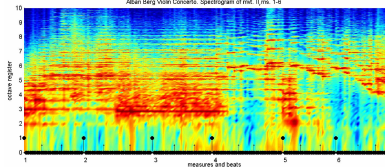
What is the sound part 2? Get Frequency



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

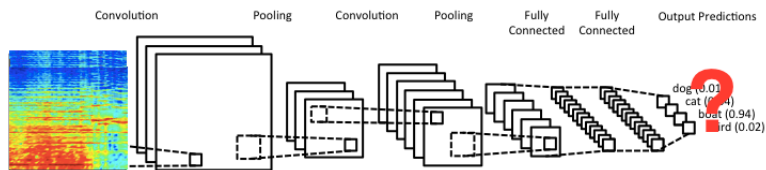
High Freq	1	2	1	1	2	1	1	2	1	1	2	1	1	2	1	1	2	1
Mid Freq	2	1	2	2	1	2	2	1	2	2	1	2	2	1	2	2	1	2
Low Freq	1	2	2	1	2	2	1	2	2	1	2	2	1	2	2	1	2	2

Alben Berg Violin Concerto, Spectrogram of mtr. I (ms. 1-8)



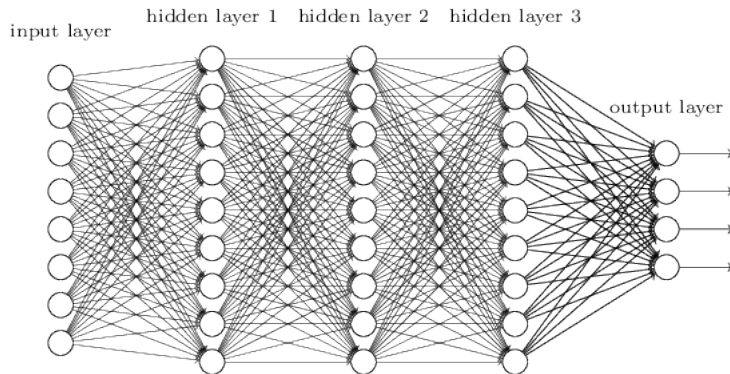
We need to train Neural Nets, but how can we do that?!

- ▶ But how can we train nets on music?
- ▶ Let's invent a fake machine learning task



- ▶ genre classification
- ▶ artist classification
- ▶ rating prediction
- ▶

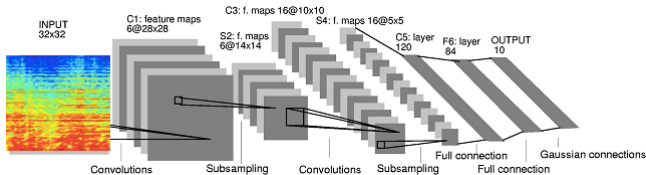
Fully connected NN



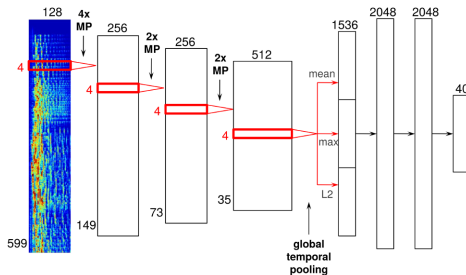
- ▶ too many parameters – number of weights = $16^4 * neurons + \dots$
- ▶ It doesn't work =(

Convolution NN

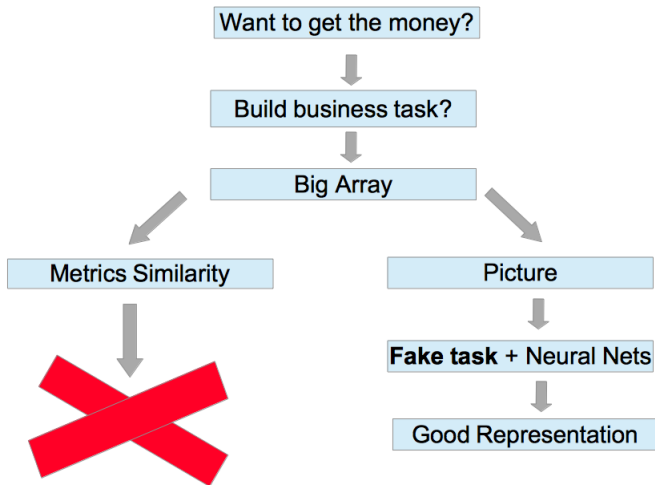
Let's invent some convolution architecture



important detail – pooling of time axis [[Spotify](#)))) Deep Learning]



General scheme, what did we do?



How to measure quality of good representation?

What we have?

- ▶ We have represented each track as a vector
- ▶ But maybe our solution is too bad, how can we understand that?
- ▶ How to test "good representation"?

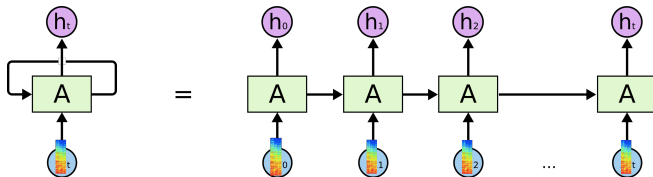
Let's invent the metrics:

- ▶ by hand
- ▶ using assessors
- ▶ recommendation quality
- ▶ using vectors to classify another labels

Let's adapt to Different length and Additional information

How to use any length?:

1. Average prediction for many patches
2. Recurrent neural net on many patches



3. Whatever?

How to take account?:

1. Lyrics

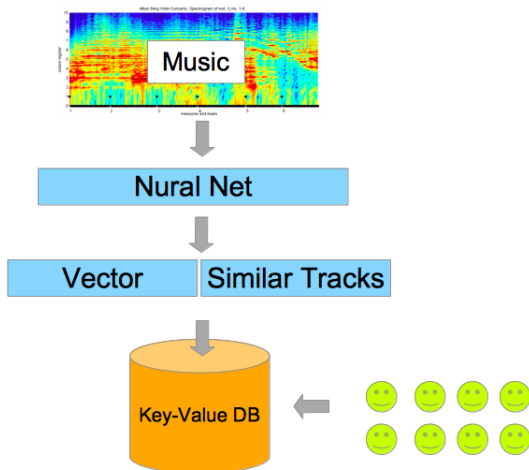
$\text{Concat}(\text{TextRNN}, \text{Conv}) \rightarrow \text{FC} \rightarrow \text{Cost}$

2. Genre, Artist, Year – embedding too, multi-cost task
3.

Technical details

How to build fast system for million users?

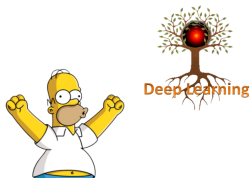
1. pre-compute vectors and tracks simulation
2. fast key-value storage



End



Current Status of your Field!
Thanks for your attention!



<https://goo.gl/wRZRgO>