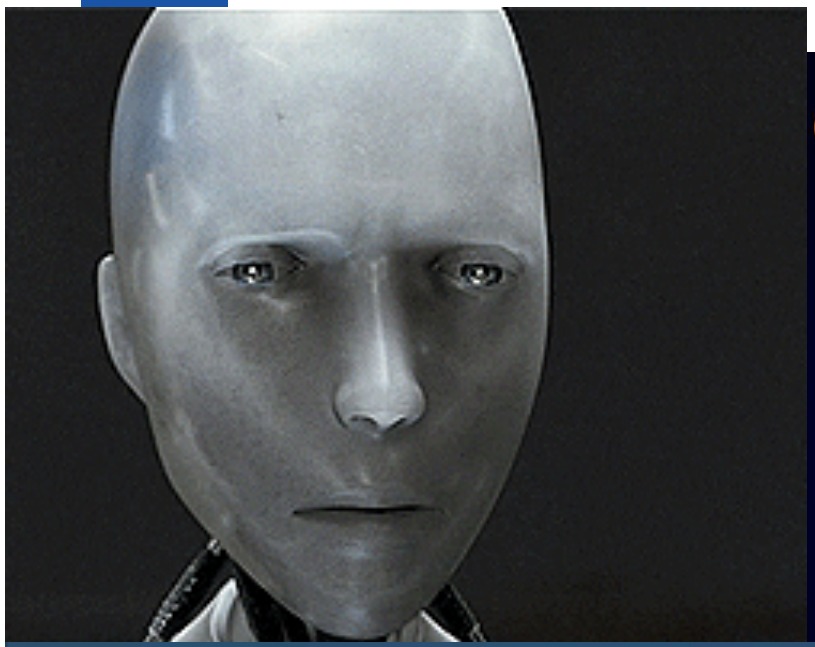
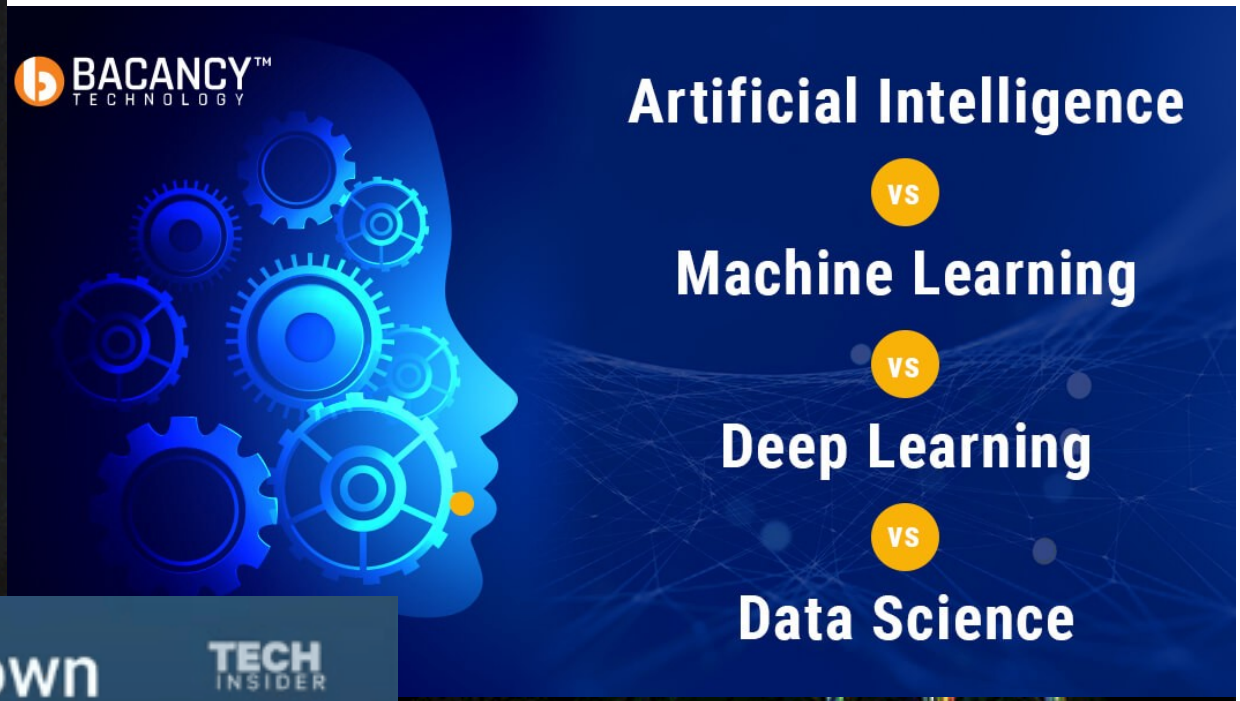
KTH ROYAL INSTITUTE
OF TECHNOLOGY

Overview of Machine Learning and Pattern Recognition

Bob L. T. Sturm (Speech, Music and Hearing)

What is machine learning? What is AI?





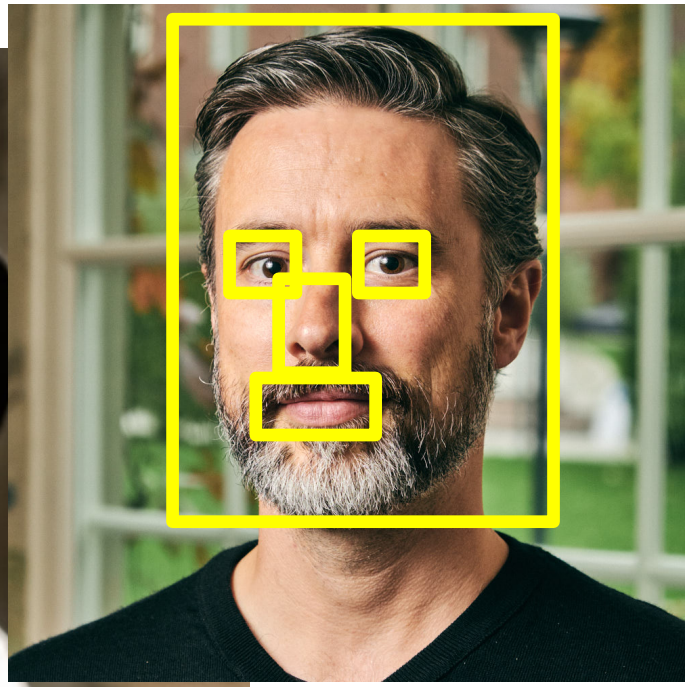
Consider taking a photograph

Is there a human face in this photograph?

Consider taking a photograph

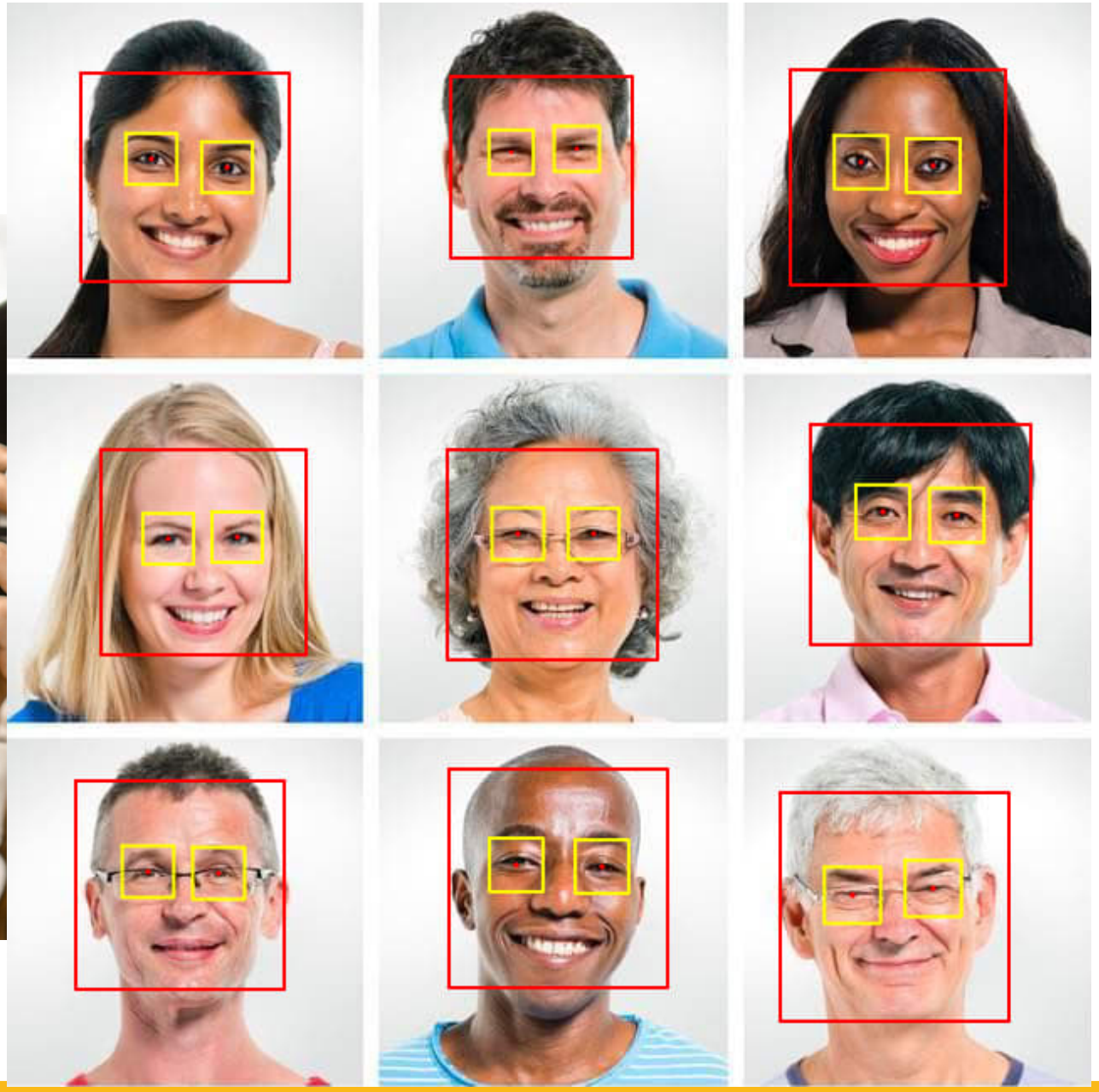
Is there a human face in this photograph?

Consider taking a photograph

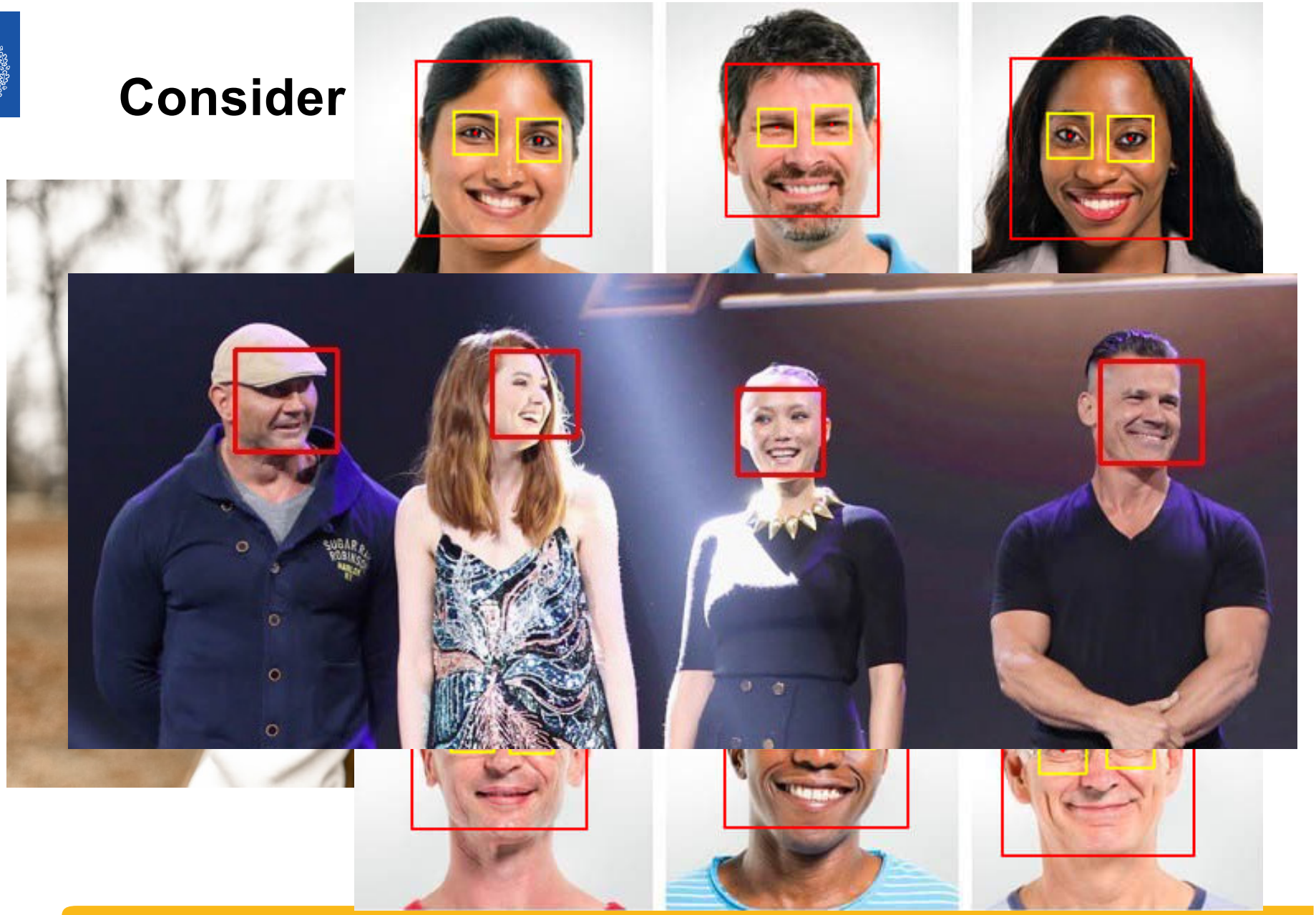


Is there a human face in this photograph?

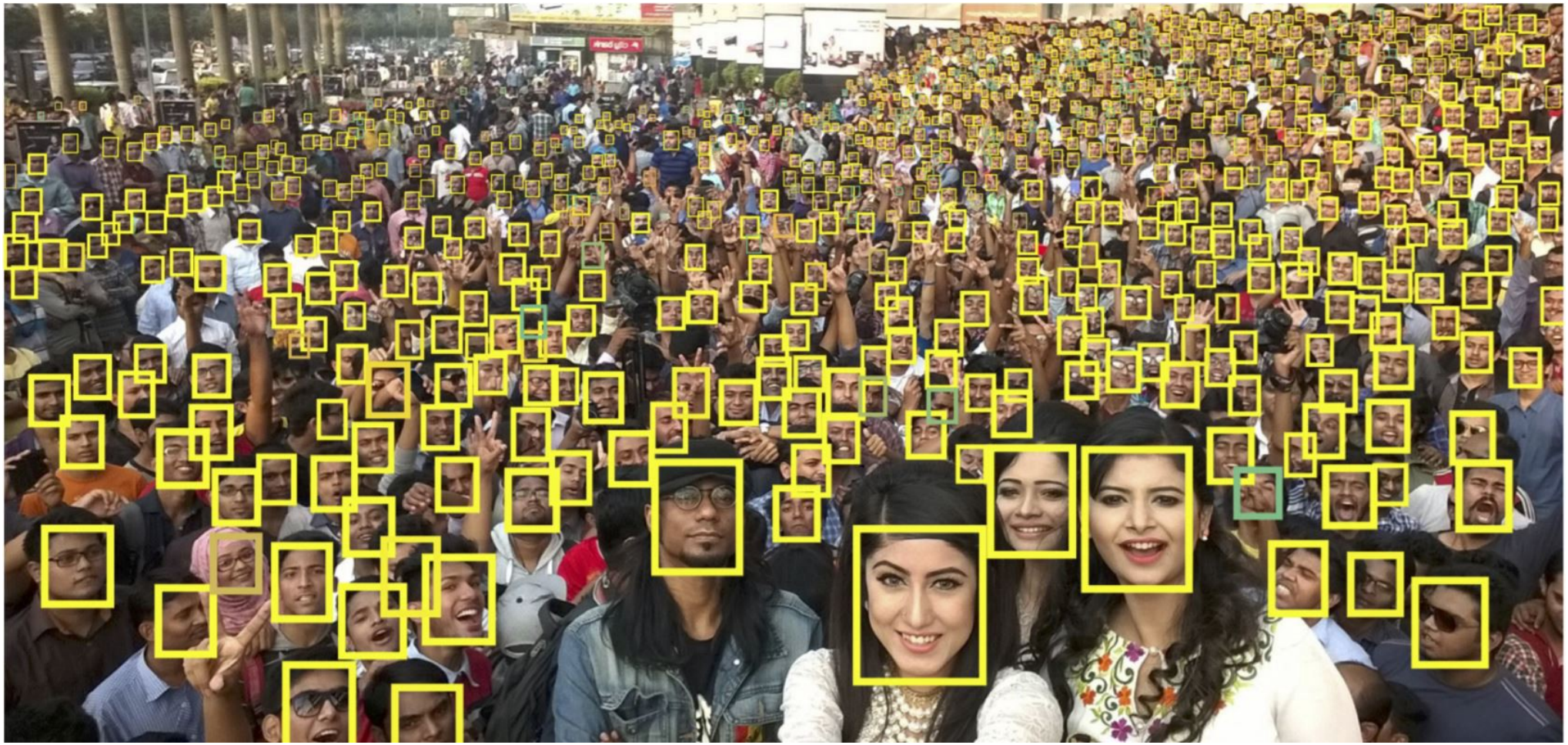
Consider



Consider



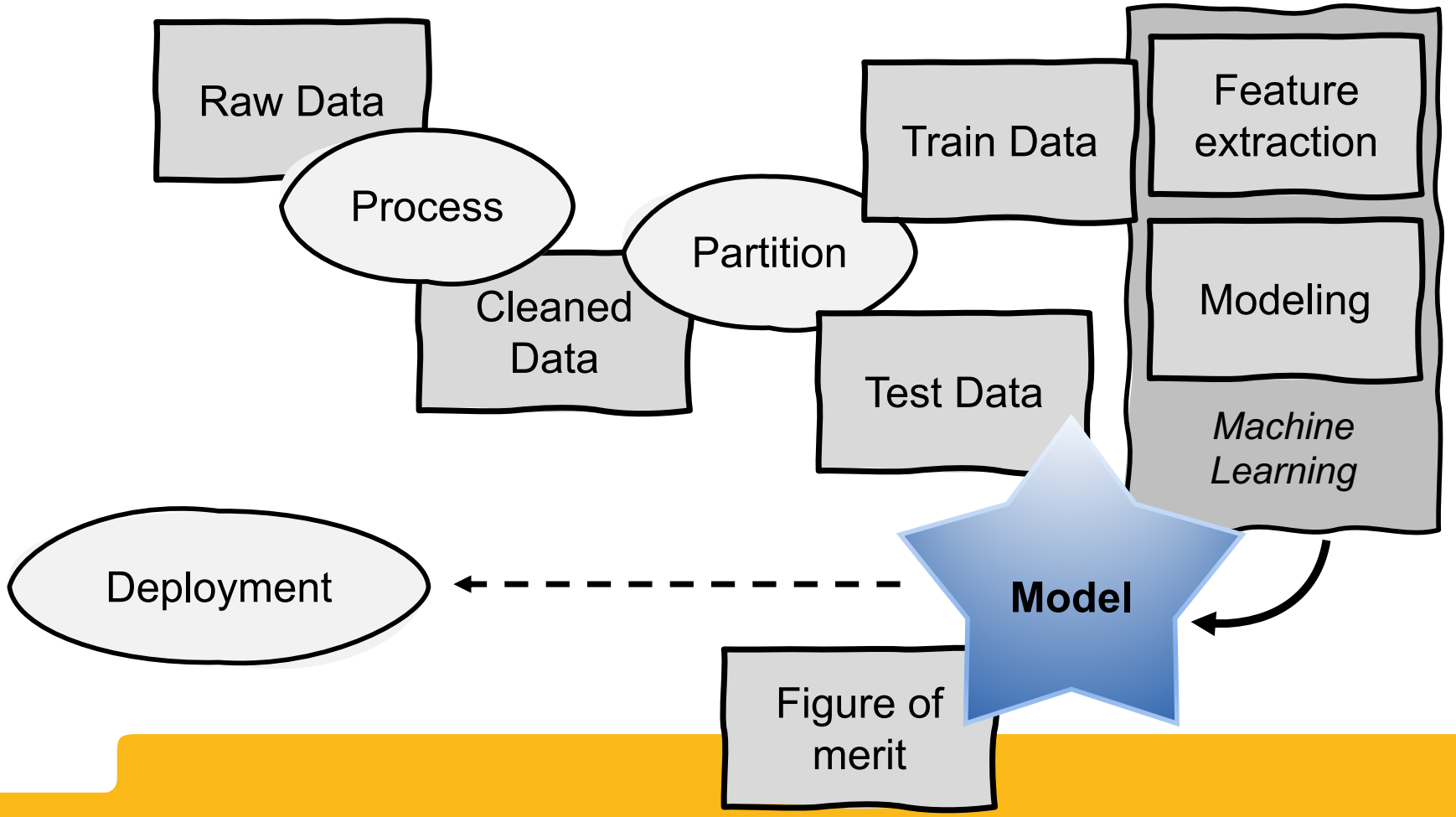
Consider taking a photograph



How many faces are there in this photograph?

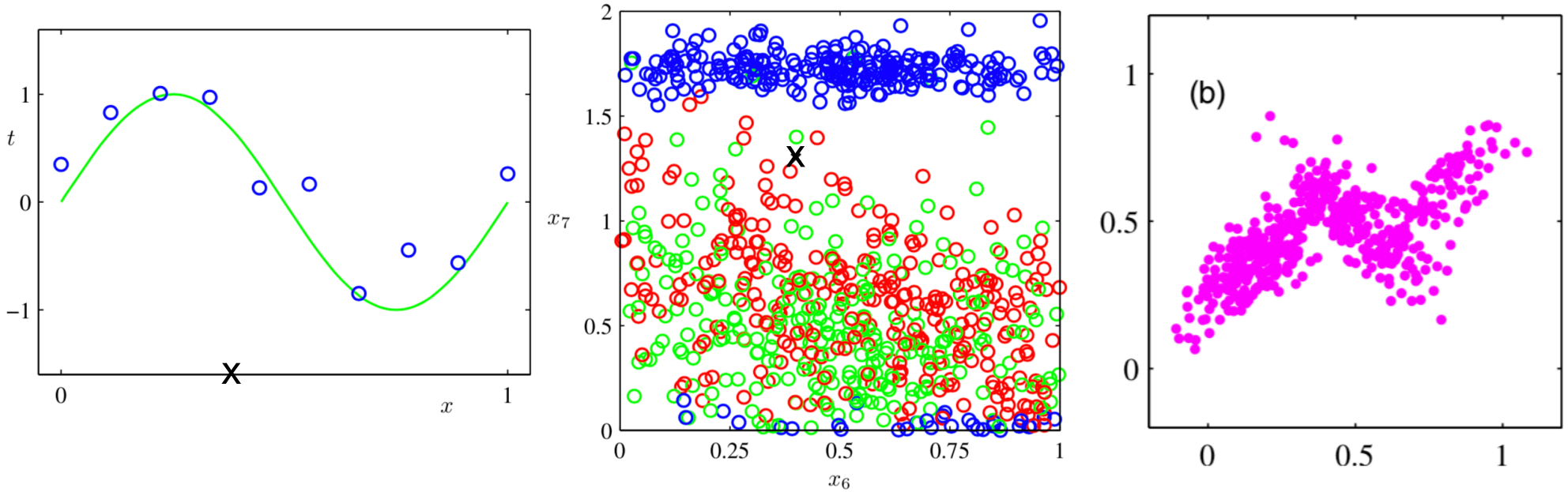
What is machine learning?

Methods for making machines learn from data.



The General Task

Given data, automatically learn about it.



Pattern Recognition: Engineering perspective

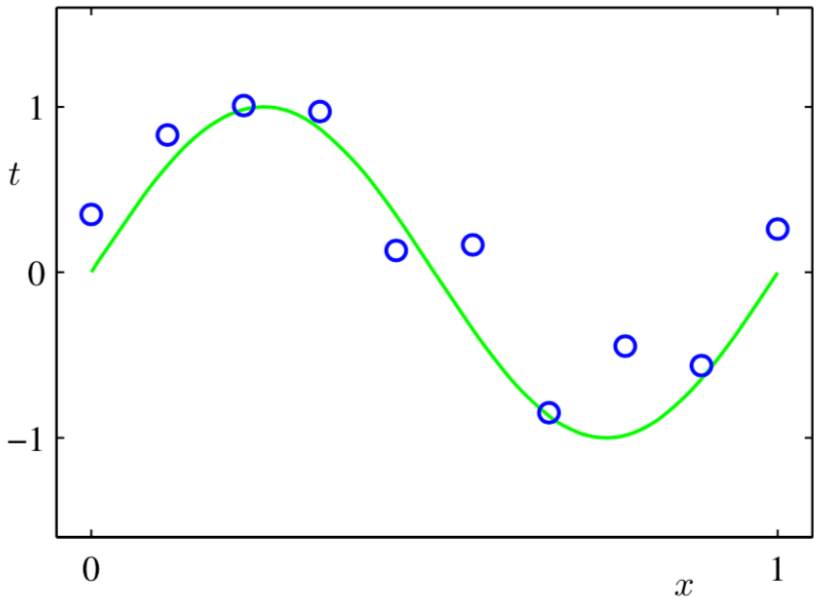
Machine Learning: Computer Science perspective

Regression

We are interested in the *relationship* between two variables, x and t.

For several x, we measure t:

$$((x_1, t_1), (x_2, t_2), \ldots, (x_i, t_i))$$



We hypothesize that these variables are related by some specific *function* y(x), e.g.,

$$y(x) = w_0 + w_1 x + w_2 x^2$$

Estimate the *parameters* (w0,w1,w2) and *evaluate* the model.

Regression Example

Regression Example

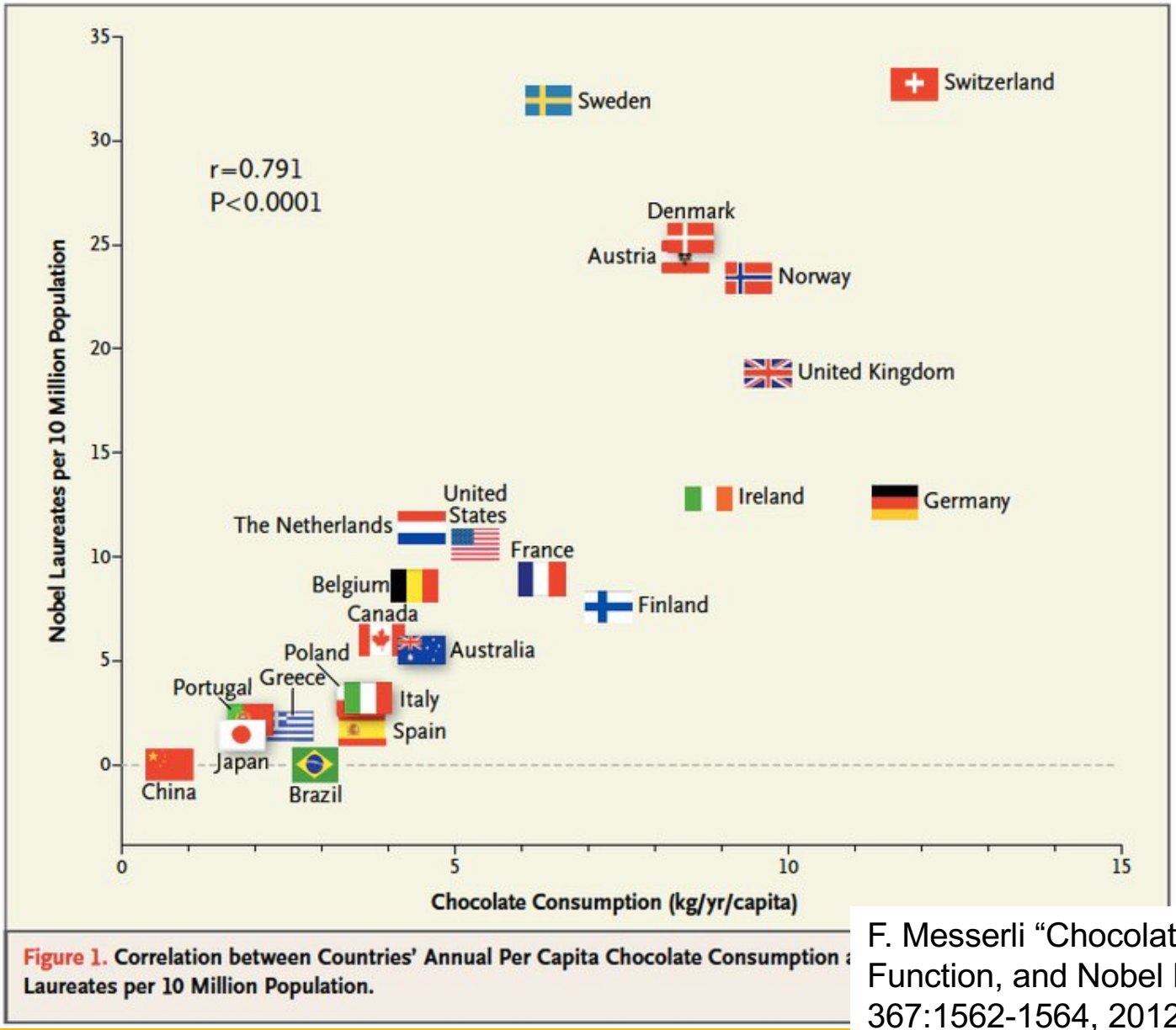


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption a
Laureates per 10 Million Population.

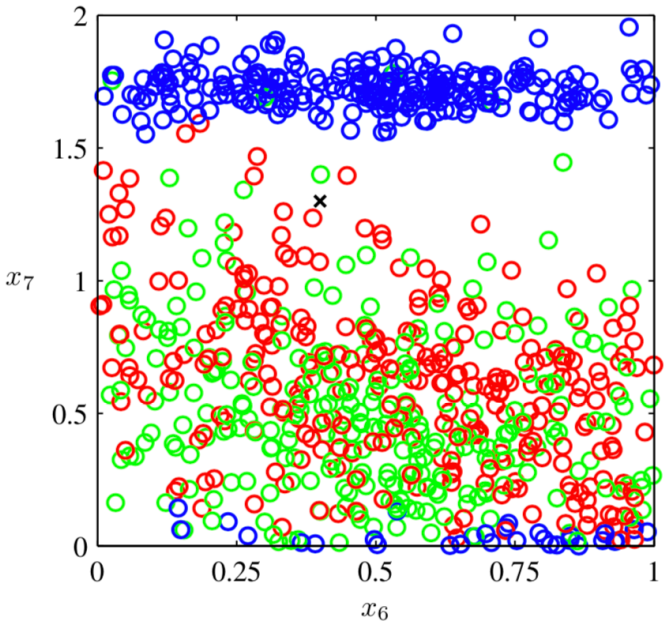
F. Messerli "Chocolate Consumption, Cognitive Function, and Nobel Laureates", *N Engl J Med* 367:1562-1564, 2012.

Classification

We are interested in building a model relating observations in two dimensions to three *classes*.

For each of several observations, we measure its class.

We hypothesize that the classes in this *feature space* relate to a contiguous *partitioning*:



scatter plot

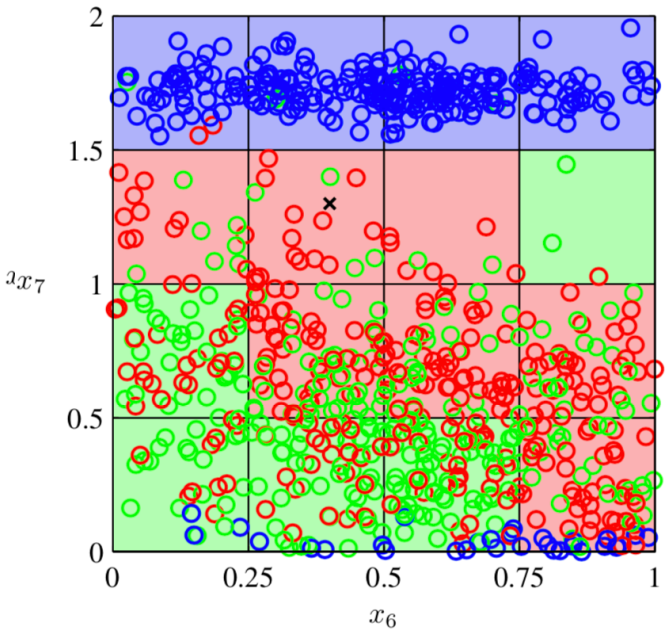
Classification

We are interested in building a model relating observations in two dimensions to three *classes*.

For each of several observations, we measure its class.

We hypothesize that the classes in this feature space relate to a contiguous *partitioning*.

We estimate this partitioning, and *evaluate* it.



scatter plot

Classification Example

Regression and Classification

Two sides of the same coin:

- Classification is regression with “discrete classes”
- Regression is classification with “continuous classes”

In both, we are choosing a class of functions (hypothesis) and estimating its parameters using data.

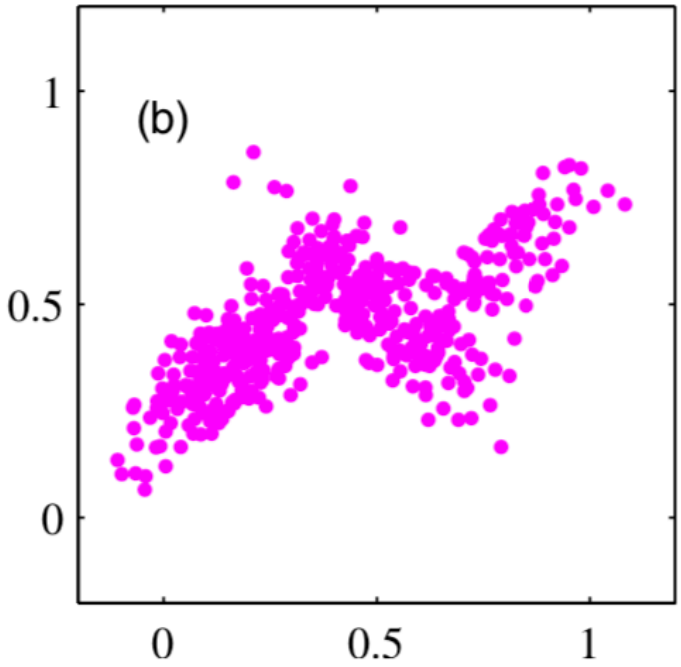
We thereby quantitatively relate a dependent variable to several other variables.

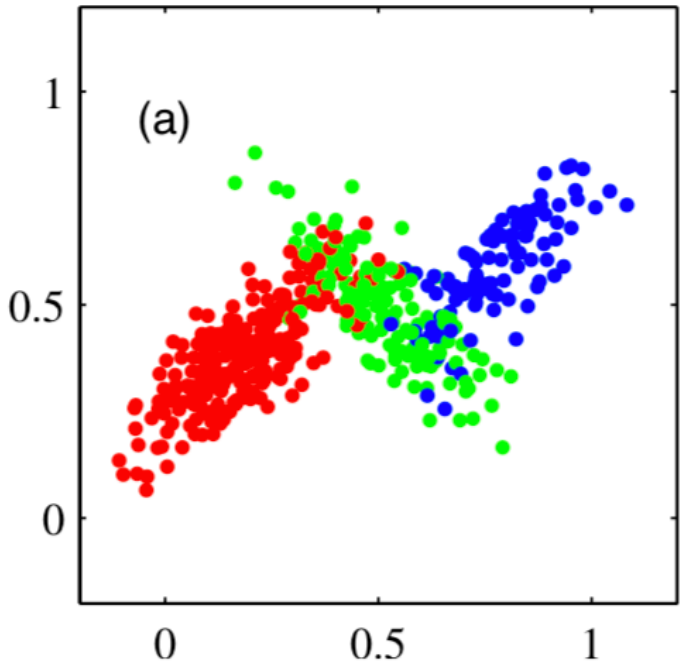
features

class

Clustering

For a given dataset, we are interested in finding relationships, e.g., structures like clusters, and estimating distributions





Clustering Example

Clustering Example: LjudMAP

Machine Learning Approaches

Supervised learning

- Fit a model to observations together with **known** classes
- Ex: Classification and regression

Unsupervised learning

- Discover structures in a dataset having **unknown** classes
- Ex: Clustering

Weakly supervised learning

- Fit a model to observations together with some known classes (which might be wrong)

Machine Learning Approaches

Deep learning

- Fit a specific type of model (multilayer perceptron, deep neural network) to observations together with known classes
- This can also be applied to unsupervised learning, e.g., *autoencoding*

Transfer learning

- Adapt a trained model to a new problem domain

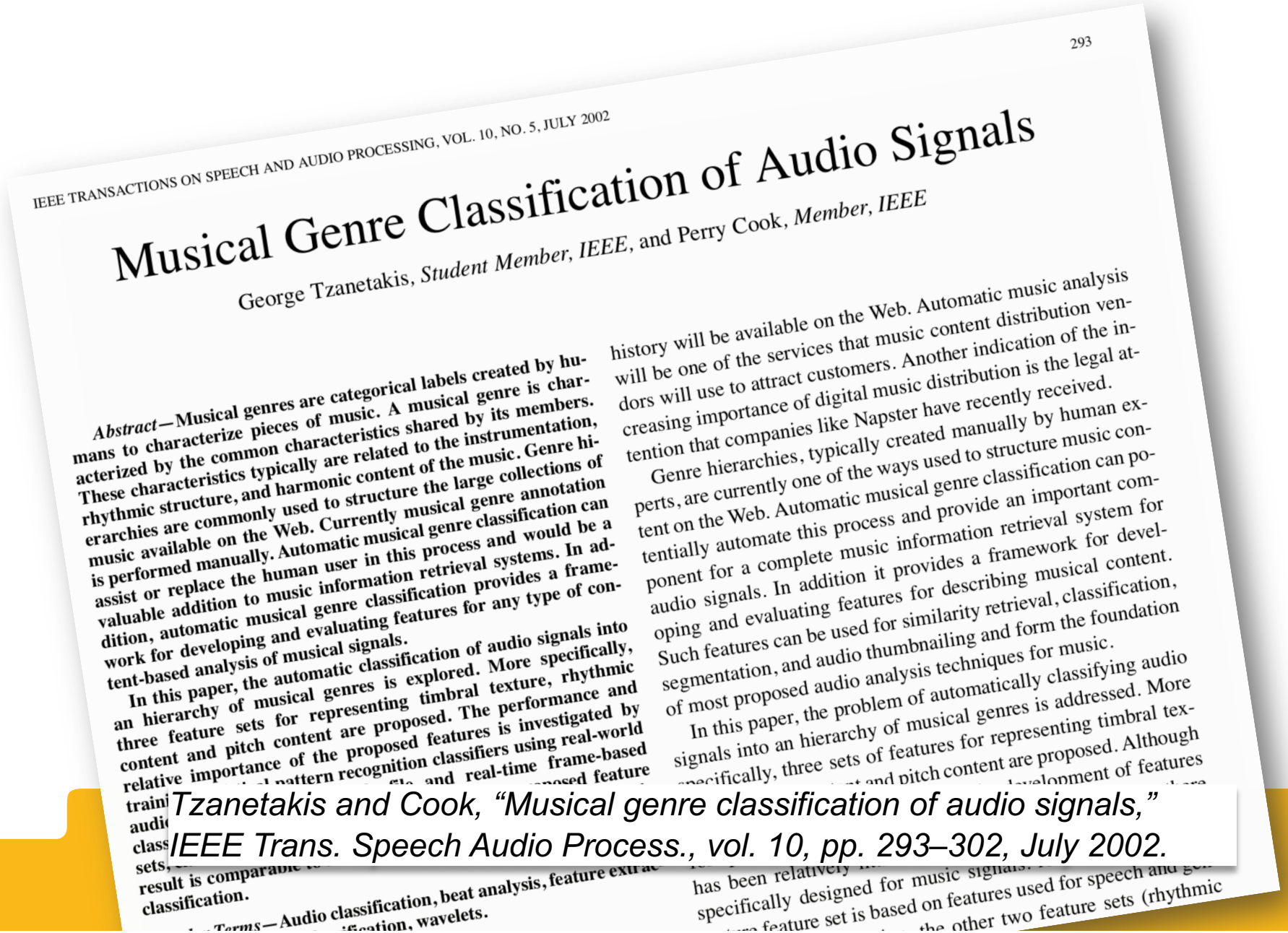
Reinforcement learning

- Learn by acting in some environment and receiving feedback

Machine learning pipeline

1. Problem is defined
2. Many observations are or have been collected
3. Dataset is constructed
4. Dataset is partitioned into multiple subsets, e.g., training/ testing, or training/validation/testing
5. Features are extracted from training dataset
6. Models are trained on training dataset
7. Models are selected using validation dataset
8. Model is evaluated on testing dataset

Let's look at a real life example from music informatics!



IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, VOL. 10, NO. 5, JULY 2002 293

Musical Genre Classification of Audio Signals

George Tzanetakis, *Student Member, IEEE*, and Perry Cook, *Member, IEEE*

***Abstract*—Musical genres are categorical labels created by humans to characterize pieces of music. A musical genre is characterized by the common characteristics shared by its members. These characteristics typically are related to the instrumentation, rhythmic structure, and harmonic content of the music. Genre hierarchies are commonly used to structure the large collections of music available on the Web. Currently musical genre annotation is performed manually. Automatic musical genre classification can assist or replace the human user in this process and would be a valuable addition to music information retrieval systems. In addition, automatic musical genre classification provides a framework for developing and evaluating features for any type of content-based analysis of musical signals.**

In this paper, the automatic classification of audio signals into an hierarchy of musical genres is explored. More specifically, three feature sets for representing timbral texture, rhythmic content and pitch content are proposed. The performance and relative importance of the proposed features is investigated by training statistical pattern recognition classifiers using real-world audio collections. Both [illegible] and real-time frame-based classification schemes are described. A proposed feature set is comparable to

***Terms*—Audio classification, beat analysis, feature extraction, wavelets.**

history will be available on the Web. Automatic music analysis will be one of the services that music content distribution vendors will use to attract customers. Another indication of the increasing importance of digital music distribution is the legal attention that companies like Napster have recently received.

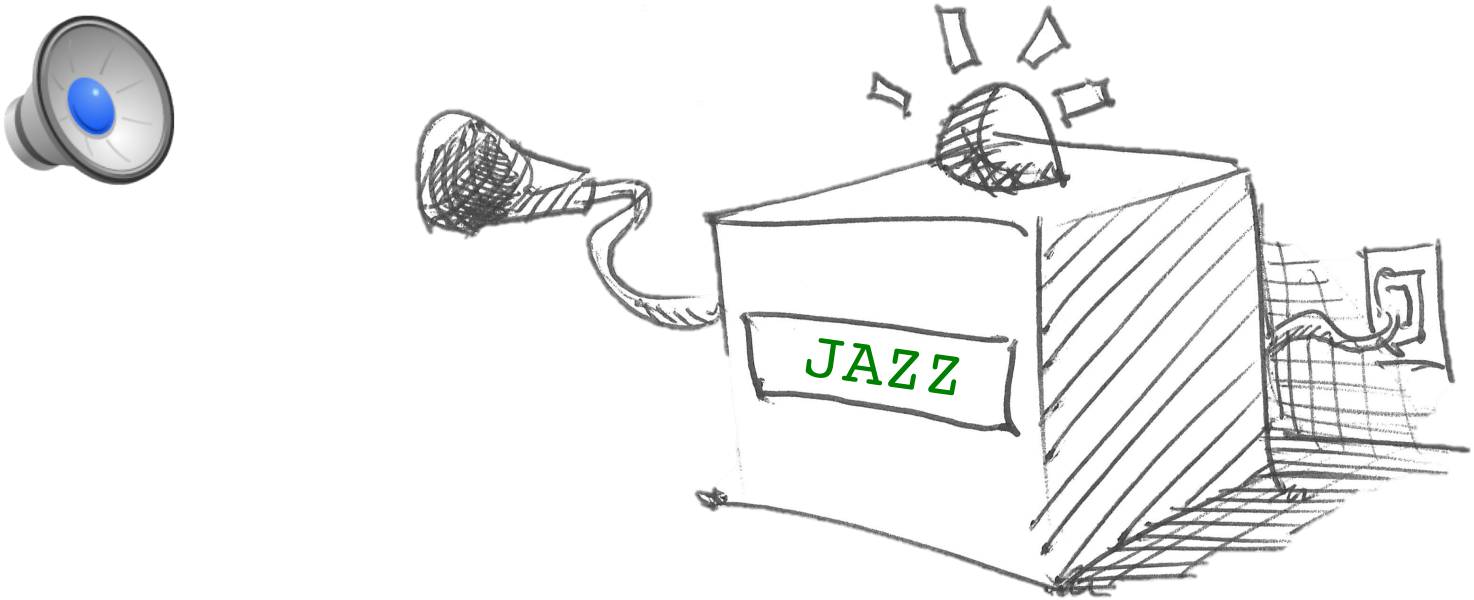
Genre hierarchies, typically created manually by human experts, are currently one of the ways used to structure music content on the Web. Automatic musical genre classification can potentially automate this process and provide an important component for a complete music information retrieval system for audio signals. In addition it provides a framework for developing and evaluating features for describing musical content. Such features can be used for similarity retrieval, classification, segmentation, and audio thumbnailing and form the foundation of most proposed audio analysis techniques for music.

In this paper, the problem of automatically classifying audio signals into an hierarchy of musical genres is addressed. More specifically, three sets of features for representing timbral texture and pitch content are proposed. Although [illegible] development of features has been relatively [illegible] specifically designed for music signals. [illegible] feature set is based on features used for speech and general [illegible] the other two feature sets (rhythmic

Tzanetakis and Cook, "Musical genre classification of audio signals," IEEE Trans. Speech Audio Process., vol. 10, pp. 293–302, July 2002.

Consider this music listening machine

Consider this music listening machine



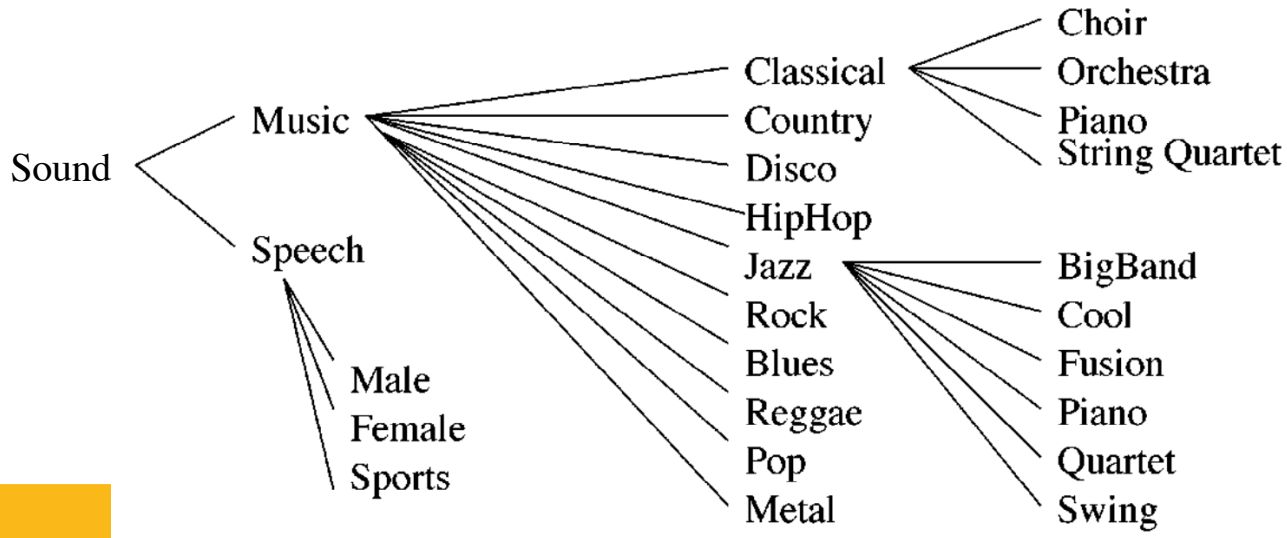
Consider this music listening machine

Consider this music listening machine

Machine learning pipeline

1. Problem is defined
2. Many observations are or have been collected

In 2000, George thinks about how music and audio might be related. He has a collection of CDs, friends with CDs, and a radio and tape recorder. He defines a taxonomy:



Machine learning pipeline

1. Problem is defined
2. Many observations are or have been collected
3. Dataset is constructed

George rips his collection of CDs, records audio from a radio, and creates 1000 30-second sound files at 22050 Hz sampling rate, 16 bit quantization. He then labels each sound file with one of ten labels:

1. Blues
2. Classical
3. Country
4. Disco
5. Hiphop
6. Jazz
7. Metal
8. Pop
9. Reggae
10. Rock

Each category has one hundred sound files.

Machine learning pipeline

1. Problem is defined
2. Many observations are or have been collected
3. Dataset is constructed
4. Dataset is partitioned into multiple subsets, e.g., training/ testing, or training/validation/testing

George randomly divides the sound files from each class into ten equal-sized subsets, and selects 9 subsets at random from each class (90 sound files) to create the training dataset (900 sound files). The remainder (10 sound files from each class, 100 sound files in total) forms the testing dataset.

Machine learning pipeline

1. Problem is defined
2. Many observations are or have been collected
3. Dataset is constructed
4. Dataset is partitioned into multiple subsets, e.g., training/ testing, or training/validation/testing
5. Features are extracted from training dataset

*For each sound file in the training dataset, George extracts a bunch of different features related to harmonic and temporal content. This reduces each 22050*30 = 661,500 sample sound file to feature vectors of up to 30 dimensions.*

Machine learning pipeline

1. Problem is defined
2. Many observations are or have been collected
3. Dataset is constructed
4. Dataset is partitioned into multiple subsets, e.g., training/ testing, or training/validation/testing
5. Features are extracted from training dataset
6. Models are trained on training dataset

George selects some different functions to model the relationships between the ten classes and the features, and then estimates the parameters of each model using the training dataset.

Machine learning pipeline

1. Problem is defined
2. Many observations are or have been collected
3. Dataset is constructed
4. Dataset is partitioned into multiple subsets, e.g., training/ testing, or training/validation/testing
5. Features are extracted from training dataset
6. Models are trained on training dataset
7. Model is evaluated on testing dataset

Using each model, George labels the features extracted from the sound files in the testing dataset, and counts the number of times the label matches the ground truth label (accuracy).

Results:

Different models and features

Model	Genres(10)	Classical(4)	Jazz(6)
Random			
RT GS	44 ± 2	61 ± 3	53 ± 4
GS	59 ± 4	77 ± 6	61 ± 8
GMM(2)	60 ± 4	81 ± 5	66 ± 7
GMM(3)	61 ± 4	88 ± 4	68 ± 7
GMM(4)	61 ± 4	88 ± 5	62 ± 6
GMM(5)	61 ± 4	88 ± 5	59 ± 6
KNN(1)	59 ± 4	77 ± 7	57 ± 6
KNN(3)	60 ± 4	78 ± 6	58 ± 7
KNN(5)	56 ± 3	70 ± 6	56 ± 6

Q: What?

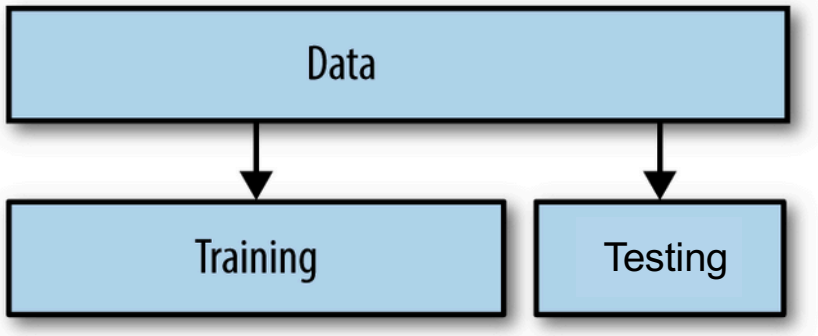
Let's dig in: *Dataset partitioning*

1. Problem is defined
2. Many observations are or have been collected
3. Dataset is constructed
4. Dataset is partitioned into multiple subsets, e.g., training/ testing, or training/validation/testing

George randomly divides the sound files from each class into ten equal-sized subsets, and selects 9 subsets at random from each class (90 sound files) to create the training dataset (900 sound files). The remainder (10 sound files from each class, 100 sound files in total) forms the testing dataset.

Partition strategy: *Holdout*

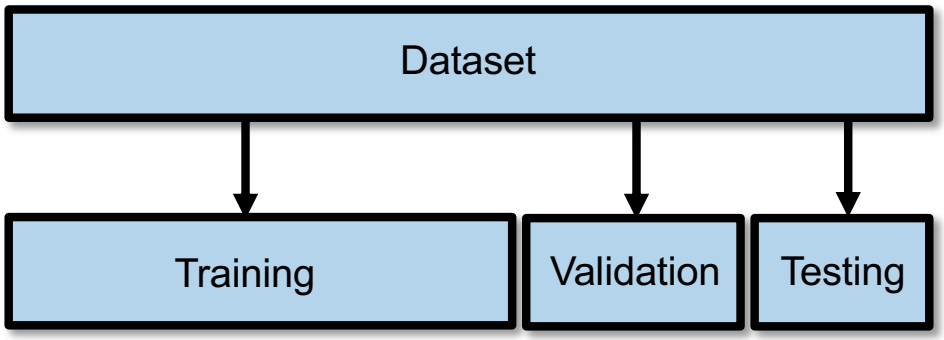
- Split dataset into two disjoint sets.
- Typical is 70/30, or 80/20.



Q: Why?

Partition strategy: *Holdout*

- Split dataset into three disjoint sets.
- Typical is 50/25/25, or 70/20/10

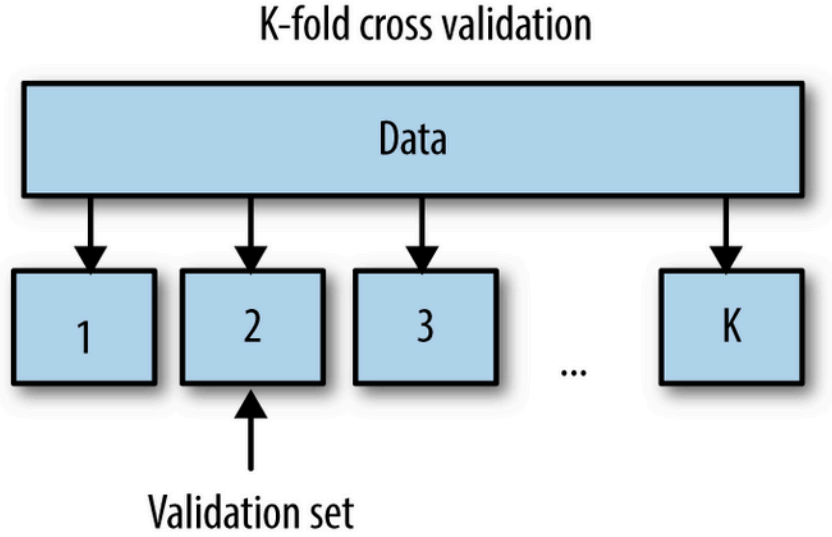


Q: Why?

Partition strategy: *K-fold Cross Validation*

- Split dataset into K disjoint pieces, then rotate which one to hold out for testing.
- Perform training and testing K times.
- Typical K is 5 and 10.

Q: Why?



- George uses 10-fold *stratified* cross-validation $\mathcal{I}_{\mathrm{tr}}, \mathcal{I}_{\mathrm{te}}$
- What is $|\mathcal{I}_{\mathrm{tr}}|$ and $|\mathcal{I}_{\mathrm{te}}|$ in each repetition?

Indexes into dataset

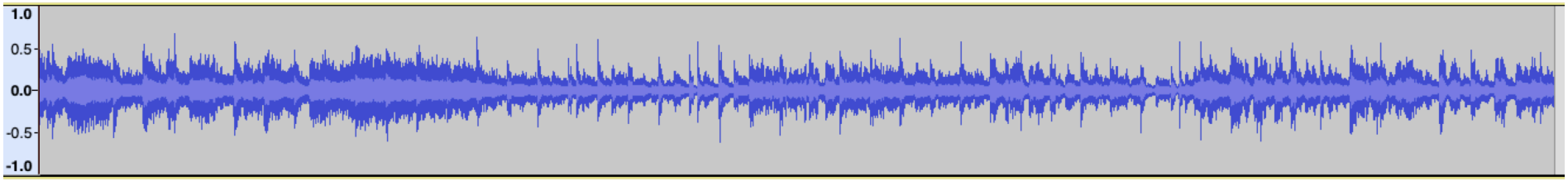
Let's dig in: *Feature extraction*

1. Problem is defined
2. Many observations are or have been collected
3. Dataset is constructed
4. Dataset is partitioned into multiple subsets, e.g., training/ testing, or training/validation/testing
5. Features are extracted from training dataset

*For each sound file in the training dataset, George extracts a bunch of different features related to harmonic and temporal content. This reduces each 22050*30 = 661,500 sample sound file to feature vectors of up to 30 dimensions.*

Feature extraction

$x[n]$



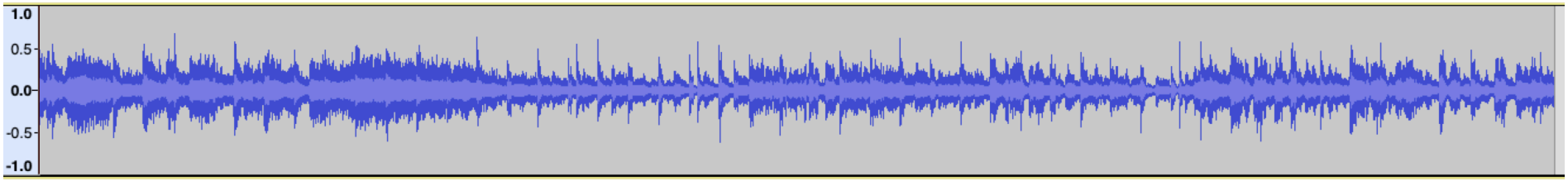
We need to *describe* each 30-second sound file in a more *meaningful* and *compact* way than just 600,000 samples.

George proposes using features related to:

1. Timbre
2. Rhythm
3. Pitch

Feature extraction: *Timbre features*

$x[n]$

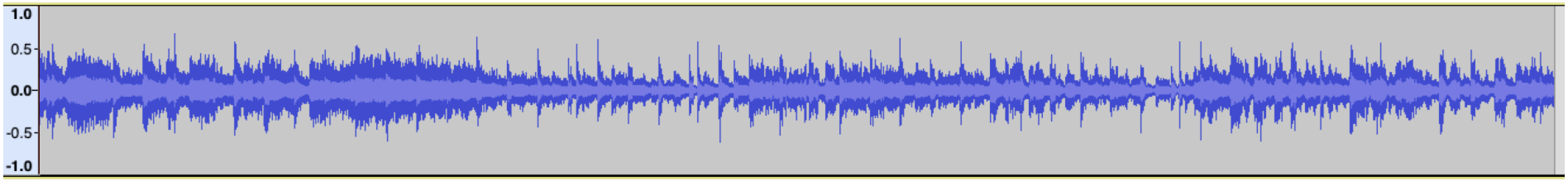


- Read sound file into memory -> $x[n]$
- Using rectangular windows of size 23 ms (how many samples when Fs=22050?), and hop 23 ms, to create *l*th analysis frame of sound file (what is the domain of *l* if sound file is 30 seconds long?) : $x_l[n]$
- Compute: *zero crossings*

$$z[l] := \frac{1}{2}\sum_{n=1}^{N-1}(\mathrm{sign}x_l[n] - \mathrm{sign}x_l[n-1])$$

Feature extraction: *Timbre features*

$x[n]$

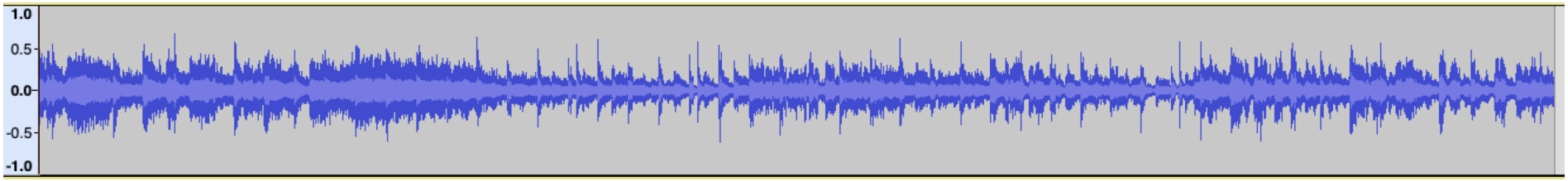


- Read sound file into memory -> $x[n]$
- Compute magnitude STFT with Hann window of size of 23 ms,hopped by 23 ms -> $|X[l,k]|$
  where l is the analysis frame *index*
- Compute: *Spectral centroid*

$$C_s[l] := \sum_{k=0}^{K/2+1} k|X[l,k]| / \sum_{k=0}^{K/2+1} |X[l,k]|$$

Feature extraction: *Timbre features*

$x[n]$

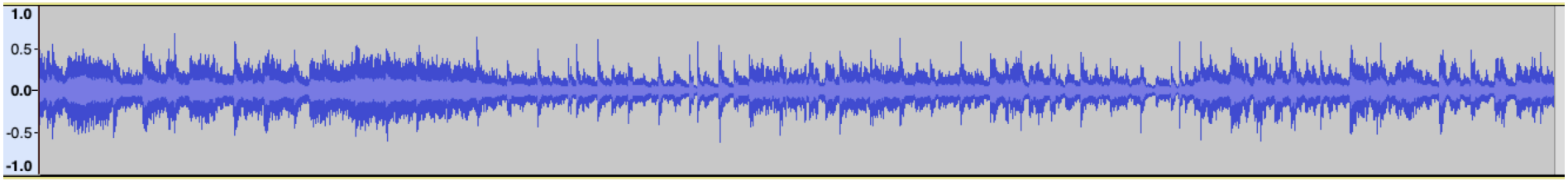


- Read sound file into memory -> $x[n]$
- Compute mag. STFT with Hann window of size of 23 ms, hopped by 23 ms -> $|X[l,k]|$
  where l is the analysis frame index
- Compute: *Spectral rolloff* $R_s[l]$

$$\sum_{k=0}^{R_s[l]} 0.85|X[l,k]| = \sum_{k=0}^{K/2+1} |X[l,k]|$$

Feature extraction: *Timbre features*

$x[n]$



- Read sound file into memory -> $x[n]$
- Compute mag. STFT with Hann window of size of 23 ms, hopped by 23 ms -> $|X[l,k]|$
  where l is the analysis frame index
- Compute: *Spectral flux*

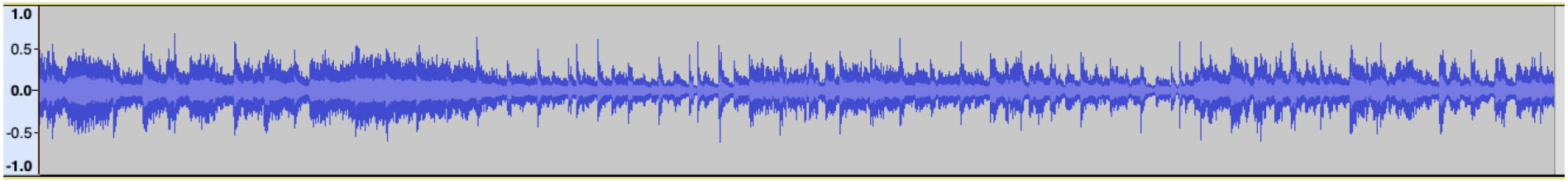
$$F_s[l] := \sum_{k=0}^{K/2+1} \left(|\hat{X}[l,k]| - |\hat{X}[l-1,k]|\right)^2$$

where

$$|\hat{X}[l,k]| := |X[l,k]| / \sum_{k=0}^{K/2+1} |X[l,k]|$$

Feature extraction

$x[n]$



Extracted from each 30-second sound file are several sequences:

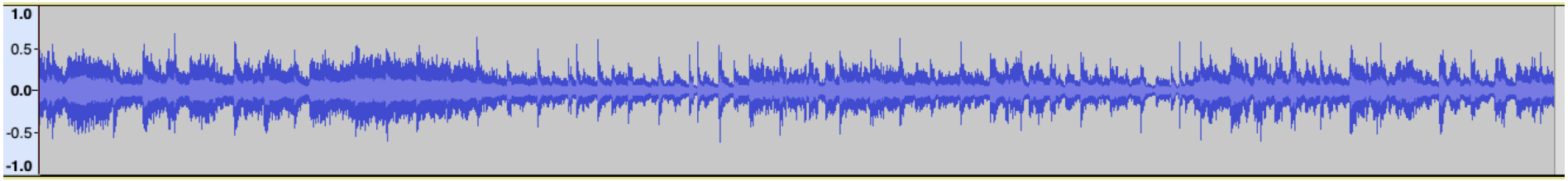
$$\{\bar{z}[l], C_s[l], R_s[l], F_s[l] : l \in [0, \ldots]\}$$

Each *l* corresponds to 512 samples (23 ms)

George computes descriptive statistics of these using *texture windows* corresponding to 1 second of analysis frames (how many?)

Feature extraction

$x[n]$



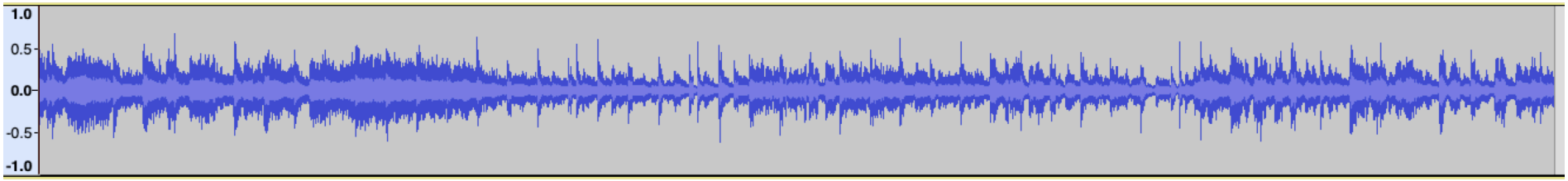
Extracted from each 30-second sound file are several sequences: $\{\bar{z}[l], C_s[l], R_s[l], F_s[l] : l \in [0, \ldots]\}$ *Analysis frame features L*

Each *l* corresponds to 512 samples of $x[n]$ (23 ms)

George computes descriptive statistics of the features extracted from 1 second's worth of analysis frames (how many analysis frames are in 1 second?)

Feature extraction

$x[n]$



Extracted from each 30-second sound file are several sequences: $\{\bar{z}[l], C_s[l], R_s[l], F_s[l] : l \in [0, \ldots]\}$ *Analysis frame features L*

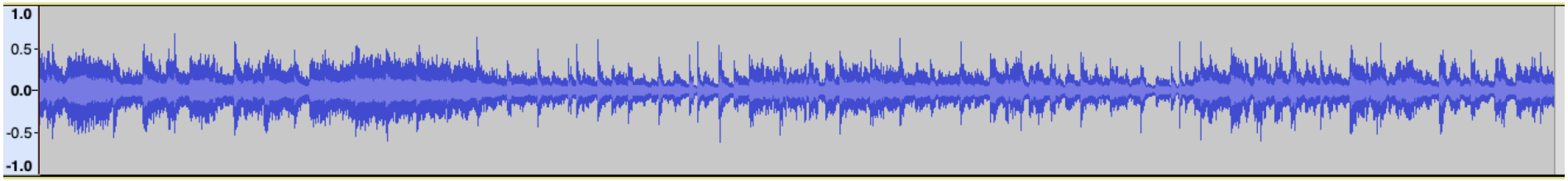
mean *variance*

$$\{\bar{z}[m], \mathrm{var}z[m] : m \in [0, \ldots]\}$$
$$\{\bar{C}_s[m], \mathrm{var}C_s[m] : m \in [0, \ldots]\}$$
$$\{\bar{R}_s[m], \mathrm{var}R_s[m] : m \in [0, \ldots]\}$$
$$\{\bar{F}_s[m], \mathrm{var}F_s[m] : m \in [0, \ldots]\}$$

Texture features M

Feature *vectors*

$x[n]$



Features are put one atop another to create vectors, e.g.,

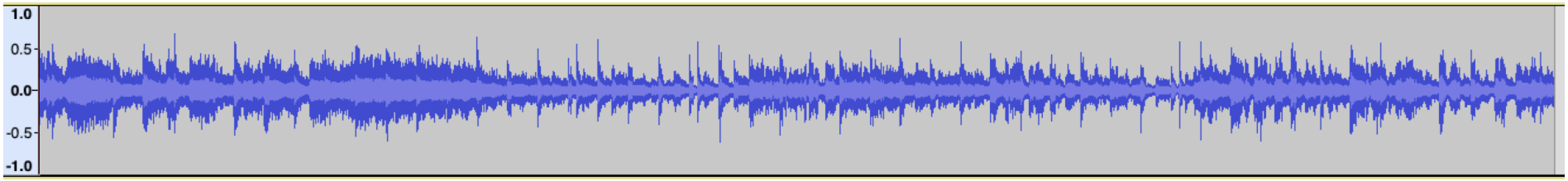
$$\mathbf{v}[m] = \begin{bmatrix} \bar{z}[m] \\ \mathrm{var} z[m] \\ \bar{C}_s[m] \\ \mathrm{var} C_s[m] \\ \bar{R}_s[m] \\ \mathrm{var} R_s[m] \\ \bar{F}_s[m] \\ \mathrm{var} F_s[m] \end{bmatrix}$$ D=8

Feature extraction

$$x[n] \Longrightarrow (\mathbf{v}[m] \in \mathbb{R}^8 : m = 0, \ldots)$$

Feature space

Feature extraction



We need to describe each 30-second sound file in a more meaningful and compact way than just 600,000 samples.

George proposes using features related to:

1. Timbre
2. Rhythm
3. Pitch

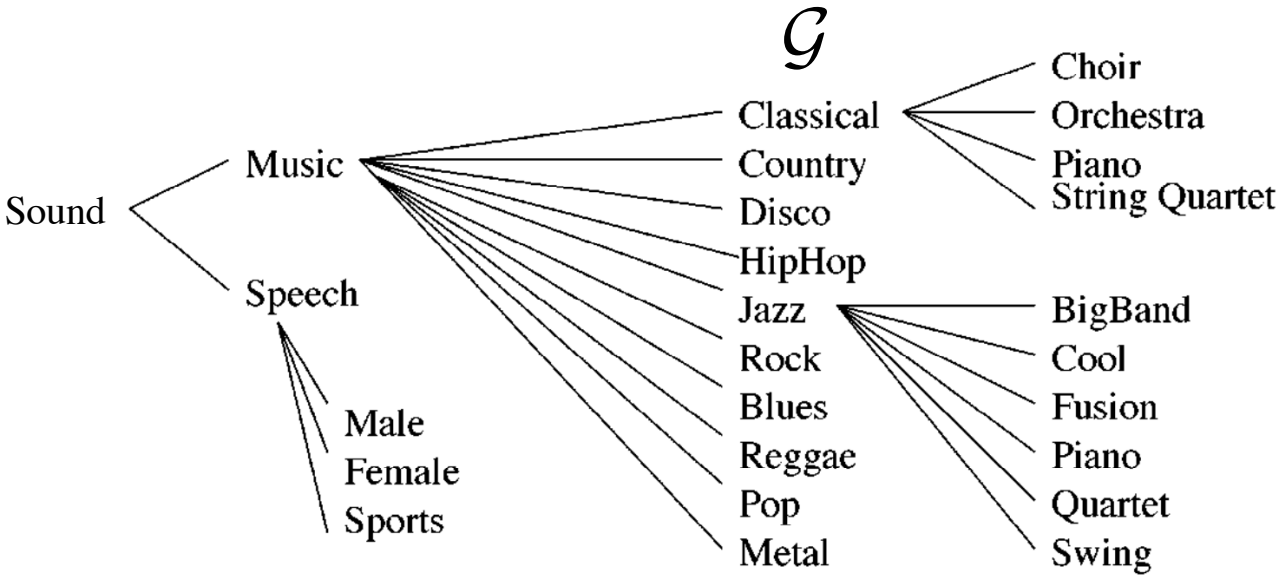
} These other features we will deal with later

Machine learning pipeline

1. Problem is defined
2. Many observations are or have been collected
3. Dataset is constructed
4. Dataset is partitioned into multiple subsets, e.g., training/ testing, or training/validation/testing
5. Features are extracted from training dataset
6. Models are trained on training dataset

George selects some different functions to model the relationships between the ten classes and the features, and then estimates the parameters of each model using the training dataset.

Let's dig in: *Modeling* feature vectors

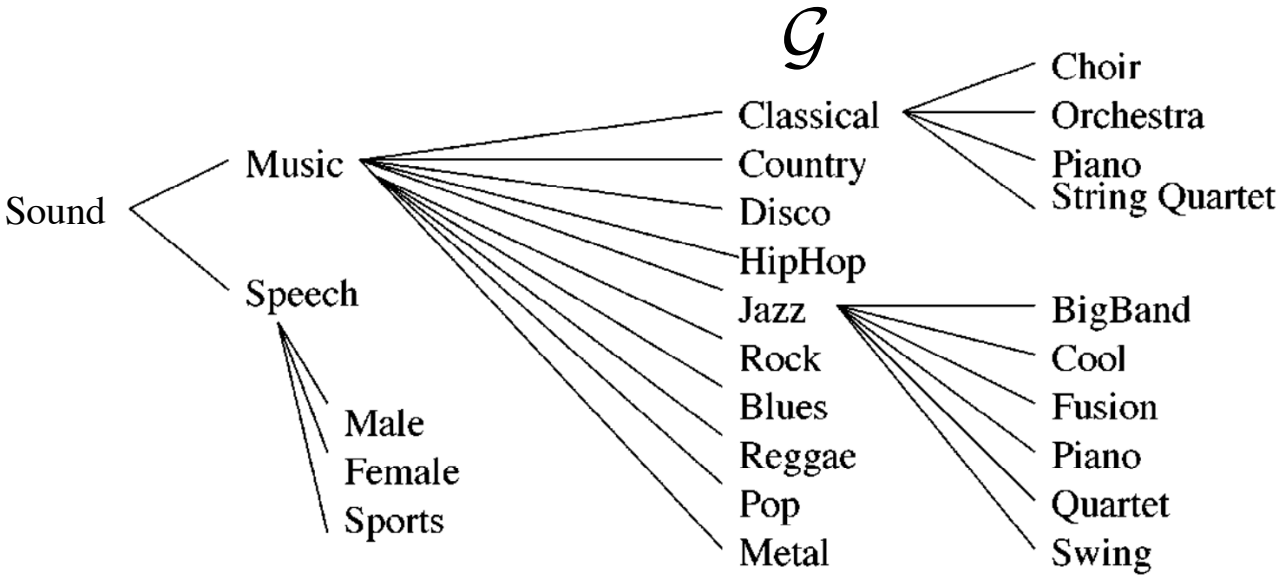


Assume features extracted from sound files in a specific class are similar.

One approach to quantify similarity is by using the notion of conditional probability distributions:

$$P[\mathbf{v}|g = \text{Disco}]$$

Modeling feature vectors



George selects a *Gaussian distribution* to model the probability distributions of D-dimensional feature vectors of class *g*:

$$P[\mathbf{v}|g \in \mathcal{G}] \sim \mathcal{N}(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$$

$$P[\mathbf{v}|g \in \mathcal{G}] = \frac{1}{\sqrt{(2\pi)^D |\boldsymbol{\Sigma}_g|}} \exp[-(\mathbf{v} - \boldsymbol{\mu}_g)^T \boldsymbol{\Sigma}_g^{-1} (\mathbf{v} - \boldsymbol{\mu}_g)]$$

parameters

Modeling feature vectors

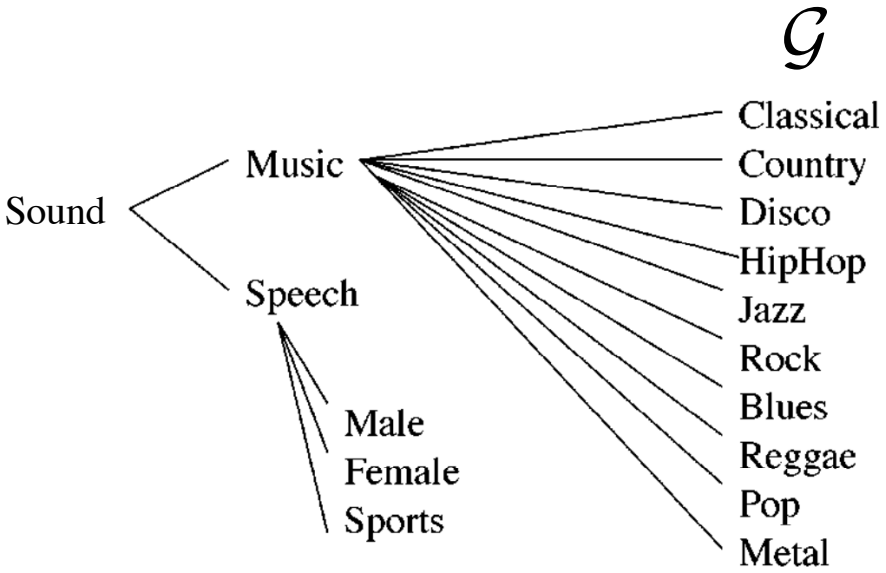
$$P[\mathbf{v}|g \in \mathcal{G}] \sim \mathcal{N}(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$$

George *estimates* the parameters of the probability distributions for each class by using the features extracted from all the sound files in the training dataset, $\mathcal{I}_{\text{tr}}$, in that class: $\mathcal{I}_{\text{tr},g}$

$$\hat{\boldsymbol{\mu}}_g = \frac{1}{M|\mathcal{I}_{\text{tr},g}|} \sum_{i \in \mathcal{I}_{\text{tr},g}} \sum_{m=0}^{M-1} \mathbf{v}_i[m]$$

$$\hat{\boldsymbol{\Sigma}}_g = \sum_{i \in \mathcal{I}_{\text{tr},g}} \frac{1}{|\mathcal{I}_{\text{tr},g}|M-1} \sum_{m=0}^{M-1} (\mathbf{v}_i[m] - \hat{\boldsymbol{\mu}}_g)(\mathbf{v}_i[m] - \hat{\boldsymbol{\mu}}_g)^T$$

Modeling feature vectors



$$\mathcal{N}(\hat{\boldsymbol{\mu}}_{\text{Classical}}, \hat{\boldsymbol{\Sigma}}_{\text{Classical}})$$
$$\mathcal{N}(\hat{\boldsymbol{\mu}}_{\text{Country}}, \hat{\boldsymbol{\Sigma}}_{\text{Country}})$$
$$\mathcal{N}(\hat{\boldsymbol{\mu}}_{\text{Disco}}, \hat{\boldsymbol{\Sigma}}_{\text{Disco}})$$
$$\ldots$$

George has 10 probability distributions, one for each class, describing how *multivariate* feature vectors are distributed in the feature space.

Machine learning pipeline

1. Problem is defined
2. Many observations are or have been collected
3. Dataset is constructed
4. Dataset is partitioned into multiple subsets, e.g., training/ testing, or training/validation/testing
5. Features are extracted from training dataset
6. Models are trained on training dataset
7. Model is evaluated on testing dataset

Using each model, George labels the features extracted from the sound files in the testing dataset, and counts the number of times the label matches the ground truth label (accuracy).

Let's dig in: *Evaluation*

For a feature vector $\mathbf{v}$ with ground truth $g_{\text{true}}(\mathbf{v})$ extracted from a sound file in the testing dataset, George finds the model with which this vector has the highest *likelihood*:

$$\hat{g}(\mathbf{v}) = \arg\max_{g \in \mathcal{G}} P[\mathbf{v}|g' = g]$$

If $\hat{g}(\mathbf{v}) \equiv g_{\text{true}}(\mathbf{v})$ then this is a win! If not, then a loss.

George does this for all feature vectors in the testing dataset, and counts the number of times the ground truth is reproduced.

The percentage of times this happens is the accuracy.

Results:

Different models and features

Model	Genres(10)	Classical(4)	Jazz(6)
Random	10	25	16
RT GS	44 ± 2	61 ± 3	53 ± 4
GS	59 ± 4	77 ± 6	61 ± 8
GMM(2)	60 ± 4	81 ± 5	66 ± 7
GMM(3)	61 ± 4	88 ± 4	68 ± 7
GMM(4)	61 ± 4	88 ± 5	62 ± 6
GMM(5)	61 ± 4	88 ± 5	59 ± 6
KNN(1)	59 ± 4	77 ± 7	57 ± 6
KNN(3)	60 ± 4	78 ± 6	58 ± 7
KNN(5)	56 ± 3	70 ± 6	56 ± 6

Other possible evaluation measures

Recall: Of the number of feature vectors actually labeled g, what percentage were labeled g?

Precision: Of the number of feature vectors labeled g, what percentage are actually g?

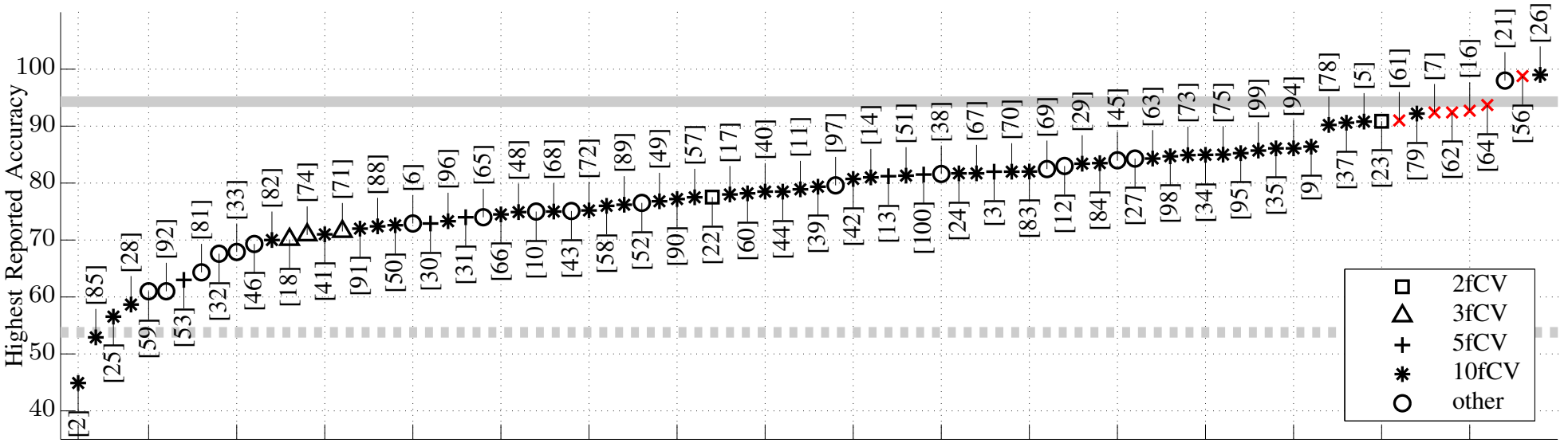
Confusion table:

true class

predicted class

	cl	co	di	hi	ja	ro	bl	re	po	me
cl	**69**	0	0	0	1	0	0	0	0	0
co	0	**53**	2	0	5	8	6	4	2	0
di	0	8	**52**	11	0	13	14	5	9	6
hi	0	3	18	**64**	1	6	3	26	7	6
ja	26	4	0	0	**75**	8	7	1	2	1
ro	5	13	4	1	9	**40**	14	1	7	33
bl	0	7	0	1	3	4	**43**	1	0	0
re	0	9	10	18	2	12	11	**59**	7	1
po	0	2	14	5	3	5	0	3	**66**	0
me	0	1	0	1	0	4	2	0	0	**53**

Many others have now used this dataset



We will return to this in the lecture about *Content-based retrieval*

Machine learning pipeline recap

1. Problem is defined
2. Observation collection
3. Dataset construction
4. Dataset partitioning into training/testing
5. Feature extraction
6. Model definition, parameter estimation from training dataset
7. Model selection from validation dataset
8. Model evaluation on testing dataset

Important points

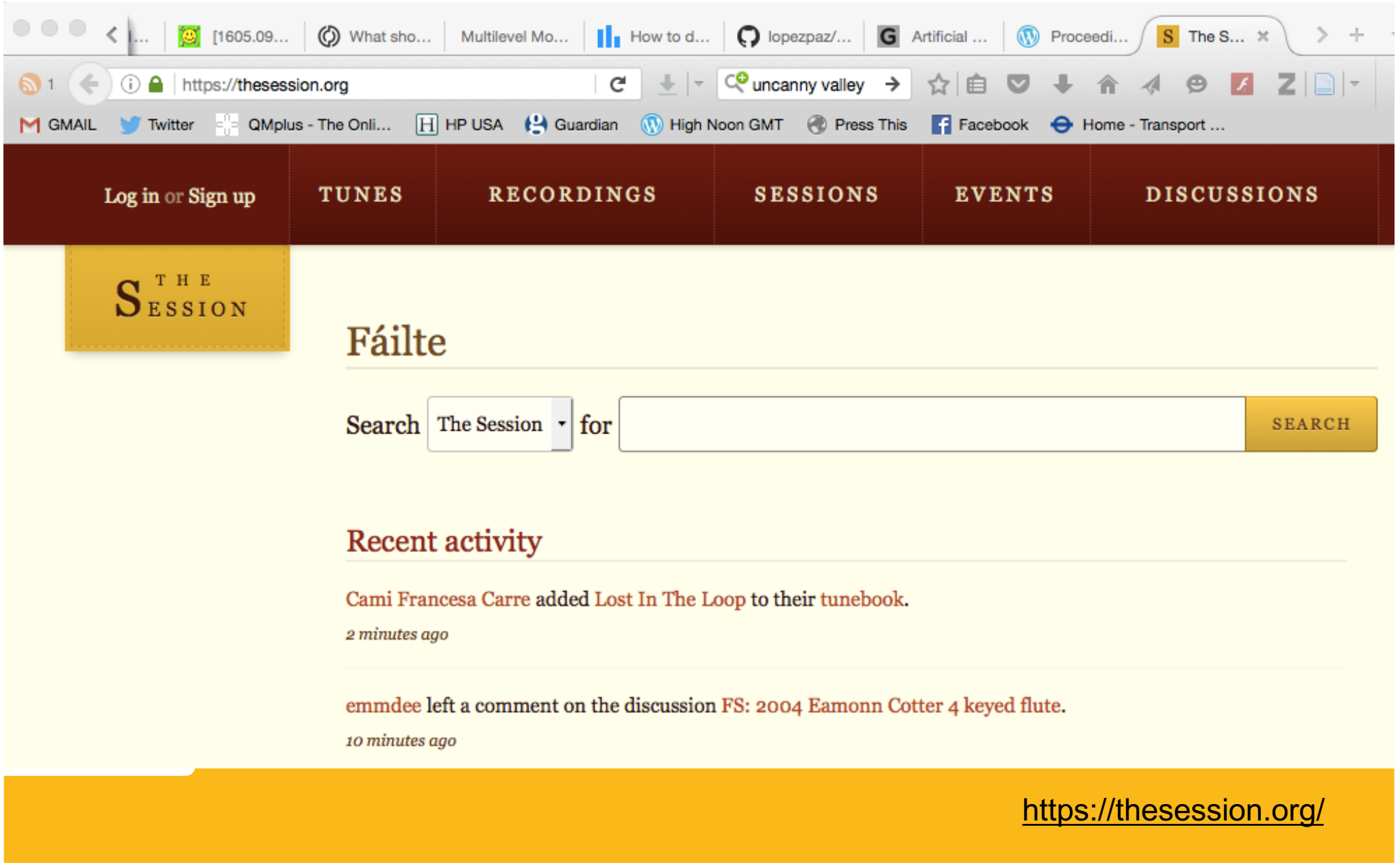
Dataset partitioning: Sanity check your train/test partitioning! Make sure you have not duplicated data (unless that is your intention).

Feature definition and extraction: Take care in choosing features, and processing your data to extract them. Sanity check your extracted features. Do they make sense? Are there any *Inf* or *NaN*?

Models: Sanity check! Do the estimated parameters make sense?

Evaluation: Have you chosen an appropriate measure for the problem definition?

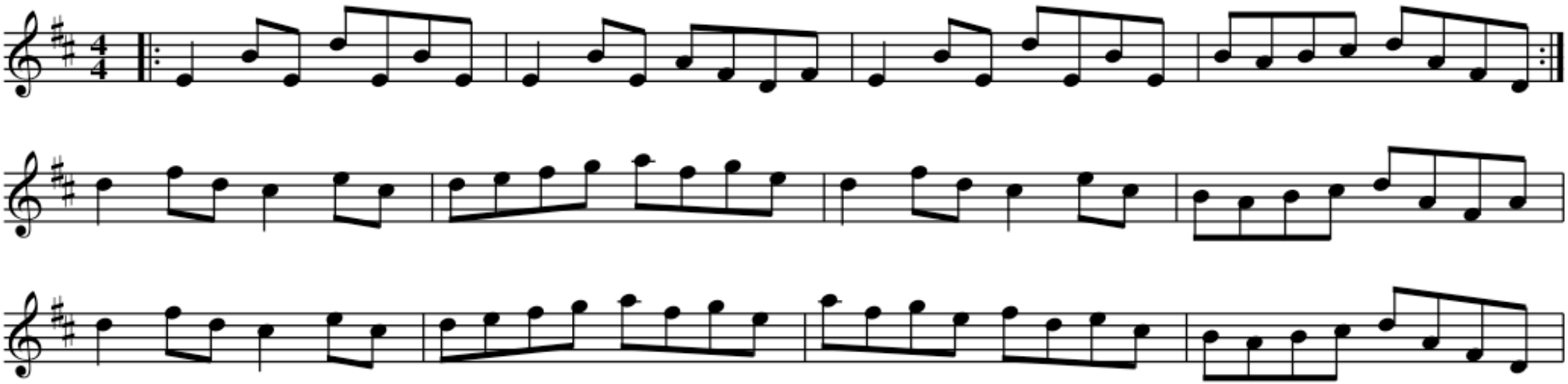
Machine learning relies on data



https://thesession.org/

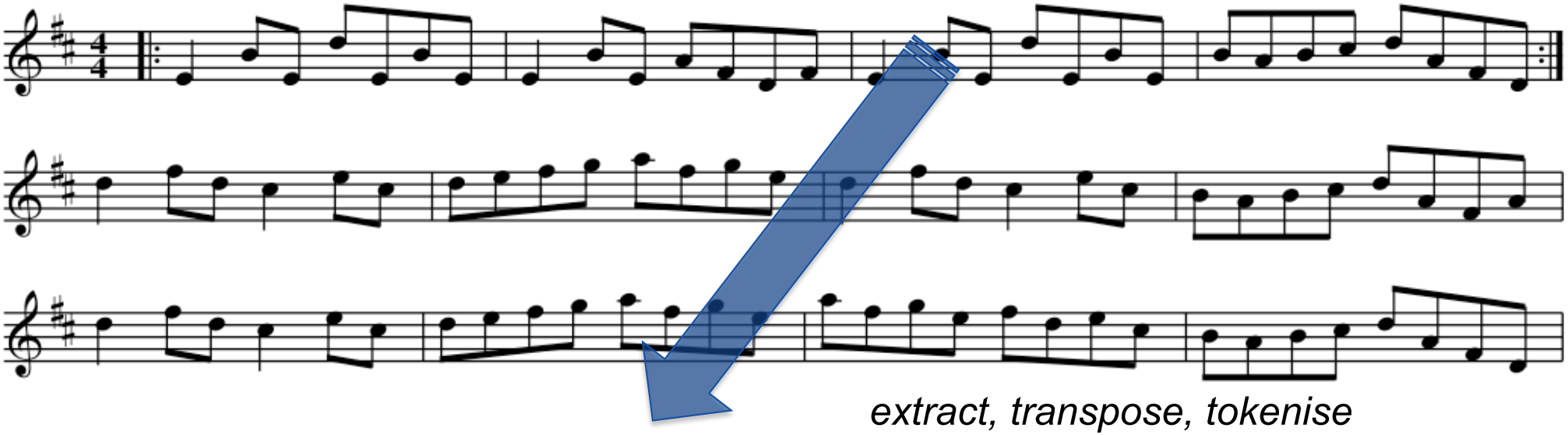
Example transcription from *thesession.org*

27,27,"Drowsy Maggie","reel","4/4","Edorian","|:E2BE dEBE|E2BE AFDF|E2BE dEBE|BABc dAFD:| d2fd c2ec|defg afge|d2fd c2ec|BABc dAFA| d2fd c2ec| defg afge|afge fdec|BABc dAFD|","2001-05-21 03:47:39","Jeremy"



Example transcription from thesession.org

27,27,"Drowsy Maggie","reel","4/4","Edorian","|:E2BE dEBE|E2BE AFDF|E2BE dEBE|BABc dAFD:| d2fd c2ec|defg afge|d2fd c2ec|BABc dAFA| d2fd c2ec| defg afge|afge fdec|BABc dAFD|","2001-05-21 03:47:39","Jeremy"

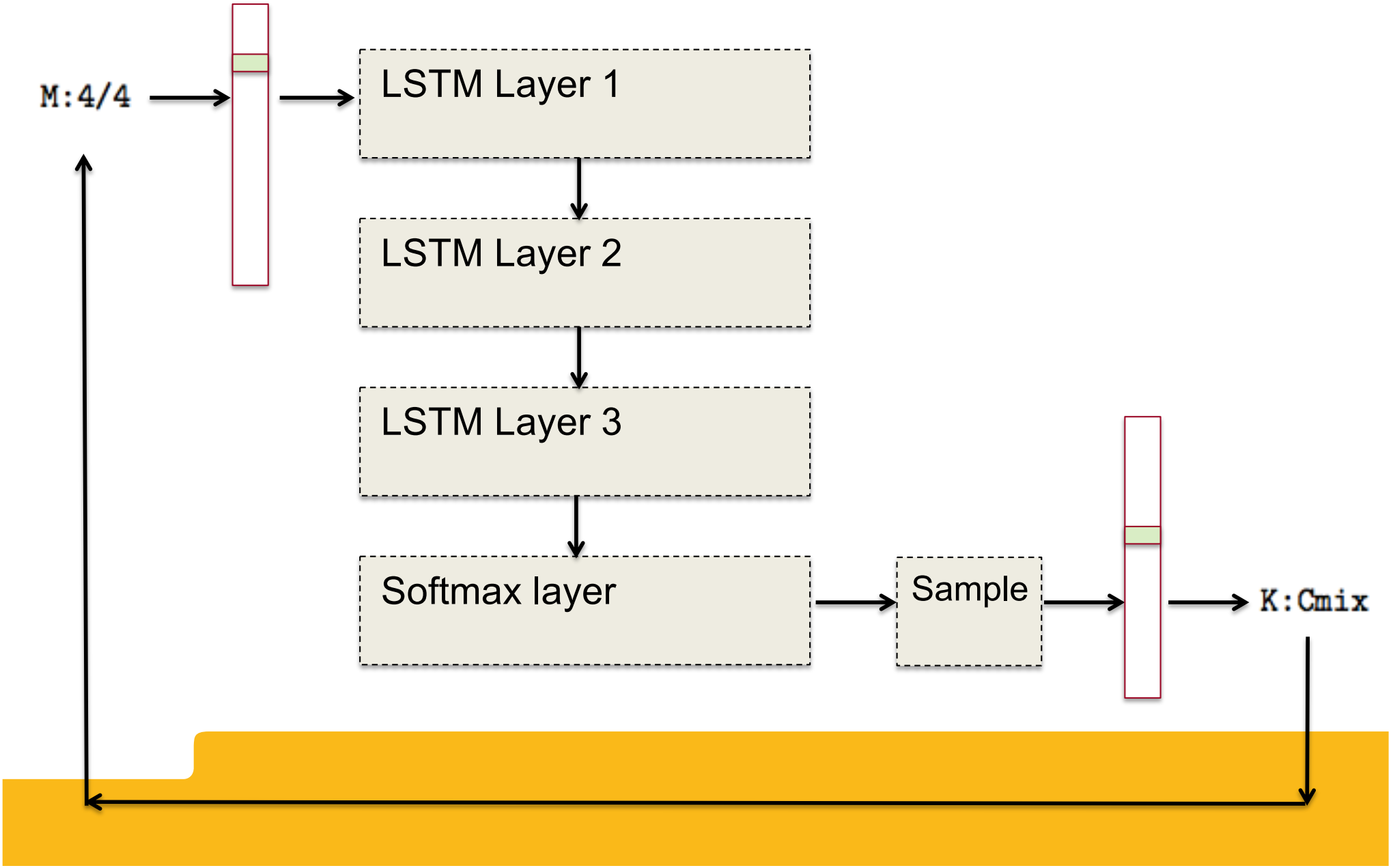


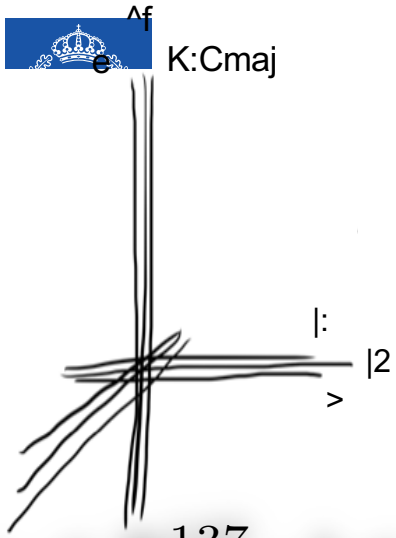
extract, transpose, tokenise

M:4/4 K:Cdor |: C 2 G C B C G C | C 2 G C F D B, D | C 2 G C B C G C | G F G A B F D B, :| B 2 d B A 2 c A | B c d e f d e c | B 2 d B A 2 c A | G F G A B F D F | B 2 d B A 2 c A | B c d e f d e c | f d e c d B c A | G F G A B F D B, |

$$P[x_7|x_6 = \mathtt{G}, x_5 = 2, x_4 = \mathtt{C}, x_3 = |:, \ldots]$$

Machine learning architecture



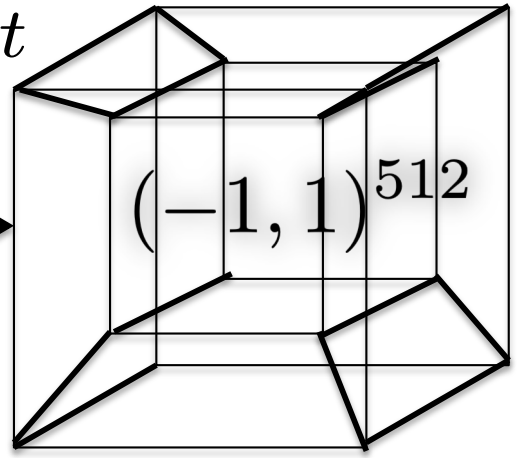


$\{\mathbf{e}_k \in \mathbb{R}^{137} : k \in [1, 137]\}$

$\mathbf{x}_t$

LSTM i=1

$\mathbf{h}_t^{(1)}$



4 gates

$$\mathbf{i}_t^{(1)} \leftarrow \sigma(\mathbf{W}_{xi}^{(1)}\mathbf{x}_t + \mathbf{W}_{hi}^{(1)}\mathbf{h}_{t-1}^{(1)} + \mathbf{b}_i^{(1)})$$

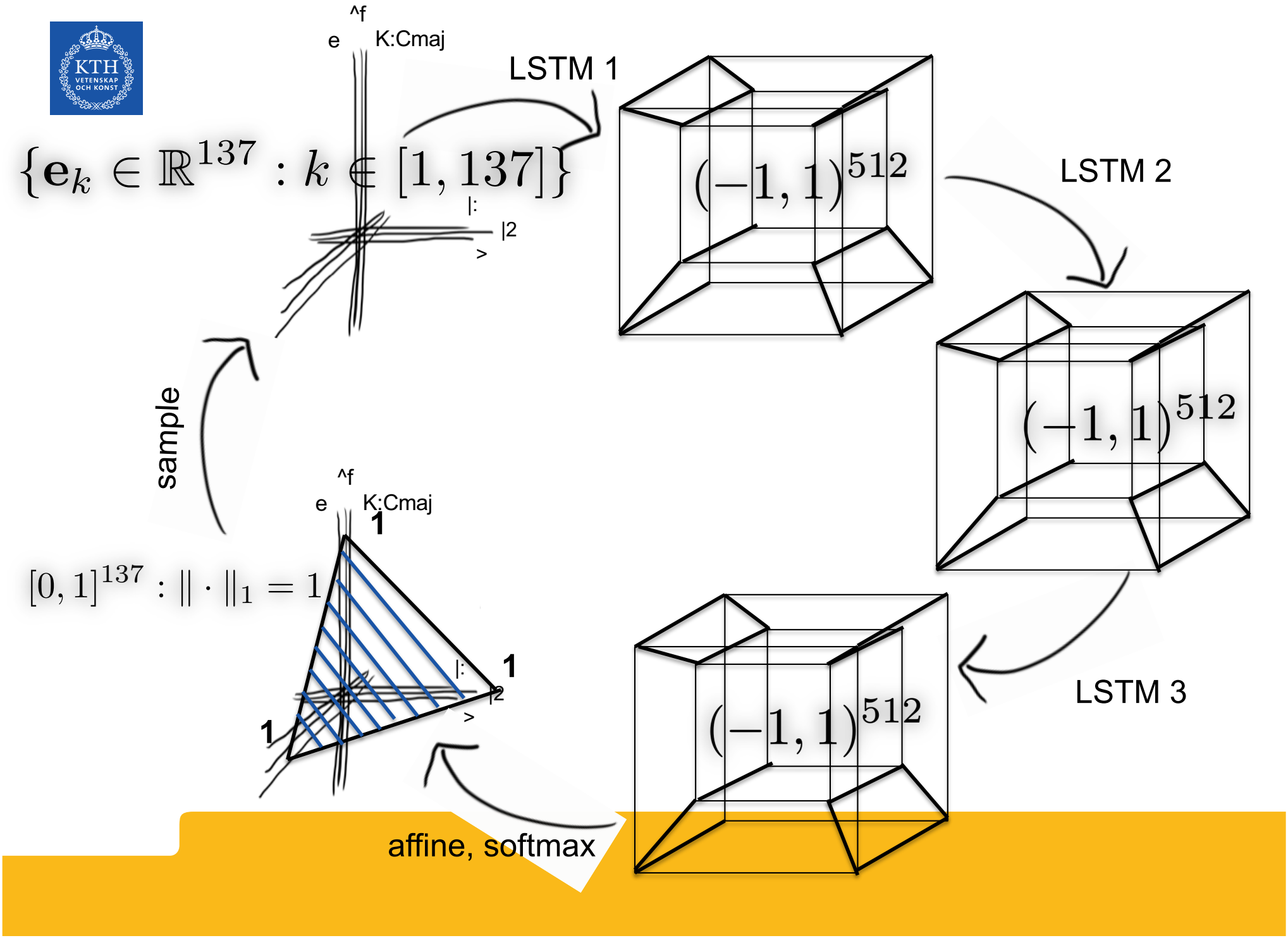
$$\mathbf{f}_t^{(1)} \leftarrow \sigma(\mathbf{W}_{xf}^{(1)}\mathbf{x}_t + \mathbf{W}_{hf}^{(1)}\mathbf{h}_{t-1}^{(1)} + \mathbf{b}_f^{(1)})$$

$$\mathbf{o}_t^{(1)} \leftarrow \sigma(\mathbf{W}_{xo}^{(1)}\mathbf{x}_t + \mathbf{W}_{ho}^{(1)}\mathbf{h}_{t-1}^{(1)} + \mathbf{b}_o^{(1)})$$

$$\mathbf{c}_t^{(1)} \leftarrow \tanh(\mathbf{W}_{xc}^{(1)}\mathbf{x}_t + \mathbf{W}_{hc}^{(1)}\mathbf{h}_{t-1}^{(1)} + \mathbf{b}_c^{(1)}) \odot \mathbf{i}_t^{(1)} + \mathbf{f}_t^{(1)} \odot \mathbf{c}_{t-1}^{(1)}$$

hidden state

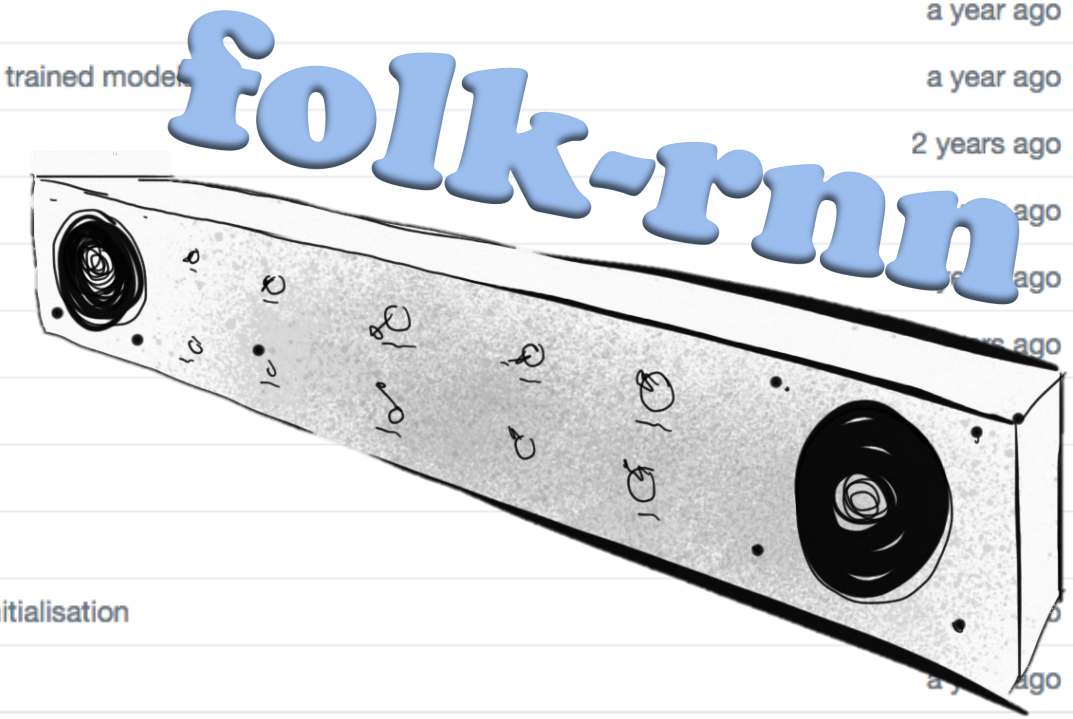
$$\mathbf{h}_t^{(1)} \leftarrow \tanh(\mathbf{c}_t^{(1)}) \odot \mathbf{o}_t^{(1)}$$


^f
e
K:Cmaj
|:
|2
>
LSTM 1
LSTM 2
LSTM 3
sample
affine, softmax
1
1
1
$\{\mathbf{e}_k \in \mathbb{R}^{137} : k \in [1, 137]\}$
$[0, 1]^{137} : \|\cdot\|_1 = 1$
$(-1, 1)^{512}$
$(-1, 1)^{512}$
$(-1, 1)^{512}$

folk-rnn: A folk music AI

boblsturm committed on **GitHub** Update README.md — Latest commit 52a7d37 22 hours ago

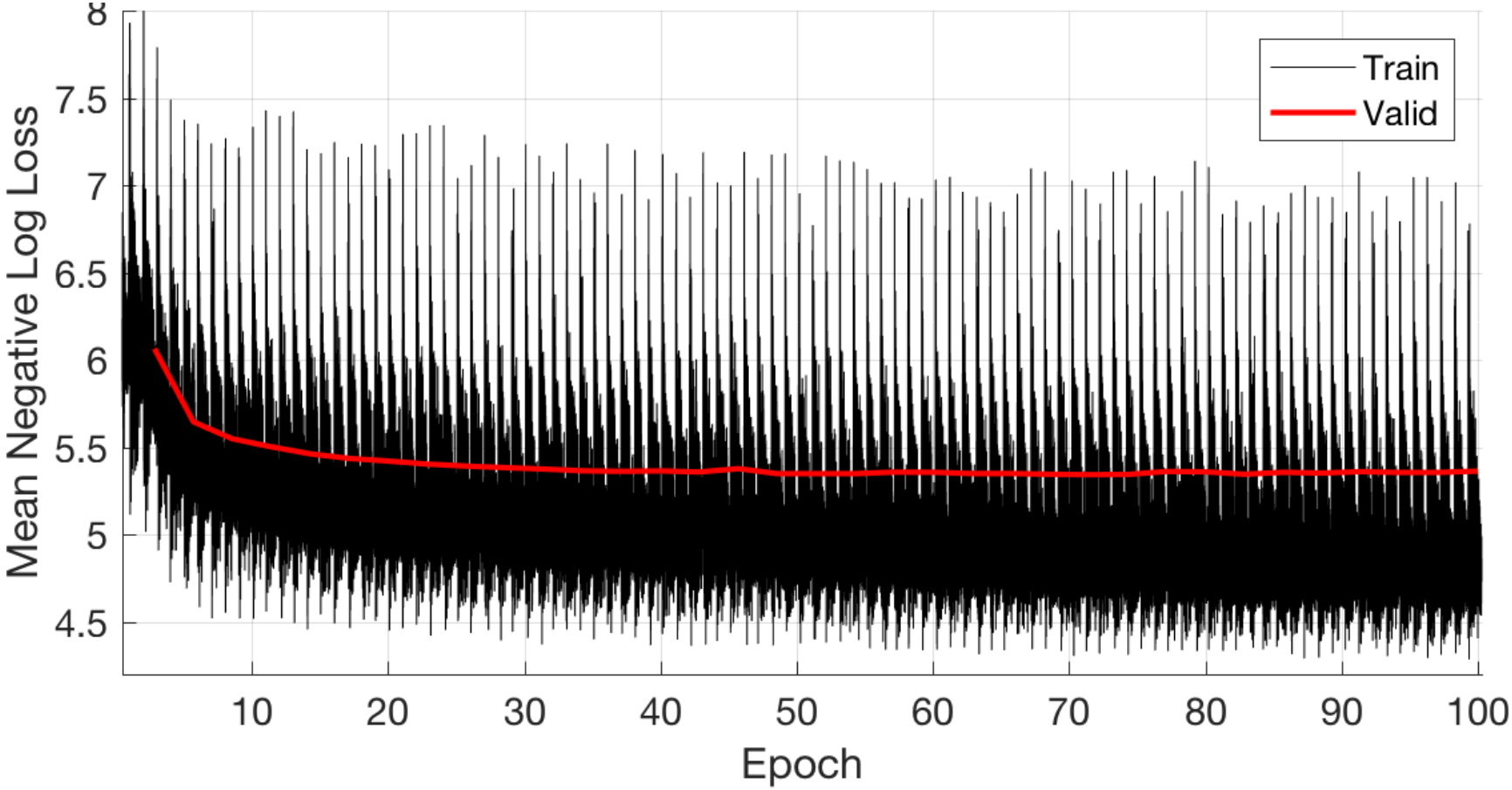
configurations	bugfix	a year ago
data	change to readme	a year ago
metadata	Updated metadata with newer trained model	a year ago
samples	add metadata and samples	2 years ago
soundexamples	Update README.md	ago
.gitignore	clean before checkout	ago
LICENSE	license	ago
README.md	Update README.md	
data_iter.py	added 1hot option	
logger.py	bugfix	
sample_rnn.py	added terminal output, fixed initialisation	
train_rnn.py	bugfix	ago



Sturm, et al., "Music transcription modelling and composition using deep learning," in *Proc. Conf. Computer Simulation of Musical Creativity*, 2016.

https://github.com/IraKorshunova/folk-rnn

Training the network decreases the loss



https://folkrnn.org/

folkRNN

generate a folk tune with a recurrent neural network

PRESS TO GENERATE TUNE

Compose

MODEL

thesession.org (w/ :|

TEMPERATURE

1

SEED

359615

METER

4/4

MODE

C Major

INITIAL ABC

Enter start of tune in ABC notation

FOLK RNN TUNE №1580

enter the folkRNN composition competition

```
X:1580
M:4/4
K:Cmaj
cGGE GFEG|(3CCCCD EGEG|cAdc ecdB|ACDE G3A|
(3cccCG EGGC|EDCD (3EEEG2|cGGG e2dc|(3ABcdB c3d:|
|:e3d cege|f3d Bdga|gfeg ecde|fegc A2GF|
EGG2 cdeg|(3fffaf dafd|e2eg fdBc|dedB c3d:|
```

*The RNN properties were **thesession_with_repeats** with seed **441885** and temperature **1**.*

*The prime tokens were **M:4/4 K:Cmaj**.*

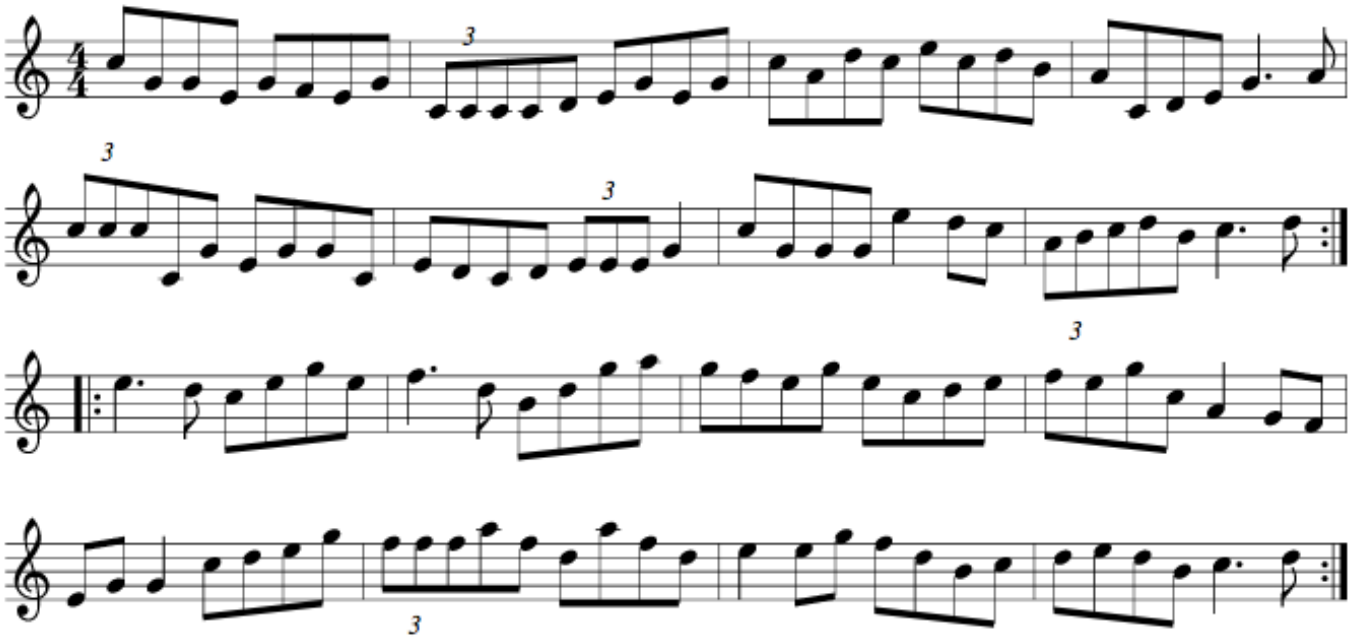
*Generated on **14/06/2018, 14:46:35**.*

HEAR IT

0:00

Download MIDI

SEE IT



https://themachinefolksession.org/

THE *machine folk* SESSION

enter the folkRNN composition competition

TUNES RECORDINGS EVENTS

Bob: log out; submit a tune; tune of the month. Help

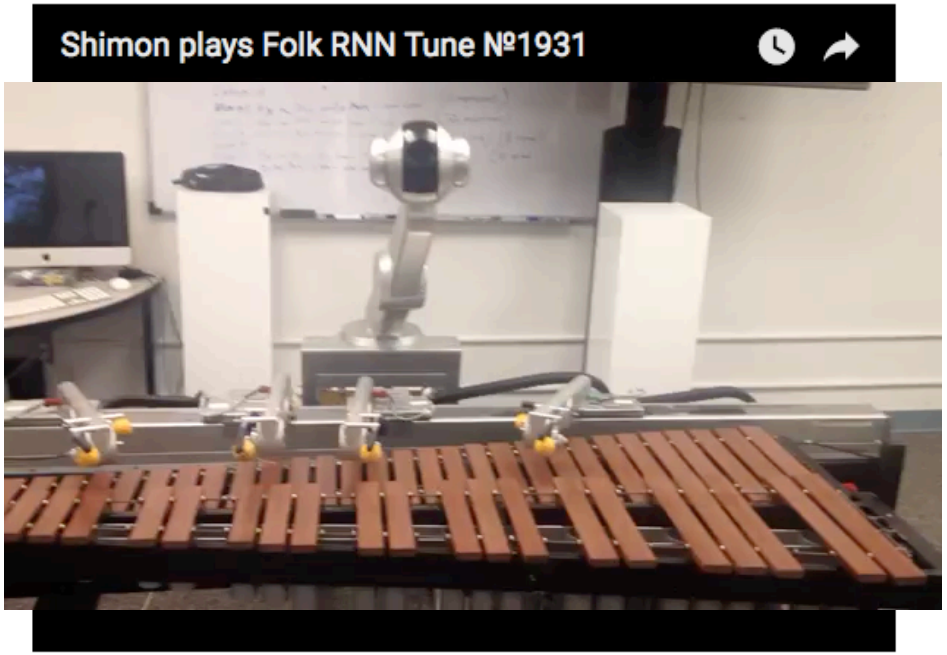
In beta! Launching soon!

Hello Bob

The Machine Folk Session is a community website dedicated to folk music generated by, or co-created with, machines.

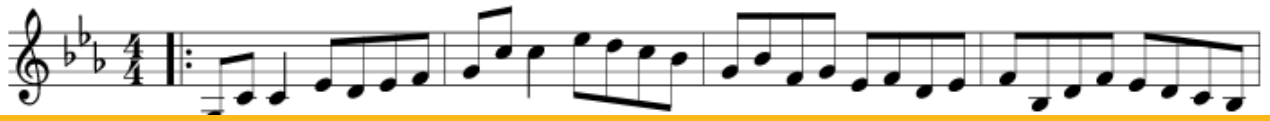
You can find tunes to play, recordings of them, or events where they're played.

machine folk, live



Popular tunes

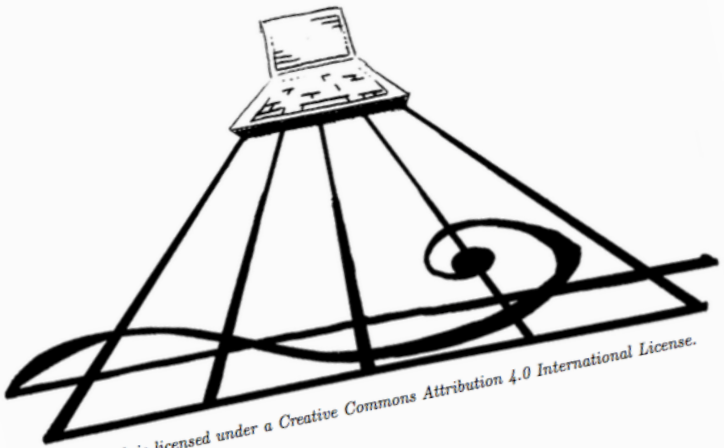
Folk RNN Tune №1931



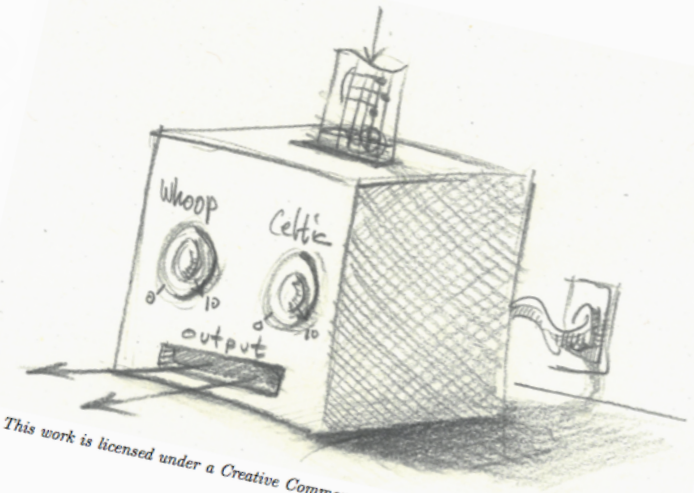
100,000 tunes in 34 volumes

The folk-rnn (v1) Session Book
Volume 1 of 20*

The folk-rnn (v3) Session Book
Volume 1 of 4*



The folk-rnn (v2) Session Book
Volume 1 of 10*



https://highnoongmt.wordpress.com/2018/01/05/volumes-1-20-of-folk-rnn-v1-transcriptions/

loquantur RHYTHM

presents

Folk-RNN

Our artist this month is "**Folk-RNN**" - a **computerized/ machine learning system** "trained" on folk music that composes ORIGINAL music!

Folk-RNN was developed in London, England at Queen Mary and Kingston Universities and has composed over 35,000 original tunes – all based on

https://soundcloud.com/sturmen-1/on-hold-millennial-whoop-reel

An Unintentional Experiment

Feedback Like 11.5M

Tuesday, Jun 20th 2017 8AM 69°F 11AM 75°F 5-Day Forecast

Daily Mail .com

Science & Tech

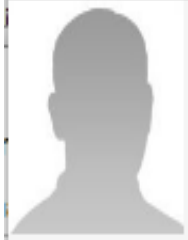
Home | U.K. | News | Sports | U.S. Showbiz | Australia | Femail | Health | Science | Money | Video | Travel | Columnists

Latest Headlines | Science | Pictures | Discounts Login

The future of music: 'Bot Dylan' AI writes its own catchy folk songs after studying 23,000 tunes

- Computer composes new tunes after being trained on 23,000 Irish folk songs
- This allowed AI to learn the patterns and structures that make for a catchy tune
- So far it has created over 100,000 new machine 'folk tunes', researchers say
- It marks a significant step forward for the capabilities of artificial intelligence

https://www.dailymail.co.uk/sciencetech/article-4544400/Researchers-create-computer-writes-folk-music.html



Duane_1981, Preston, United Kingdom, 11 months ago

It's sounds very neat. It's missing the "human" element.

PaxRomana, Novi, 11 months ago

That's it?!!! I'm not impressed.

paevo, USA, United States, 11 months ago

Sounds like a robotic Irish jig....

Mikeyt1941, London, Canada, 11 months ago

Totally lifeless without warmth. Mind you much human tuneless junk that passes for music today isn't much better.

Click to rate 11 1

Fabrice, Manchester, United Kingdom, 11 months ago

No no no.

rocksnoop1, dover, United Kingdom, 11 months ago

Isn't music robotic enough these days?

pen, somewhere, United Kingdom, 11 months ago

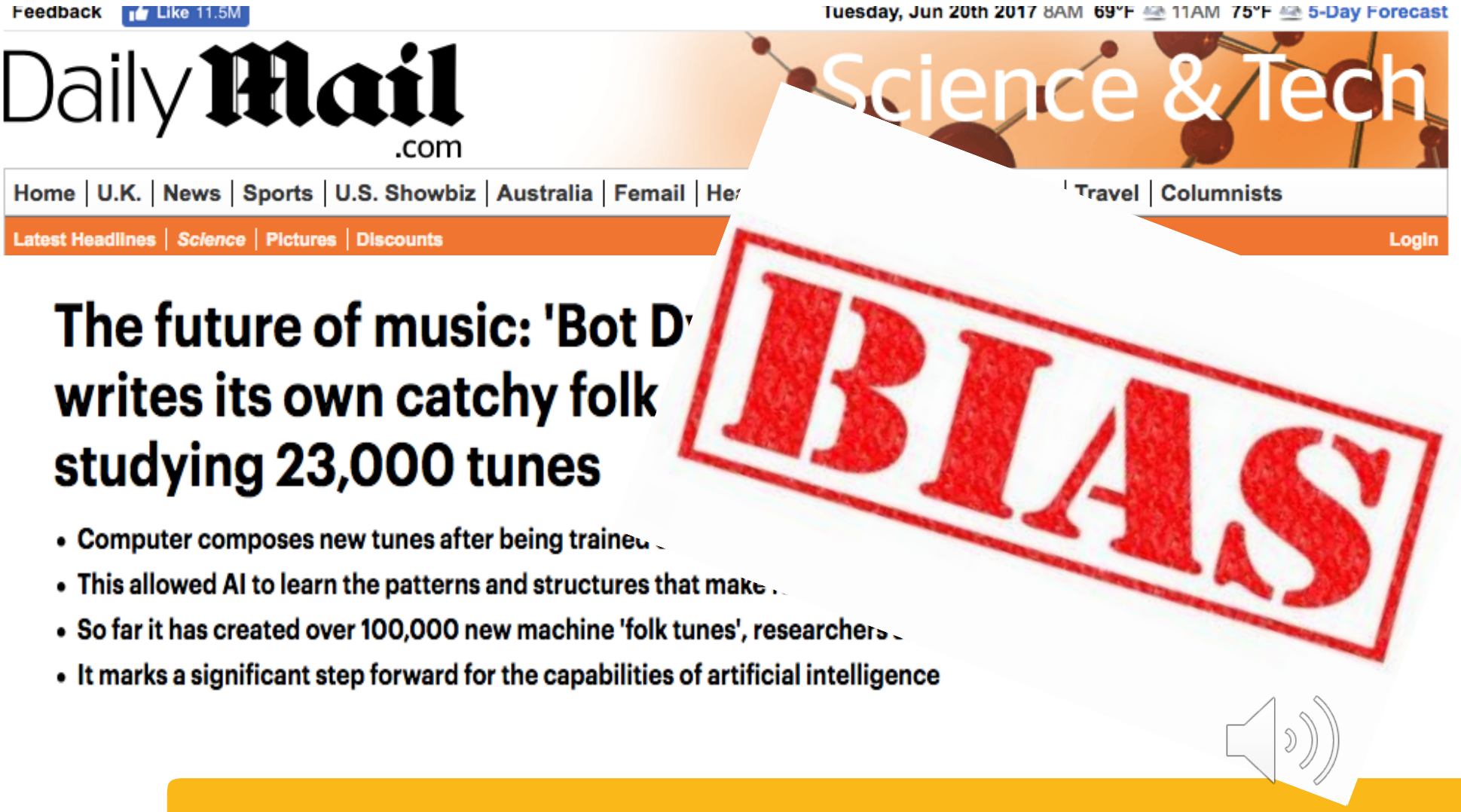
Let's make all humans redundant, brilliant! Has everybody really lost their soul?!

Radar Also, Hemet, 11 months ago

This computerized "AI" is just so non musically untalented lazy nerds can infiltrate the world of true musicians who love, created, and write the music from the joy, hurt, and life emanating from their hearts.

Click to rate 7 1

An Unintentional Experiment



https://www.dailymail.co.uk/sciencetech/article-4544400/Researchers-create-computer-writes-folk-music.html

An Intentional Experiment

How difficult will it be for a professional musician to produce an album using material generated by our AI that will be judged successful within the idiom of Irish traditional music?


The reveal is
coming soon ...

The Resulting Album

Track listing:

1. Gan Ainm, Gan Ainm, Gan Ainm
2. The Drunken Landlady, Gan Ainm, Gan Ainm
3. Gan Ainm, Gan Ainm, Gan Ainm
4. Battle Of Aughrim, Gan Ainm, Lord Mayo
5. Gan Ainm, Gan Ainm, Tom Billy's
6. Girls Of Banbridge, Gallowglass, Gan Ainm
7. The Blackbird, Gan Ainm, Mrs Galvin's
8. Gan Ainm
9. Gan Ainm, Bunch of Green Rushes, Gan Ainm
10. Gan Ainm, Gan Ainm, Anthony Frowley's
11. Gan Ainm, Toss the Feathers (II), Gan Ainm

https://soundcloud.com/oconaillfamilyandfriends

Sturm, B. L. and Ben-Tal, O. (2018). *Let's Have Another Gan Ainm: An experimental album of Irish traditional music and computer-generated tunes*. Technical report, KTH.

https://soundcloud.com/oconaillfamilyandfriends

KTH Översikt Campi Lätt svenska Lexin Shipley The Local NYT WashPost Guardian blog Twitter fb Instagram

Home Stream Library Search Try Pro Upload Ó Conaill Family and F...

All Tracks Albums Playlists Reposts

Station Share Edit

Spotlight (0/5)

Edit Spotlight

Highlight your best tracks and playlists: put them in Spotlight so that your audience will find them first when they visit your profile.

Recent

Ó Conaill Family and Friends
Let's have another Gan Ainm Album · 2018
11 months ago
Folk & Singer-Songw...

3:24

1 01 – Gan Ainm, Gan Ainm, Gan Ainm
2 02 – The Drunken Landlady, Gan Ainm, Gan Ainm

Followers	Following	Tracks
47	1	11

This is an experimental album of Irish traditional music and computer-generated tunes. Each "Gan Ainm" comes from material generated by a computer trained on over 23,000 transcriptions of traditional music from Ireland and the UK. More

Show more

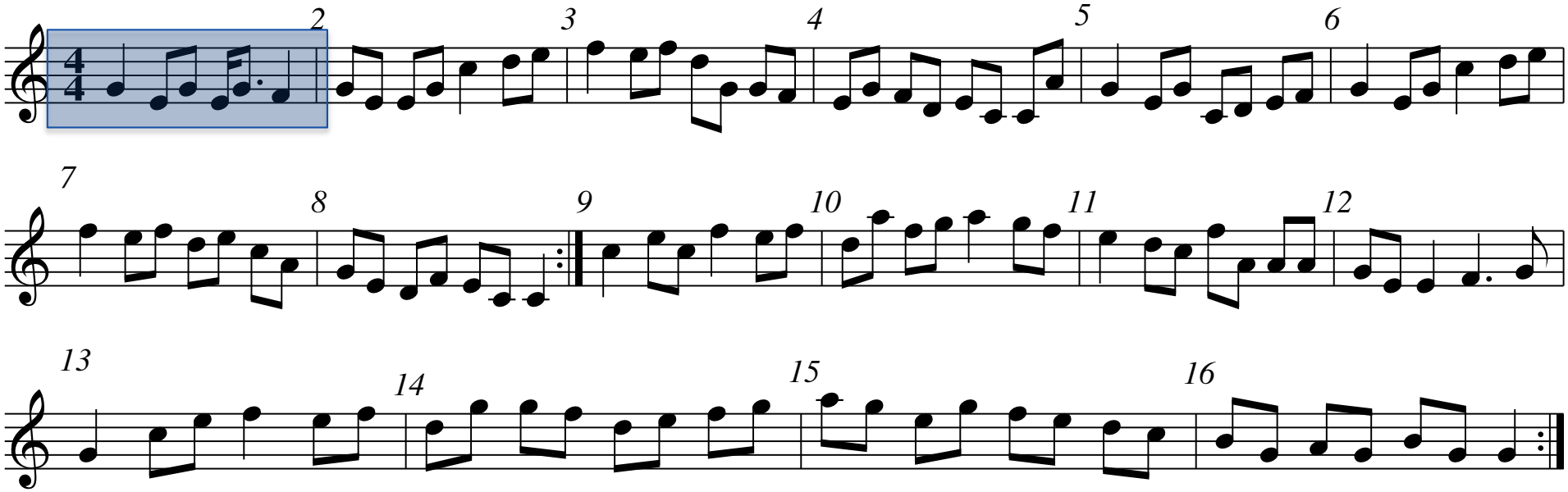
Stats View all

Plays last 24 hours	Plays last 7 days
13	138

15,035 plays in total

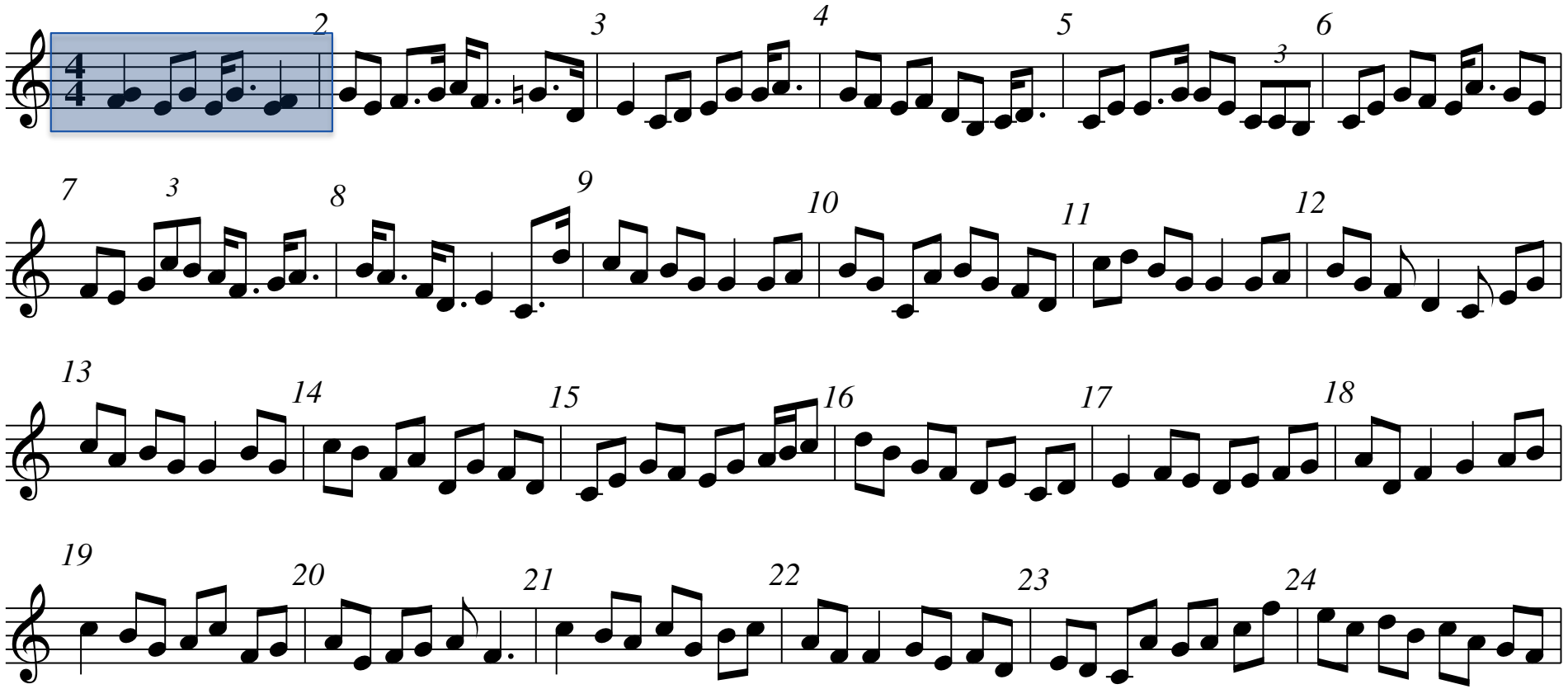
Does it *really* know about music?

Seed: “M:4/4 K:Cmaj G 2 E G E < G F 2 |”



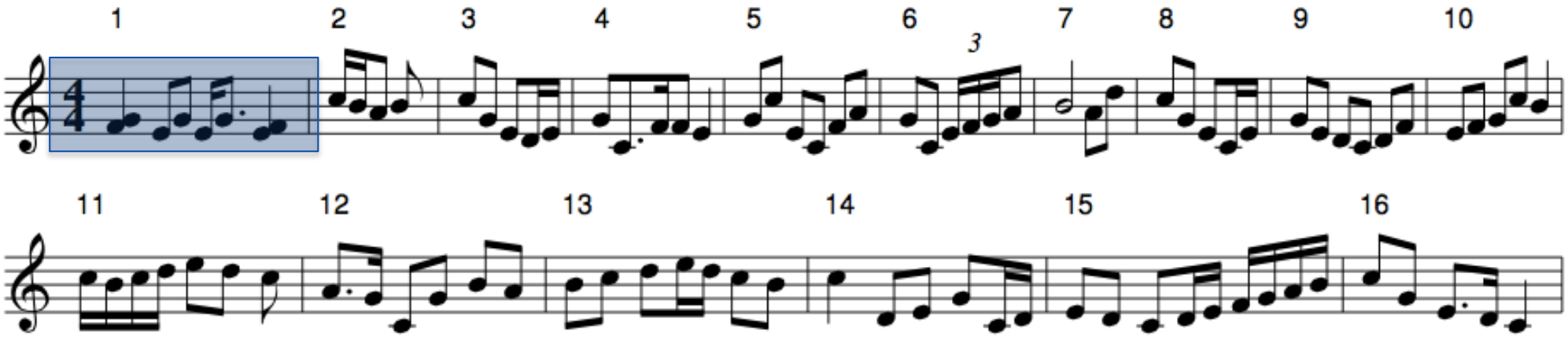
Does it *really* know about music?

Seed: “M:4/4 K:Cmaj [G F] 2 E G E < G [E F] 2 |”



Does it *really* know about music?

Seed: “M:4/4 K:Cmaj [G F] 2 E G E /2 G > [E F] 4 |”



Failure can be more interesting

"The Humours of Time Pigeon"

https://youtu.be/1xBisQK8-3E

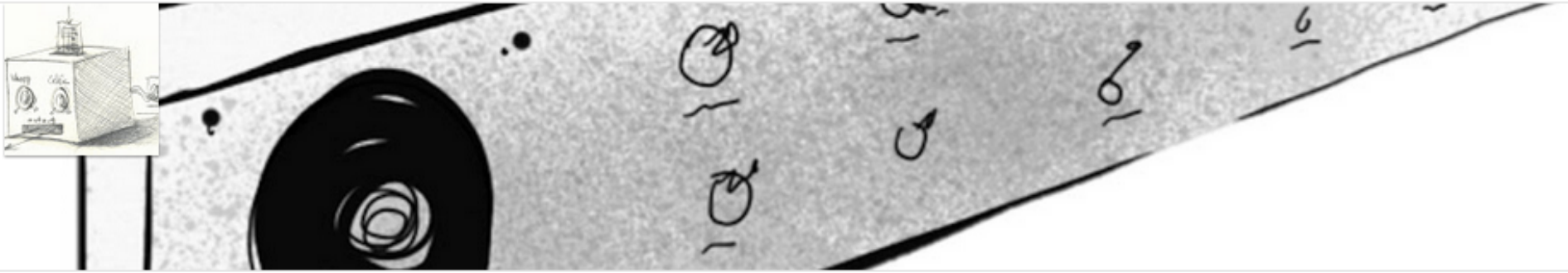
"A Fhsoilah Kilnie"

https://youtu.be/RaO4HpM07hE

YouTube GB

Search

32 subscribers 6,520 views Video Manager https://www.youtube.com/channel/UC7wzmG64y2IbTUeWji_qKhA



The Bottomless Tune Box

Subscribe 32

The music on this channel arises from creative partnerships between musicians and a deep learning system that models music transcription ... Show more

Uploads Public

Set #3 (fast reels)
31 views • 1 week ago

Set #1 (jigs)
69 views • 3 weeks ago

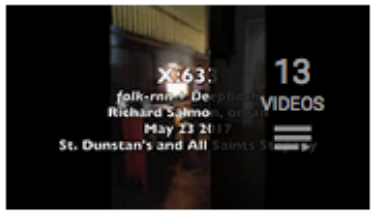
"Chicken Bits and Bits and Bobs" by Bob L. Sturm + folk-rnn
68 views • 3 weeks ago

"Interlude" by Bob L. Sturm + folk-rnn
26 views • 3 weeks ago

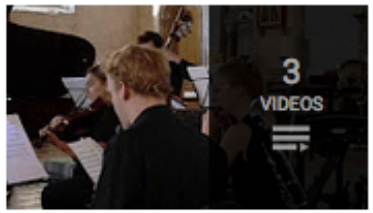
"The Humours of Time Pigeon" by Bob L. Sturm + folk-rnn
41 views • 3 weeks ago

Created playlists Public

C4DM concert (QMUL Nov. 18 2016)



Partnerships concert (May 23 2017)



Two short pieces and an Interlude in Concert

Bastard Tunes in Concert

C4DM concert (QMUL Nov 23 2015)

KTH Hub @ Ars Electronica 2020

https://ars.electronica.art/keplersgardens/en/kth/

The 2020 Joint Conference on AI Music Creativity, Oct 19-23

The 2020 Joint Conference on AI Music Creativity

October 19–23, 2020 organized and hosted virtually by the Royal Institute of Technology (KTH), Stockholm, Sweden

Programme
Online Exhibit
Call for papers, tutorials, panels and musical works
AI Music Generation Challenge 2020
Registration
Organization
Questions

Financial support comes from ERC-2019-COG No. 864189 MUSAiC: Music at the Frontiers of Artificial Creativity and Criticism

The 2020 Joint Conference on AI Music Creativity will be entirely virtual.

This conference brings together for the first time two overlapping research forums: The Computer Simulation of Music Creativity conference (est. 2016), and The International Workshop on Musical Metacreation (est. 2012). The principal goal is to bring together scholars and artists interested in the virtual emulation of musical creativity and its use for music creation, and to provide an interdisciplinary platform to promote, present and discuss their work in scientific and artistic contexts.

The computational simulation of musical creativity continues to be an exciting and significant area of academic research, and is now making impacts in commercial realms. Such systems pose several theoretical and technical challenges, and are the result of an interdisciplinary effort that encompasses the domains of music, artificial intelligence, cognitive science and philosophy. This can be seen within the broader realm of Musical Metacreation, which studies the design and use of such generative tools and theories for music making: discovery and exploration of novel musical styles and content, collaboration between human performers and creative software "partners", and design of systems in gaming and entertainment that dynamically generate or modify music.

The 2020 Joint Conference on AI Music Creativity will be **virtual**. It will consist of synchronous and asynchronous events. The five-day program will feature tutorials, research paper presentations, discussion panels, an online exhibition, nine invited spotlight presentations and two keynotes.

https://boblsturm.github.io/aimusic2020/

BEWARE OF SKITSNACK!

WE'RE LIVING IN THE FUTURE NOW!

FACEPTION
Facial Personality Analytics

Our Technology | Verticals | About us | News & Events | Blog | Contact us

FACEPTION IS A FACIAL PERSONALITY ANALYTICS TECHNOLOGY COMPANY

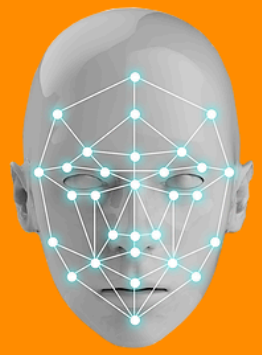
We reveal personality from facial images at scale to revolutionize how companies, organizations and even robots understand people and dramatically improve public safety, communications, decision-making, and experiences.

WE'RE LIVING IN THE FUTURE NOW!

FACEPTION
Facial Personality Analytics

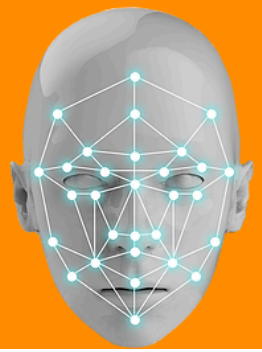
Our Technology | Verticals | About us | News & Events | Blog | Contact us

OUR CLASSIFIERS



HIGH IQ

Endowed with a reasoning skills, like logic, spatial skills. Self-made people, free-thinkers and entrepreneurs. Exceptionally gifted, tend to be less socially oriented, value truth, facts and logic more than emotional relations. Creative and independent minded, with exceptional concentration abilities, a high intellect and mental capacity



Academic Researcher

Endowed with sequential thinking, high analytical abilities, a multiplicity of ideas, deep thoughts and seriousness. Creative, with a high concentration ability, high mental capacity, and interest in data and information.

WE'RE LIVING IN THE FUTURE NOW!

FACEPTION
Facial Personality Analytics

Our Technology | Verticals | About us | News & Events | Blog | Contact us

OUR CLASSIFIERS

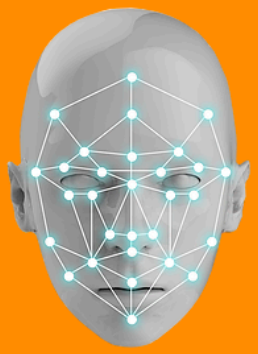
Professional Poker Player

Endowed with a high concentration ability, perseverance and patience. Goal-oriented, analytical, with a dry sense of humor. Silent, devoid of emotion and emotional expression, strict and sharp minded, with high critical perception.

Bingo Player

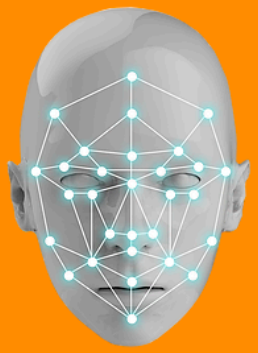
Endowed with a high mental ceiling, high concentration, adventurousness, and strong analytical abilities. Tends to be creative, with a high originality and imagination, high conservation and sharp senses.

and information.



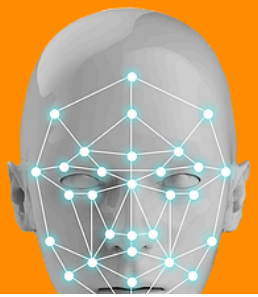
White-Collar Offender

Tends to have a low self-esteem, a high IQ and charisma. Anxious, tensed and frustrated, competitive, ambitious and dominant. Usually loves to take risks and have a dry sense of humor.



Terrorist

Suffers from a high level of anxiety and depression. Introverted, lacks emotion, calculated, tends to pessimism, with low self-esteem, low self image and mood swings.



Pedophile

Suffers from a high level of anxiety and depression. Introverted, lacks emotion, calculated, tends to pessimism, with low self-esteem, low self image and mood swings.

WE'RE LIVING IN THE FUTURE NOW!

FACEPTION
Facial Personality Analytics

Our Technology Verticals About us News & Events Blog t us

SKITSNACK

ACIAL PERSONALITY ANALYTICS TECHNOLOGY COMPANY

We reveal personality from facial images at scale to revolutionize how companies, organizations and even robots understand people and dramatically improve public safety, communications, decision-making, and experiences.