Today

Lecture 1

Using Machine Learning to Explore Neural Data

Lecture 2

Magnets, Machines, Brains

Friday

Lecture 3

Deep Learning

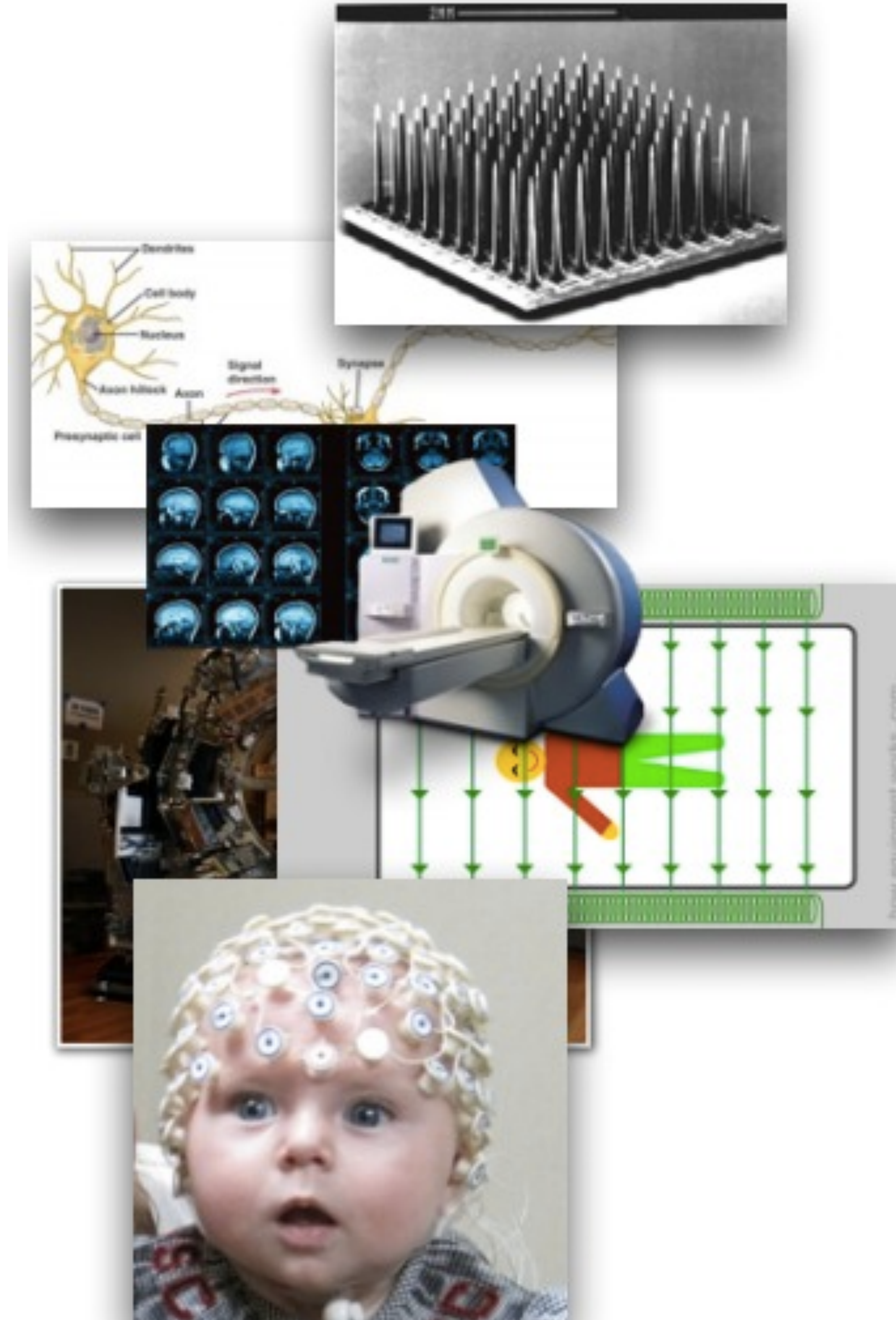
Using Machine Learning to Explore Neural Data

Raul Vicente

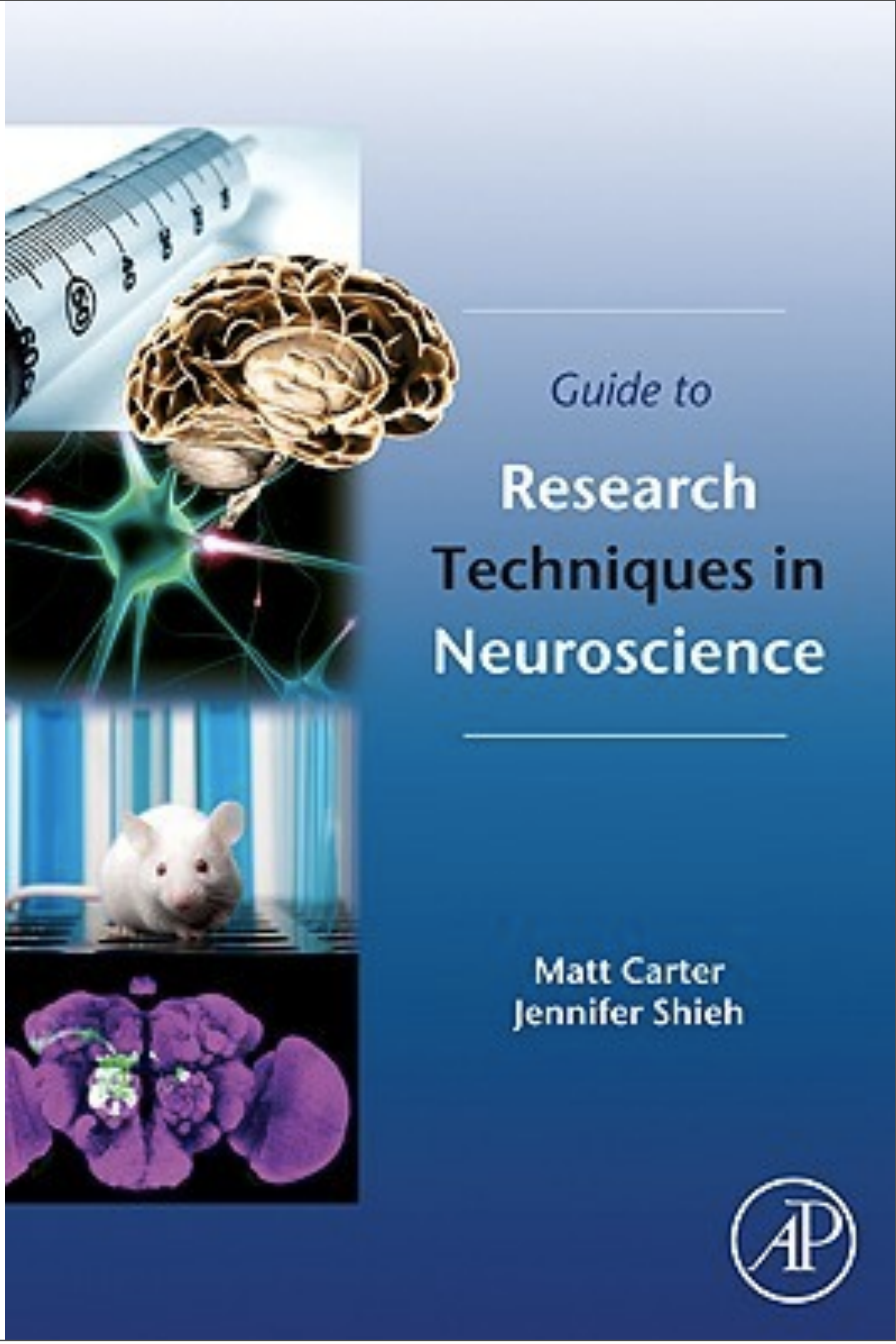
Institute of Computer Science, University of Tartu

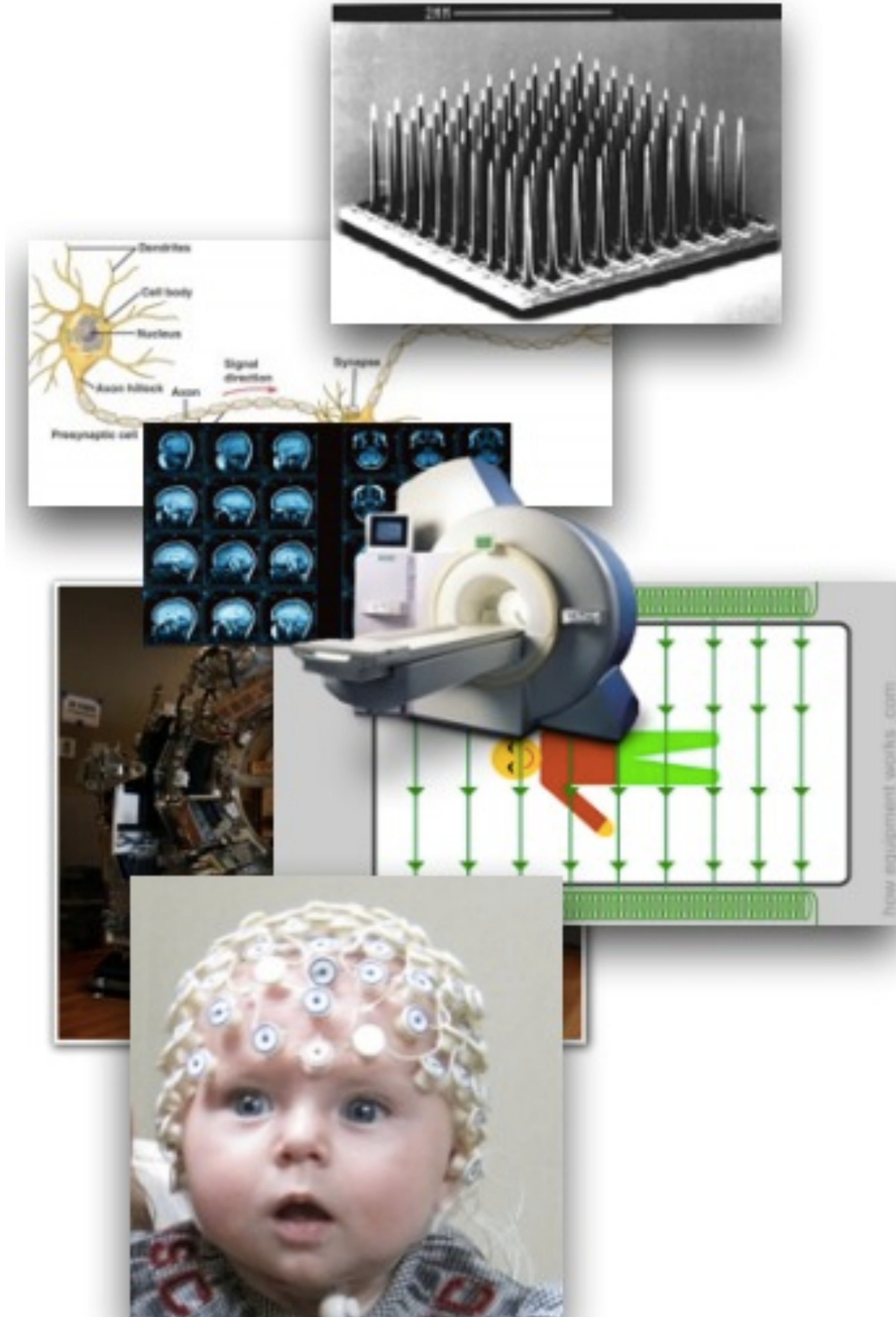
08	02	22	97	38	15	00	40	00	75	04	05	07	78	52	12	50	77	91	08
49	49	99	40	17	81	18	57	60	87	17	40	98	43	69	48	04	56	62	00
81	49	31	73	55	79	14	29	93	71	40	67	53	88	30	03	49	13	36	65
52	70	95	23	04	60	11	42	69	24	68	56	01	32	56	71	37	02	36	91
22	31	16	71	51	67	63	89	41	92	36	54	22	40	40	28	66	33	13	80
24	47	32	60	99	03	45	02	44	75	33	53	78	36	84	20	35	17	12	50
32	98	81	28	64	23	67	10	26	38	40	67	59	54	70	66	18	38	64	70
67	26	20	68	02	62	12	20	95	63	94	39	63	08	40	91	66	49	94	21
24	55	58	05	66	73	99	26	97	17	78	78	96	83	14	88	34	89	63	72
21	36	23	09	75	00	76	44	20	45	35	14	00	61	33	97	34	31	33	95
78	17	53	28	22	75	31	67	15	94	03	80	04	62	16	14	09	53	56	92
16	39	05	42	96	35	31	47	55	58	88	24	00	17	54	24	36	29	85	57
86	56	00	48	35	71	89	07	05	44	44	37	44	60	21	58	51	54	17	58
19	80	81	68	05	94	47	69	28	73	92	13	86	52	17	77	04	89	55	40
04	52	08	83	97	35	99	16	07	97	57	32	16	26	26	79	33	27	98	66
88	36	68	87	57	62	20	72	03	46	33	67	46	55	12	32	63	93	53	69
04	42	16	73	38	25	39	11	24	94	72	18	08	46	29	32	40	62	76	36
20	69	36	41	72	30	23	88	34	62	99	69	82	67	59	85	74	04	36	16
20	73	35	29	78	31	90	01	74	31	49	71	48	86	81	16	23	57	05	54
01	70	54	71	83	51	54	69	16	92	33	48	61	43	52	01	89	19	67	48

08	02	22	97	38	15	00	40	00	75	04	05	07	78	52	12	50	77	91	08
49	49	99	40	17	81	18	57	60	87	17	40	98	43	69	48	04	56	62	00
81	49	31	73	55	79	14	29	93	71	40	67	53	88	30	03	49	13	36	65
52	70	95	23	04	60	11	42	69	24	68	56	01	32	56	71	37	02	36	91
22	31	16	71	51	67	63	89	41	92	36	54	22	40	40	28	66	33	13	80
24	47	32	60	99	03	45	02	44	75	33	53	78	36	84	20	35	17	12	50
32	98	81	28	64	23	67	10	26	38	40	67	59	54	70	66	18	38	64	70
67	26	20	68	02	62	12	20	95	63	94	39	63	08	40	91	66	49	94	21
24	55	58	05	66	73	99	26	97	17	78	78	96	83	14	88	34	89	63	72
21	36	23	09	75	00	76	44	20	45	35	14	00	61	33	97	34	31	33	95
78	17	53	28	22	75	31	67	15	94	03	80	04	62	16	14	09	53	56	92
16	39	05	42	96	35	31	47	55	58	88	24	00	17	54	24	36	29	85	57
86	56	00	48	35	71	89	07	05	44	44	37	44	60	21	58	51	54	17	58
19	80	81	68	05	94	47	69	28	73	92	13	86	52	17	77	04	89	55	40
04	52	08	83	97	35	99	16	07	97	57	32	16	26	26	79	33	27	98	66
88	36	68	87	57	62	20	72	03	46	33	67	46	55	12	32	63	93	53	69
04	42	16	73	38	25	39	11	24	94	72	18	08	46	29	32	40	62	76	36
20	69	36	41	72	30	23	88	34	62	99	69	82	67	59	85	74	04	36	16
20	73	35	29	78	31	90	01	74	31	49	71	48	86	81	16	23	57	05	54
01	70	54	71	83	51	54	69	16	92	33	48	61	43	52	01	89	19	67	48



Go and read this book!



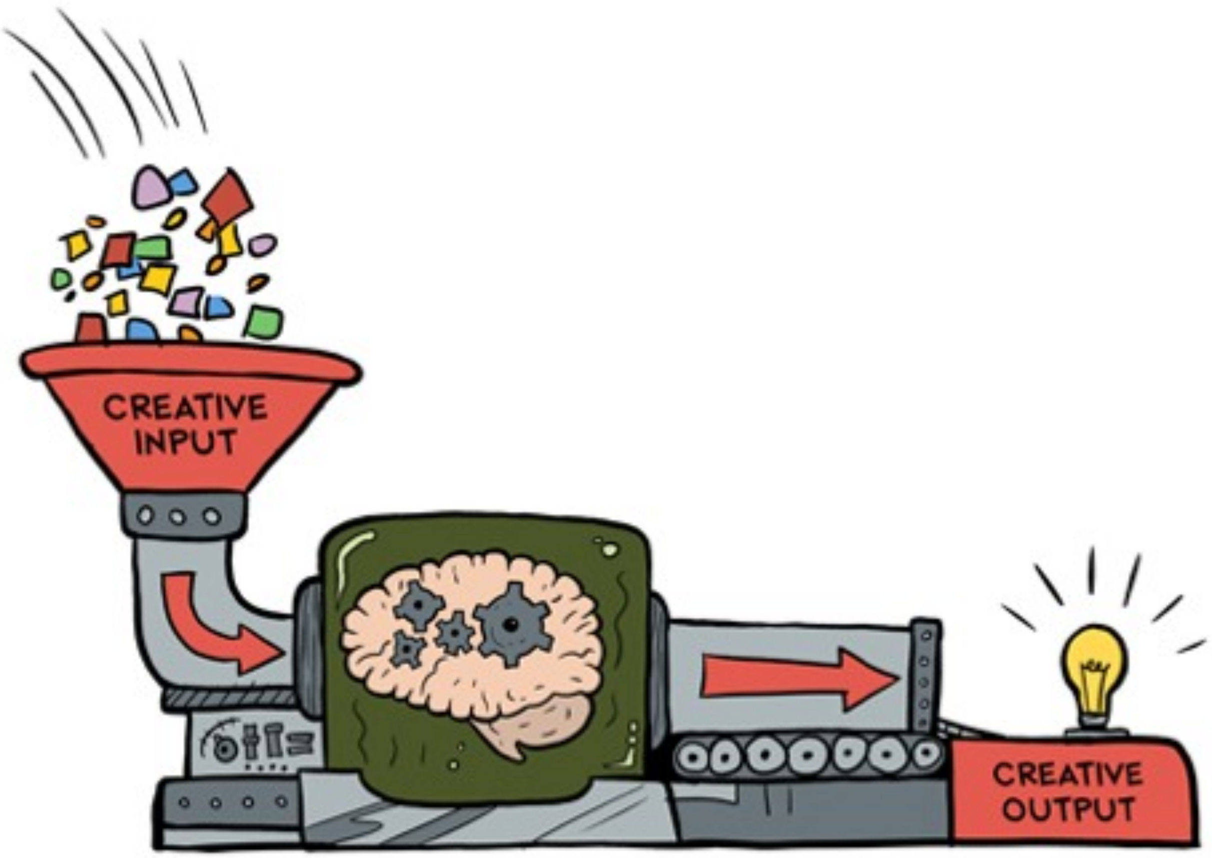

Dendrites
Cell body
Nucleus
Signal direction
Synapse
Axon hillock
Axon
Presynaptic cell

Stimuli

Photons

Pressure

Chemicals



Behavior

Movement

Disease state

Neuroimaging

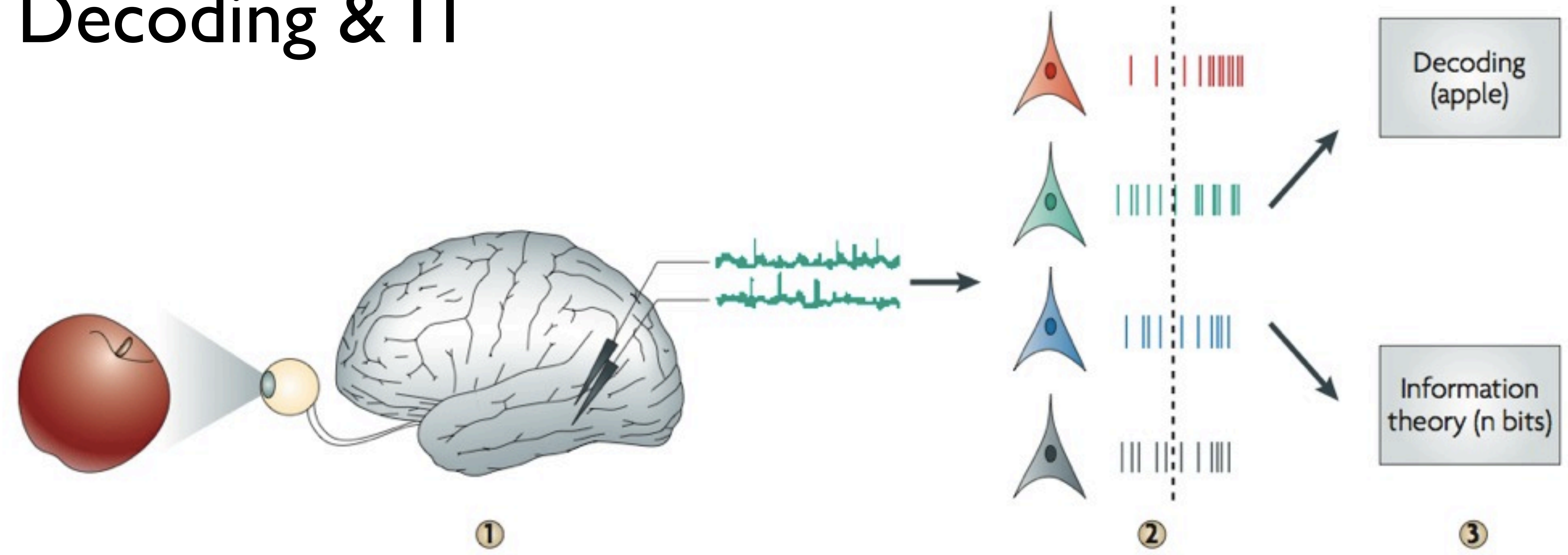
EEG / LFP / MEG

SUA / MUA

Ca^{++} / VSD

(f)MRI

Decoding & IT

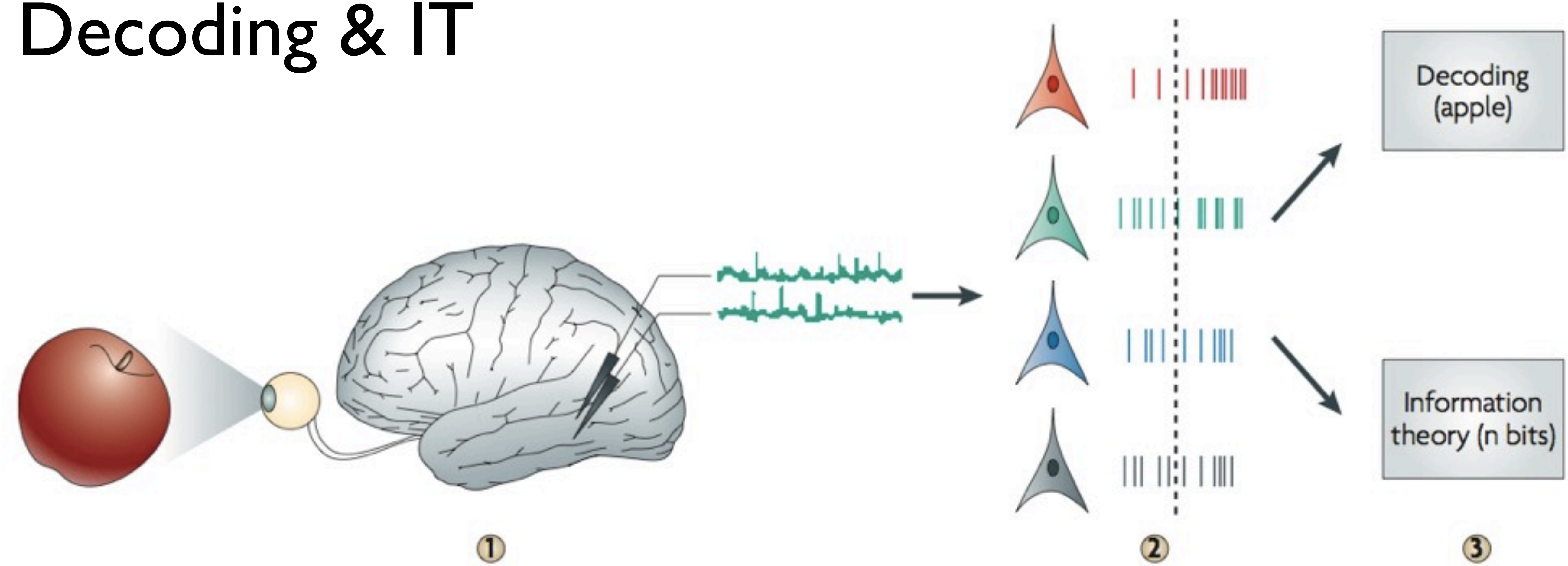


Extracting information from neuronal populations: information theory and decoding approaches

Rodrigo Quian Quiroga & Stefano Panzeri

Nature Reviews Neuroscience **10**, 173-185 (March 2009)

Decoding & IT



Extracting information from neuronal populations: information theory and decoding approaches
Rodrigo Quian Quiroga & Stefano Panzeri
Nature Reviews Neuroscience **10**, 173-185 (March 2009)

D: predict which stimulus or behavior elicits an observed neural response (ex., classifiers in % accuracy)

IT: reduction of uncertainty about the stimulus obtained by knowing the neural response (ex., mutual information in bits)

Why decoding?

Basic

- What aspects of a stimulus are important for each brain area?
- Where? When?

Practical

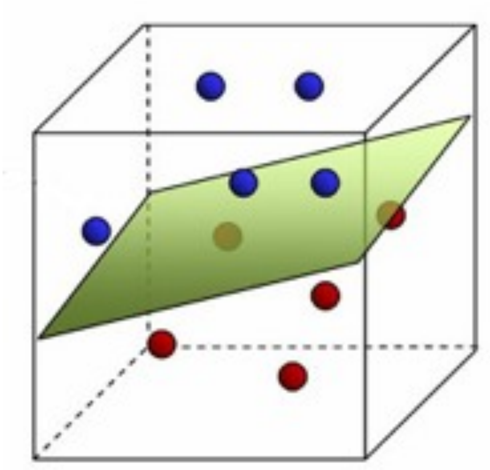
- Neural prosthetics

    neural signals into movements

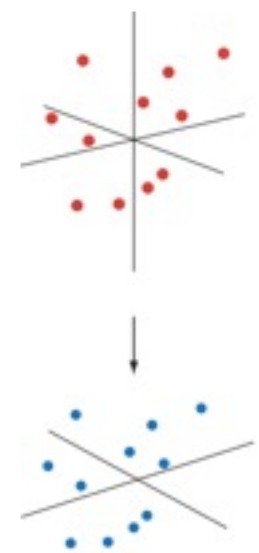
    stimuli into neural signals

3 themes

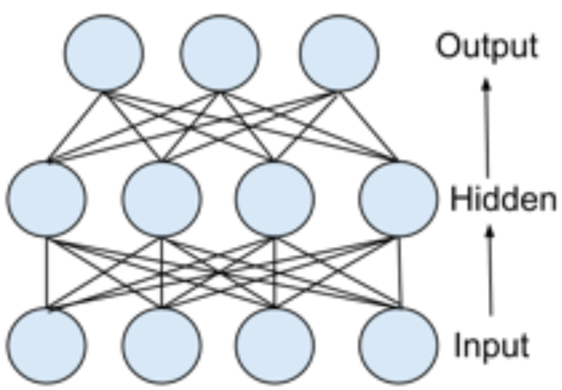
Machine learning



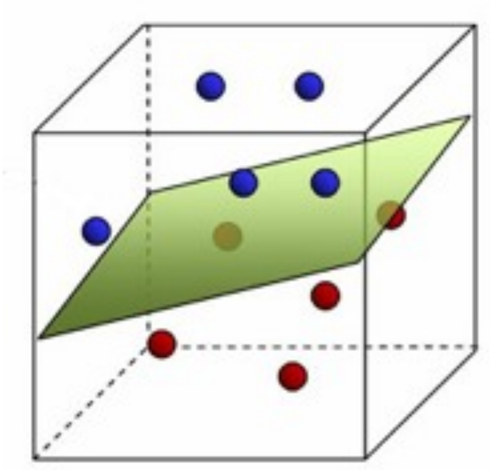
Dimensionality reduction



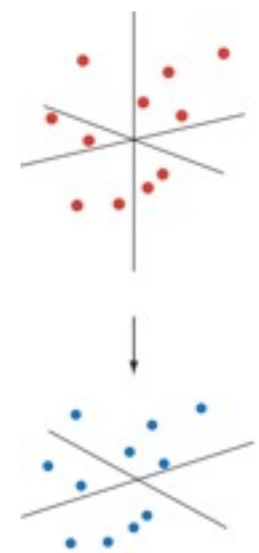
Caveats



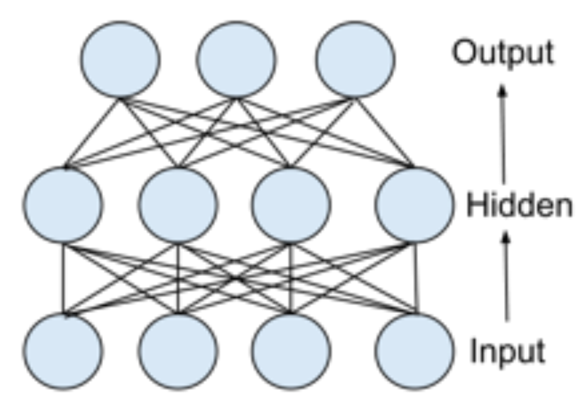
Machine learning



Dimensionality reduction

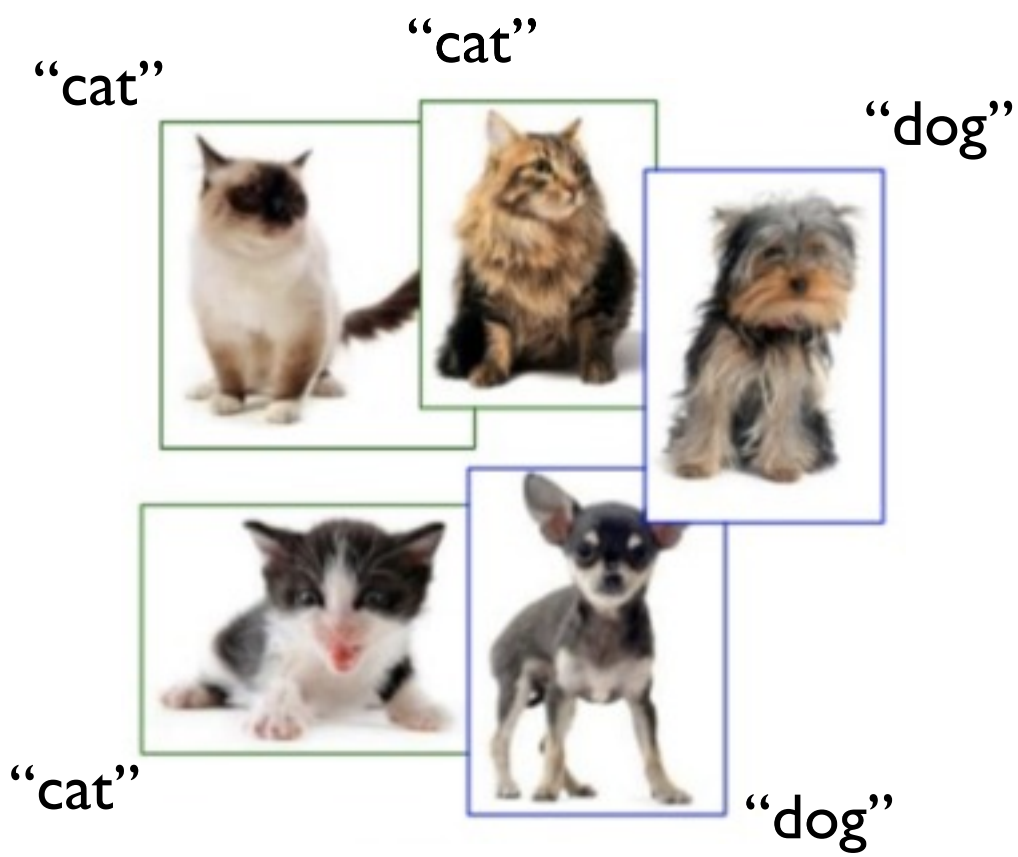


Caveats



Machine Learning: algorithms that learn from data

- Supervised



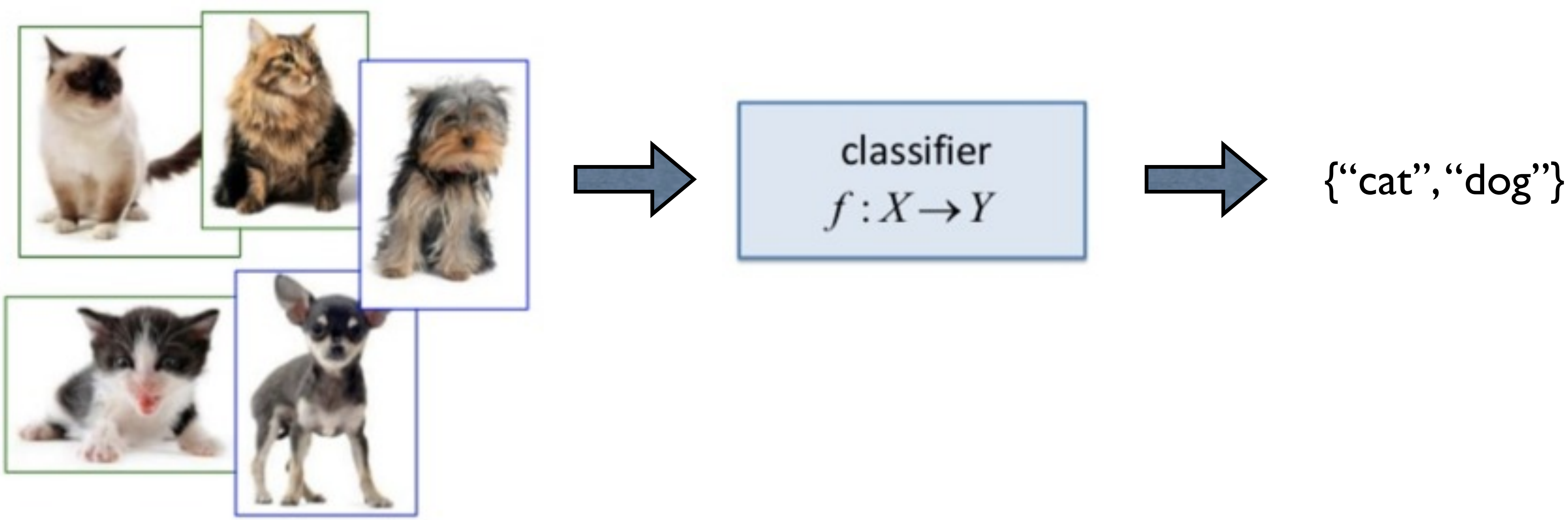
Machine Learning: algorithms that learn from data

- Supervised

{"cat", "dog"}

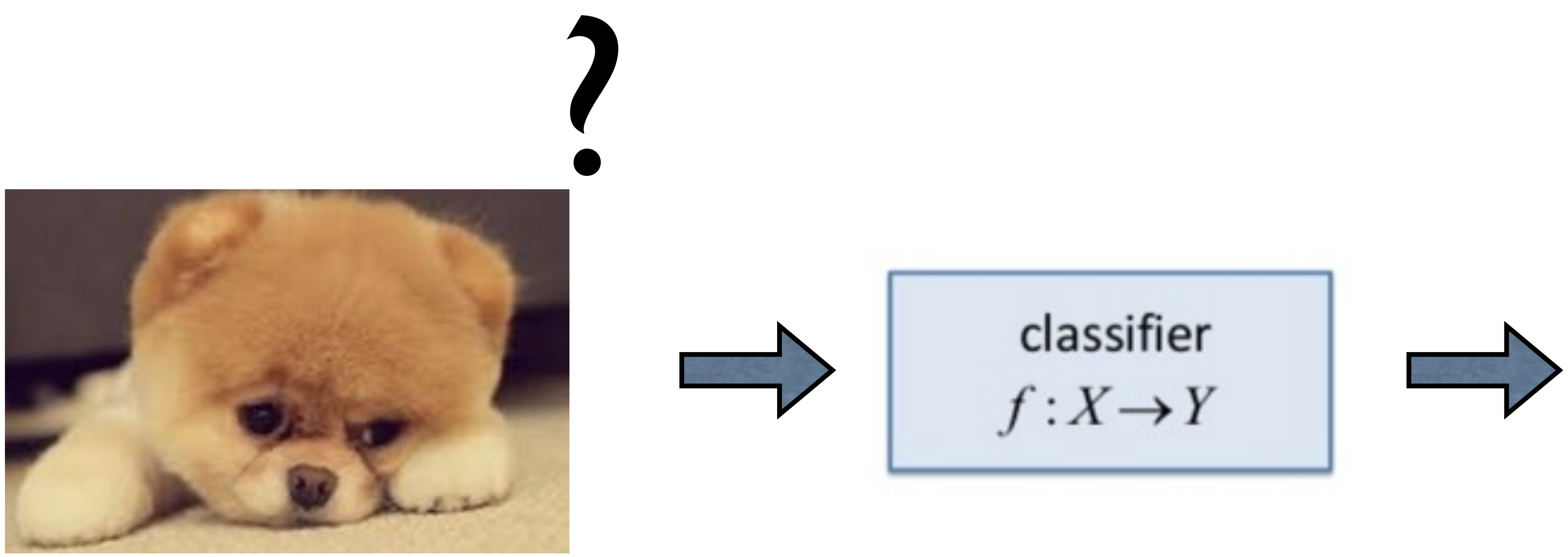
Machine Learning: algorithms that learn from data

- Supervised



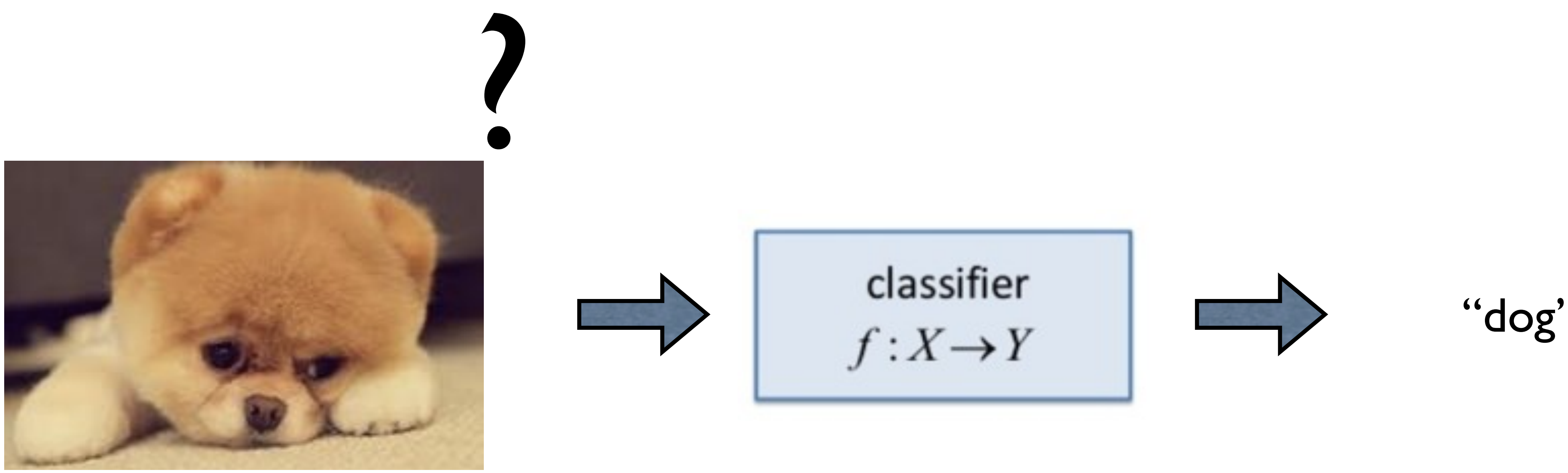
Machine Learning: algorithms that learn from data

- Supervised



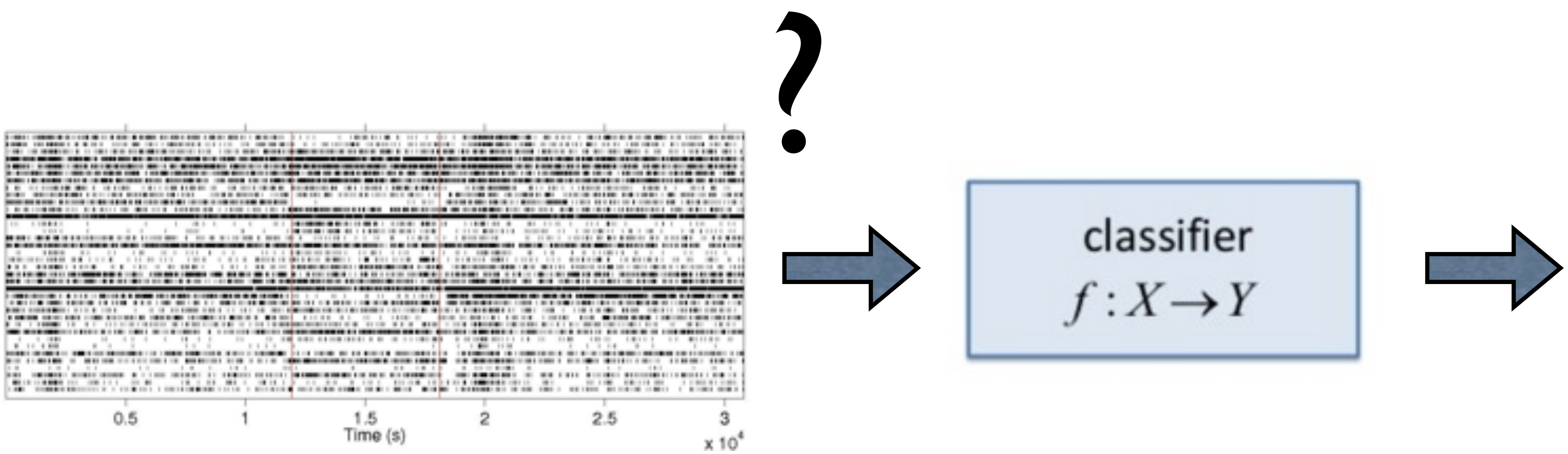
Machine Learning: algorithms that learn from data

- Supervised



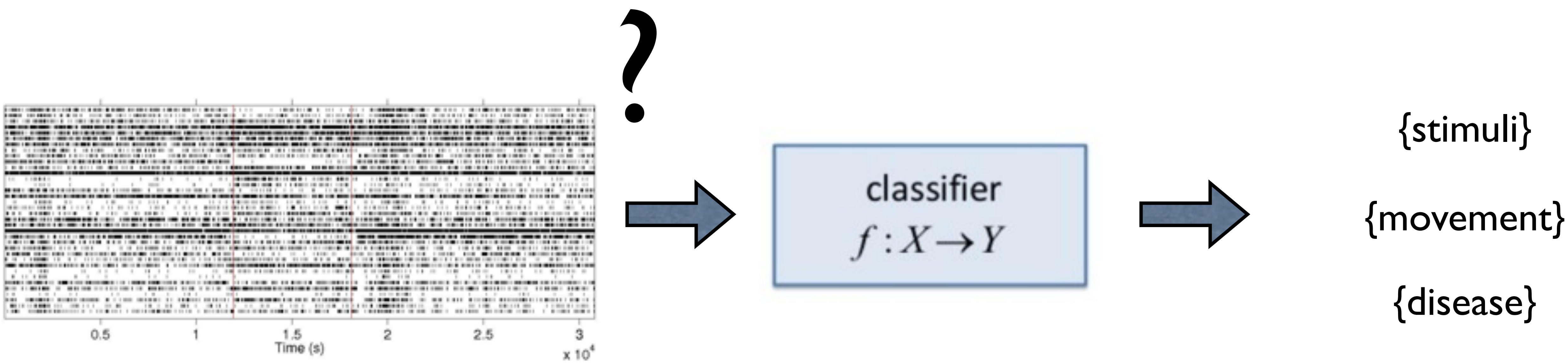
Machine Learning: algorithms that learn from data

- Supervised



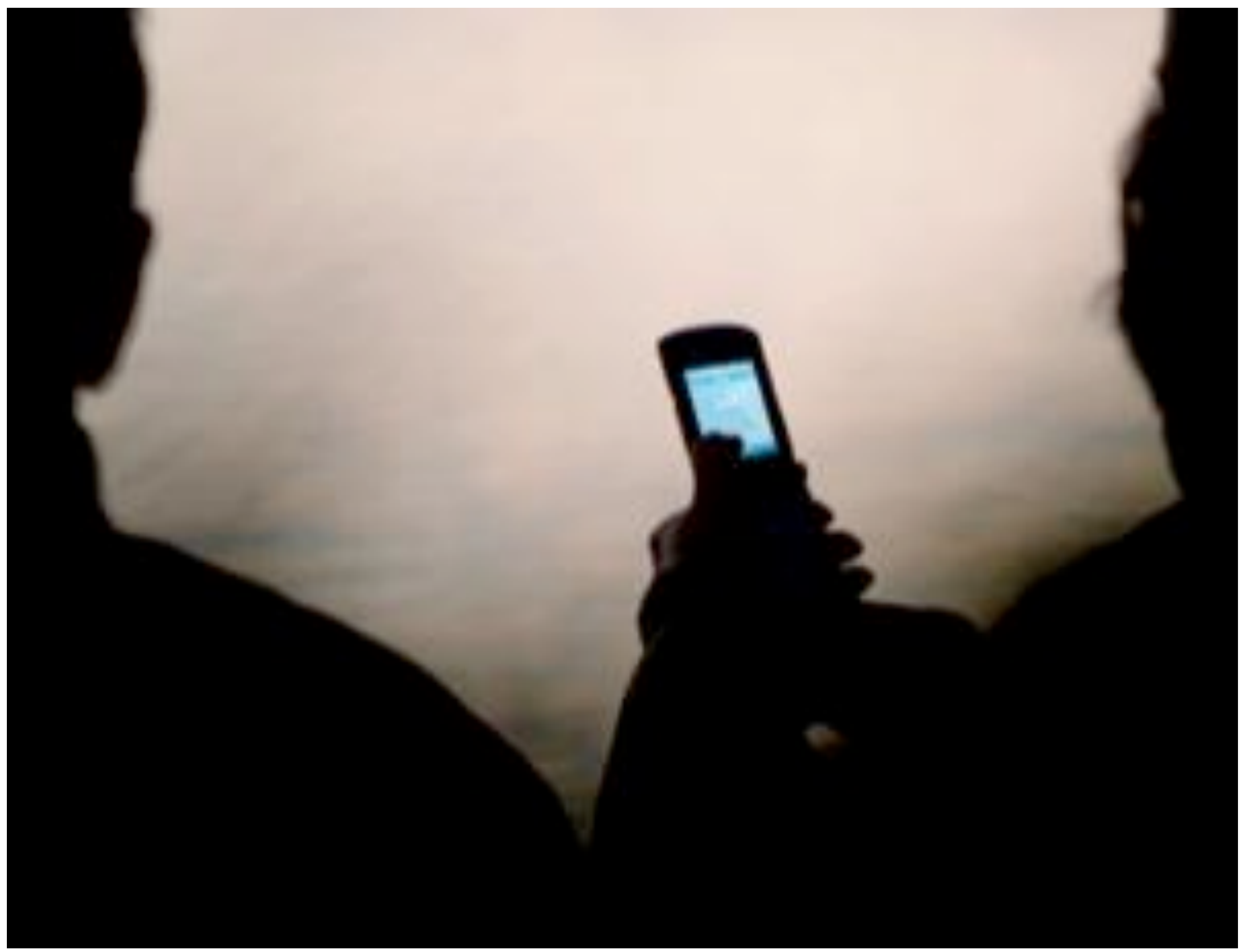
Machine Learning: algorithms that learn from data

- Supervised



Machine Learning: algorithms that learn from data

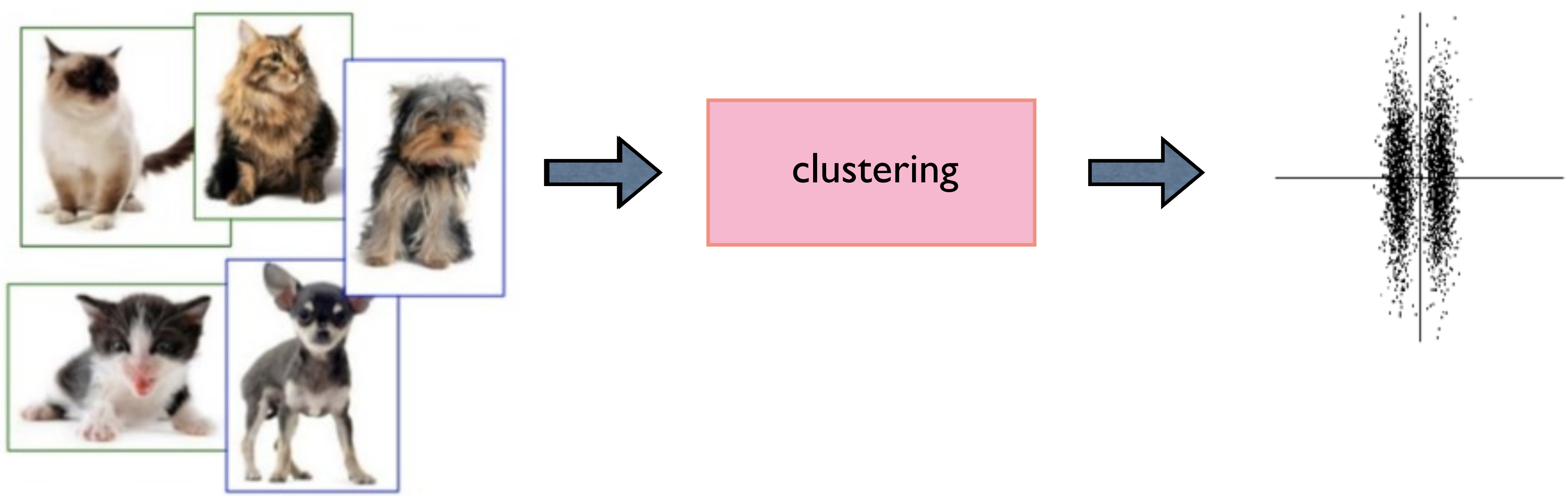
- Supervised



Parkinson tested from phone calls... >95% accuracy

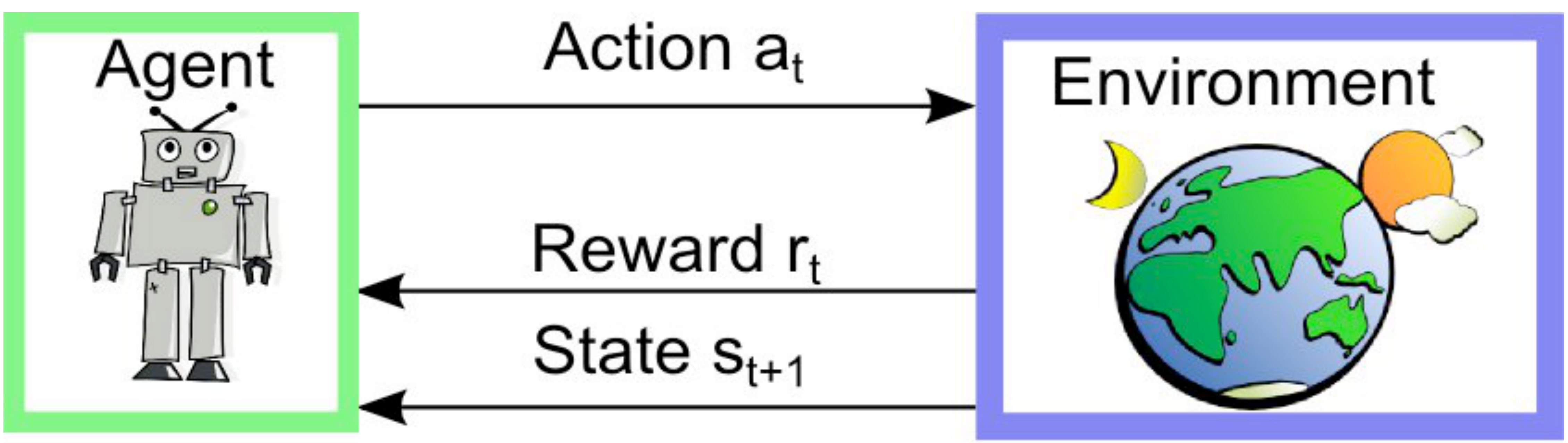
Machine Learning: algorithms that learn from data

- Supervised
- Unsupervised

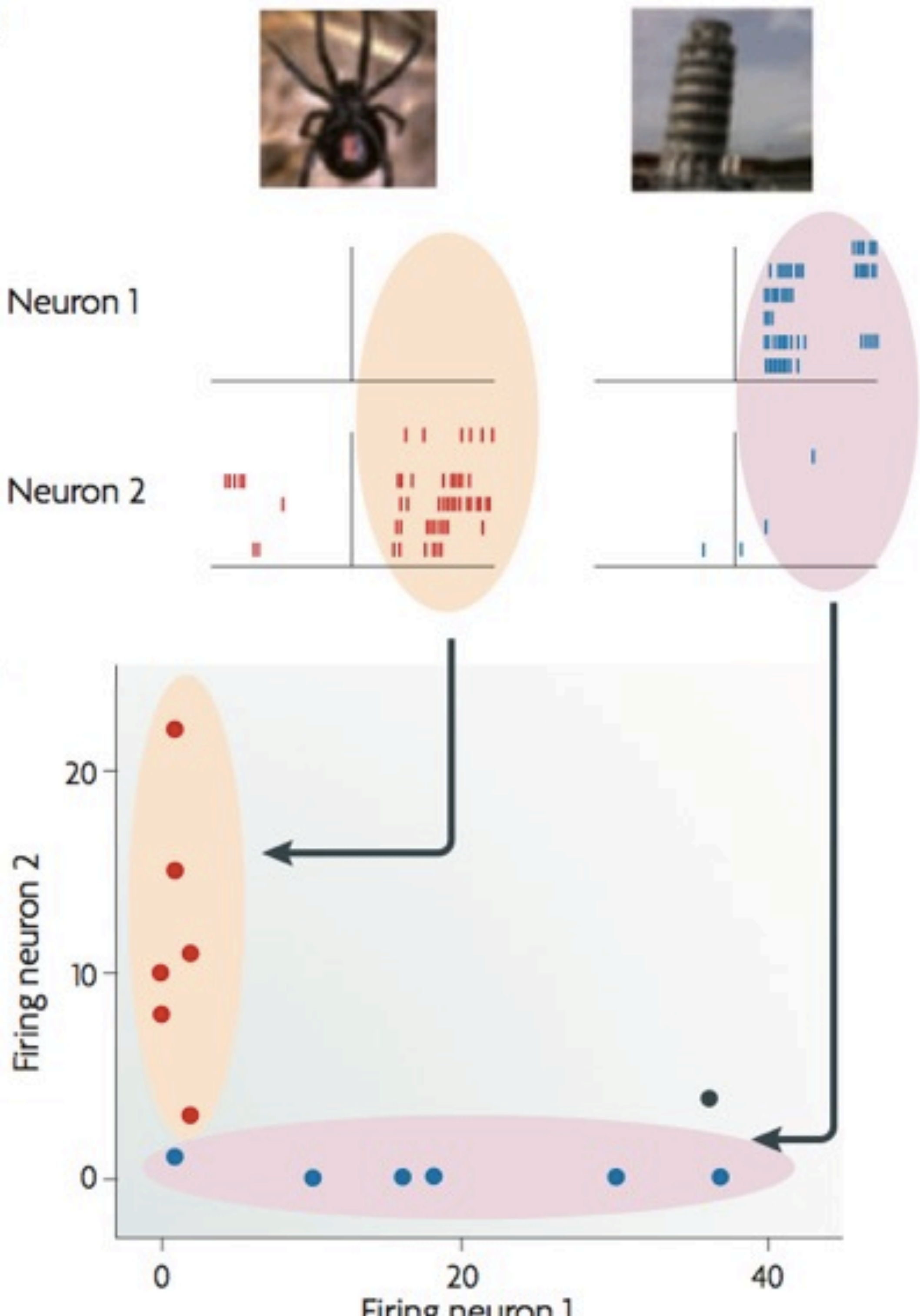


Machine Learning: algorithms that learn from data

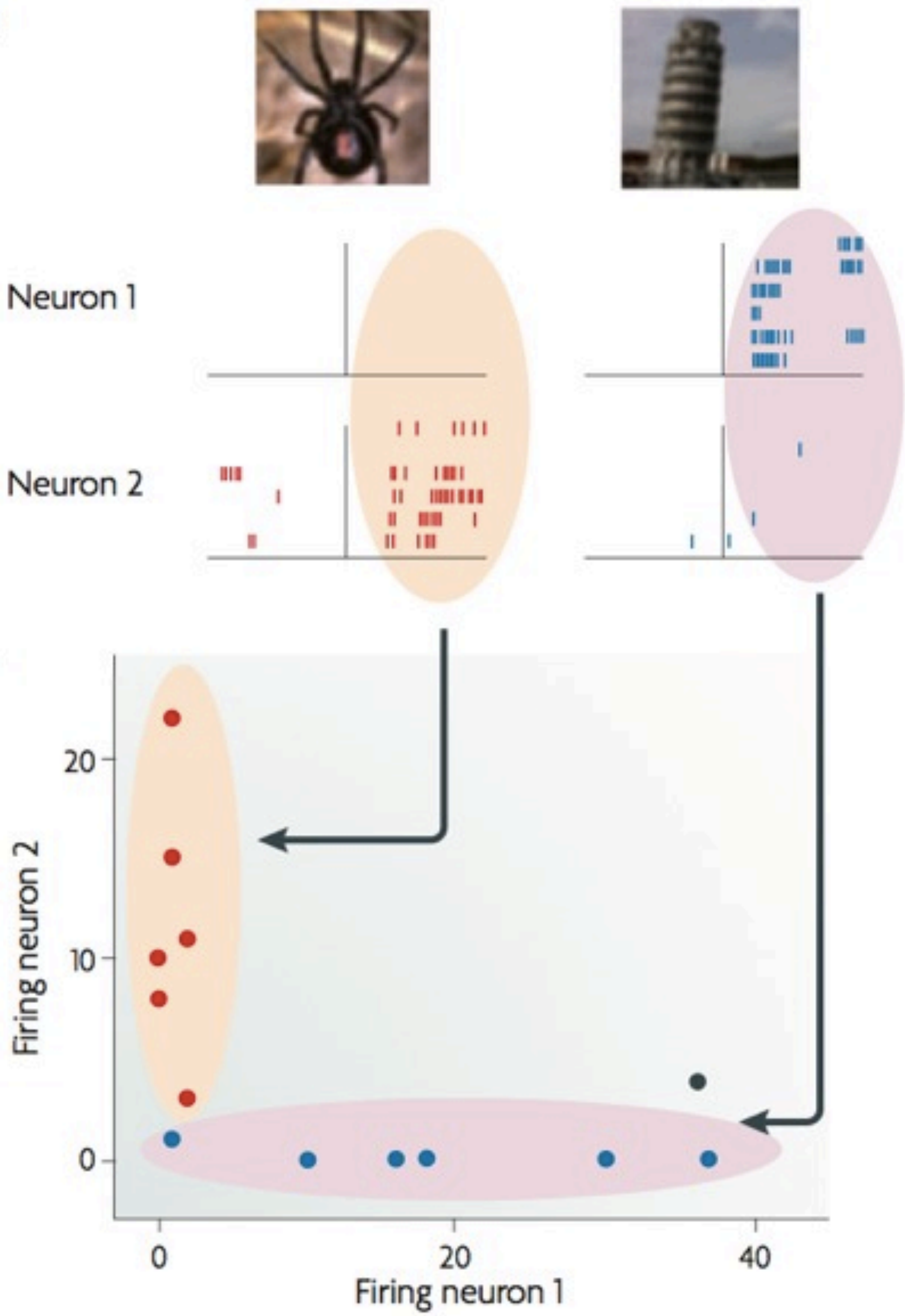
- Supervised
- Unsupervised
- Reinforcement



Supervised learning task

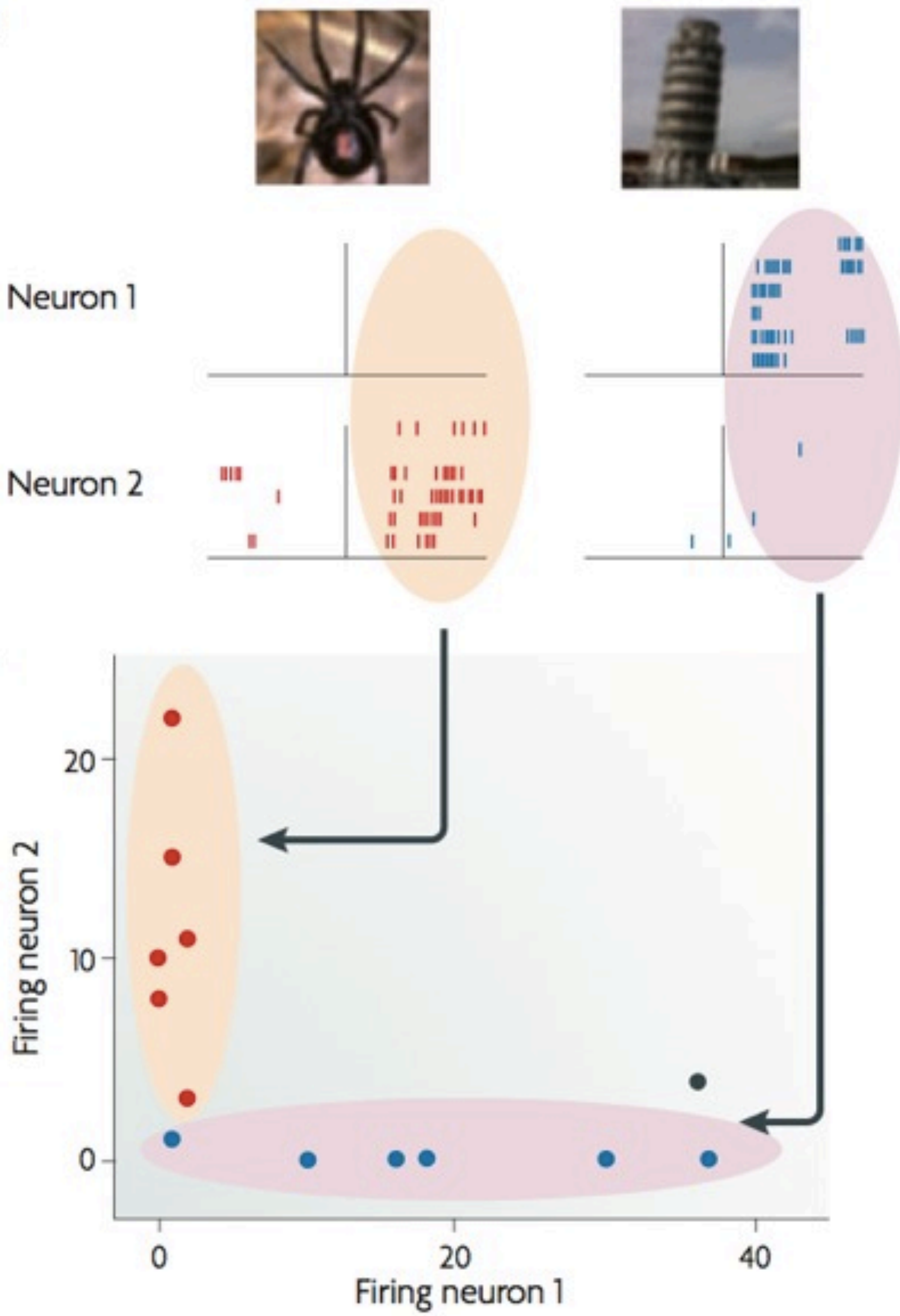


Supervised learning task



Prediction: given the firing rate of the two neurons...which class was the stimulus?

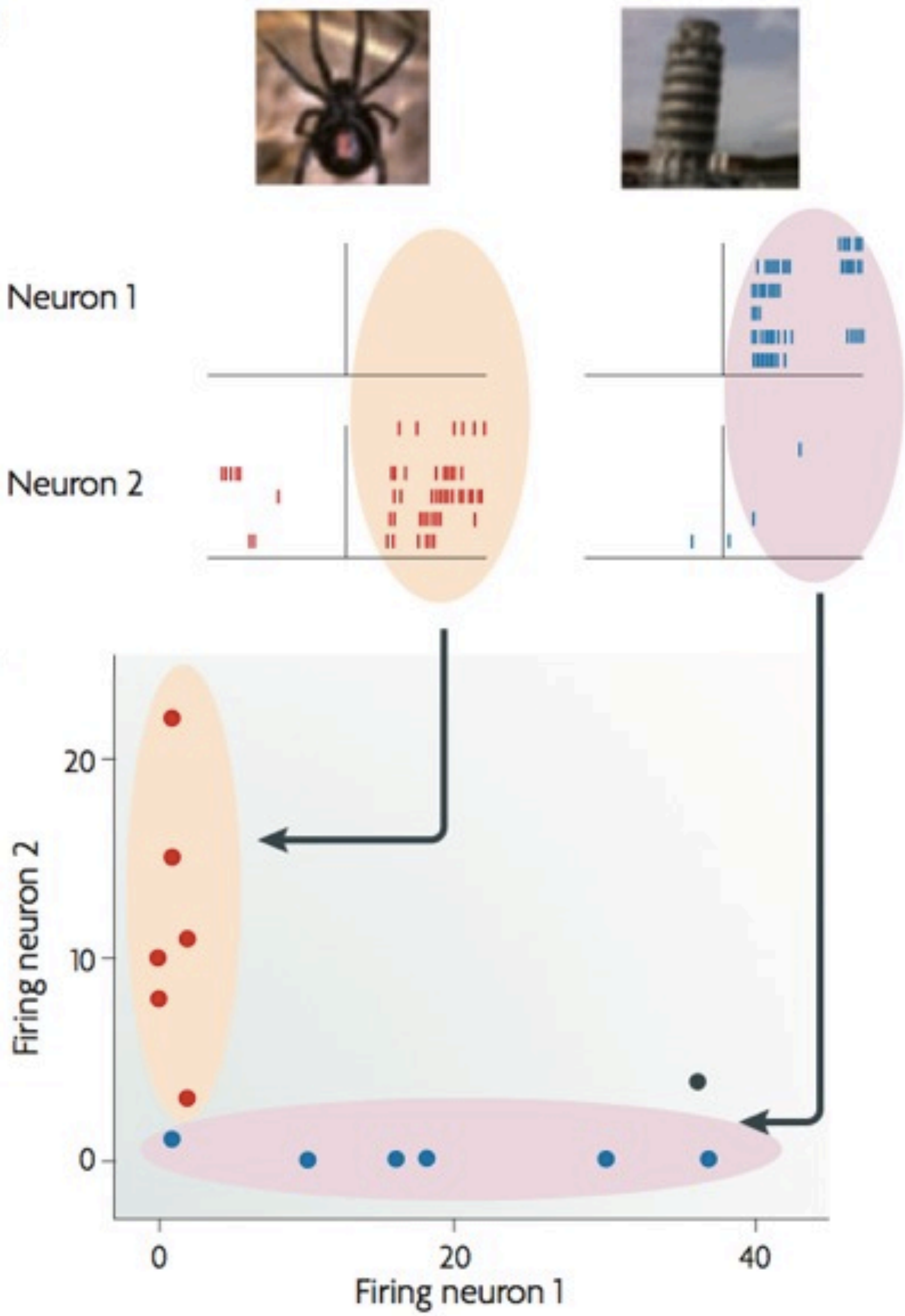
Supervised learning task



Prediction: given the firing rate of the two neurons...which class was the stimulus?

Features: $\{\mathbf{r_1}, \mathbf{r_2}\}$

Supervised learning task

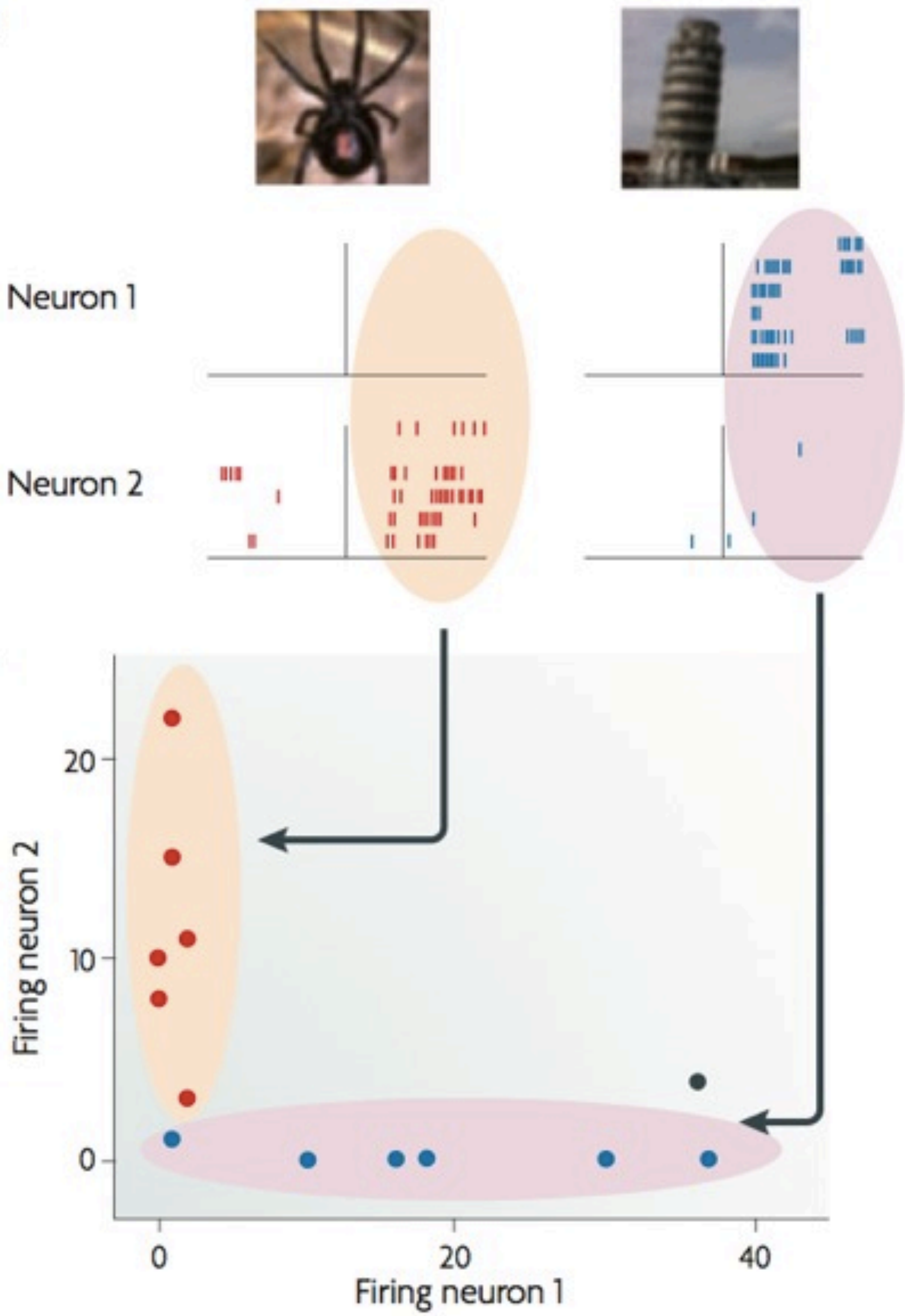


Prediction: given the firing rate of the two neurons...which class was the stimulus?

Features: $\{\mathbf{r_1}, \mathbf{r_2}\}$

Labels: {**0**,**1**} (**animal**,**building**)

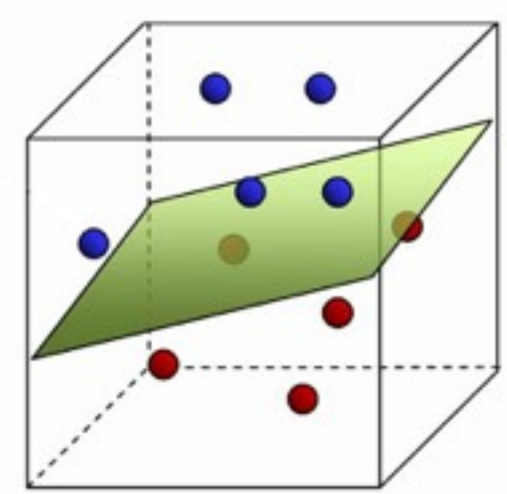
Supervised learning task



Prediction: given the firing rate of the two neurons...which class was the stimulus?

Features: $\{\mathbf{r_1}, \mathbf{r_2}\}$

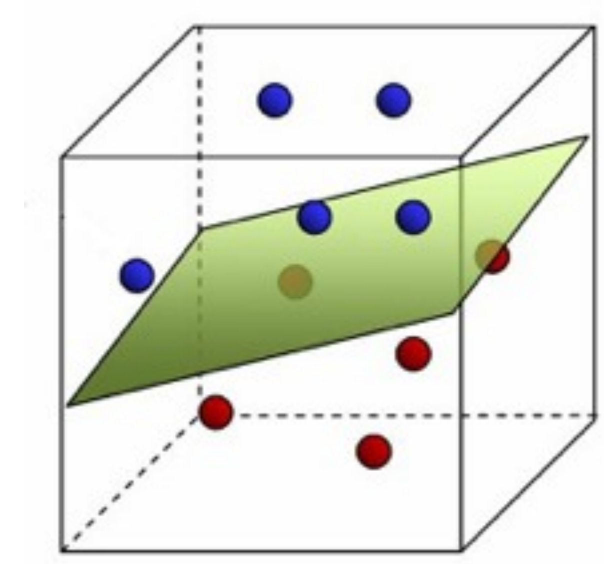
Labels: {**0**,**1**} (**animal**,**building**)



Idea: train a classifier to discriminate between different classes of stimuli (or decisions) and used to predict novel examples

Basic pipeline

1. Defining features and classes
2. Feature selection
3. Choosing a classifier
4. Training and testing a classifier
5. Examining results



1) Feature and classes

Features: could be firing rates in intracraneal recordings, power of oscillations in EEG, voxel activity in fMRI,...

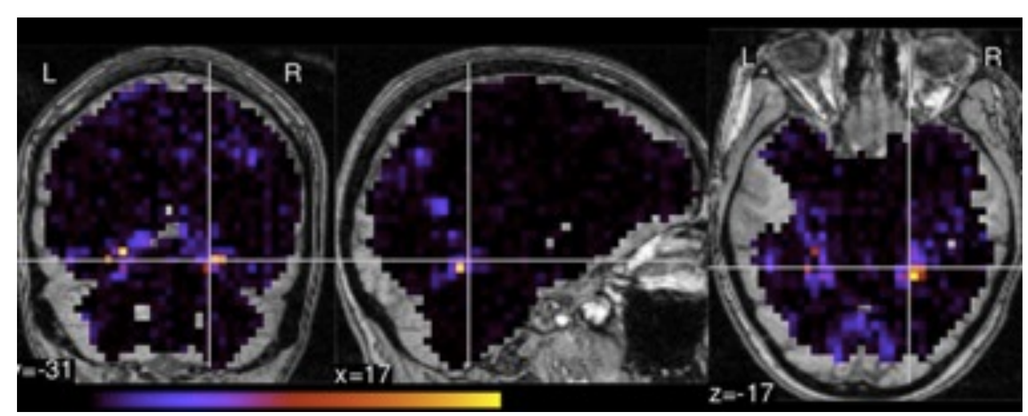
$$\mathbf{x} = (x_1, x_2, ..., x_v)$$

Classes: can type of stimuli, movement, disease,...

E.g. Vision:

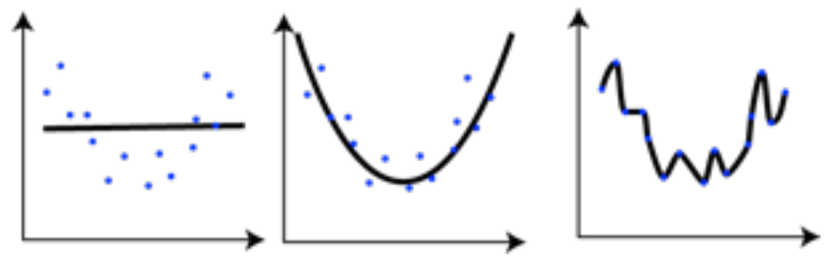
features: voxels in visual region

classes: object shown



2) Feature selection

Too many features: overfitting & curse of dimensionality



Solution: select only the most informative features (2nd part of the talk)

#features < 20 x #samples

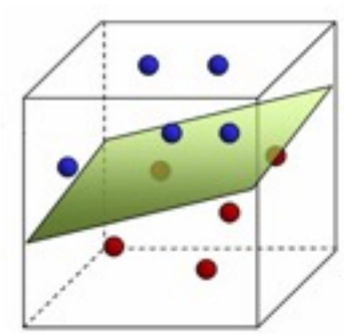
E.g. Vision:

100000 voxels: select only those modulated by the stimuli

3) Choose a classifier

Classifier: a function f(.) that takes the values of the observed features (ex., voxels) and predicts to which class *y* the observation belongs

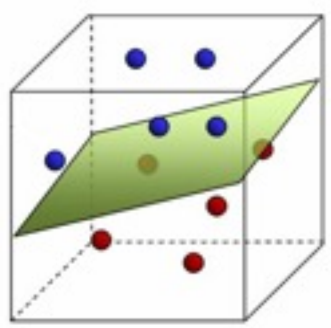
$$y = f(\mathbf{x})$$



3) Choose a classifier

Classifier: a function f(.) that takes the values of the observed features (ex., voxels) and predicts to which class *y* the observation belongs

$$y = f(\mathbf{x}, u)$$



- Decision trees
- C4.5
- Random forests
- Bayesian networks
- Hidden Markov models
- Artificial neural network
- Data clustering
- Expectation-maximization algorithm
- Self-organizing map
- Radial basis function network
- Vector Quantization
- Generative topographic map
- Information bottleneck method
- IBSEAD
- Apriori algorithm
- Eclat algorithm
- FP-growth algorithm
- Single-linkage clustering
- Conceptual clustering
- K-means algorithm
- Fuzzy clustering
- Temporal difference learning
- Q-learning
- Learning Automata

- AODE
- Artificial neural network
- Backpropagation
- Naive Bayes classifier
- Bayesian network
- Bayesian knowledge base
- Case-based reasoning
- Decision trees
- Inductive logic programming
- Gaussian process regression
- Gene expression programming
- Group method of data handling (GMDH)
- Learning Automata
- Learning Vector Quantization
- Logistic Model Tree
- Decision trees
- Decision graphs
- Lazy learning
- Monte Carlo Method
- SARSA

- Instance-based learning
- Nearest Neighbor Algorithm
- Analogical modeling
- Probably approximately correct learning (PAC
- Symbolic machine learning algorithms
- Subsymbolic machine learning algorithms
- Support vector machines
- Random Forests
- Ensembles of classifiers
- Bootstrap aggregating (bagging)
- Boosting (meta-algorithm)
- Ordinal classification
- Regression analysis
- Information fuzzy networks (IFN)
- Linear classifiers
- Fisher's linear discriminant
- Logistic regression
- Naive Bayes classifier
- Perceptron
- Support vector machines
- Quadratic classifiers
- k-nearest neighbor
- Boosting

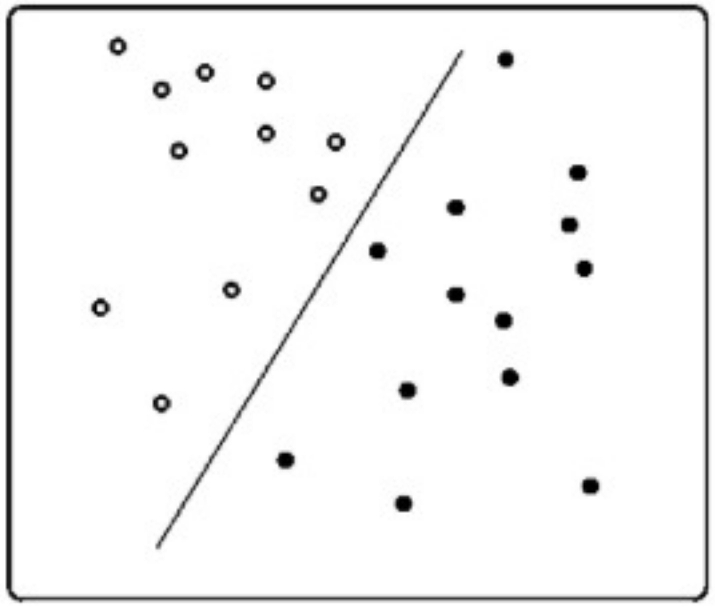
Linear:

Naive Bayes (NB)

Support Vector Machines (SVM)

Logistic Regression (LR)

Linear Discrimination Analysis (LDA)

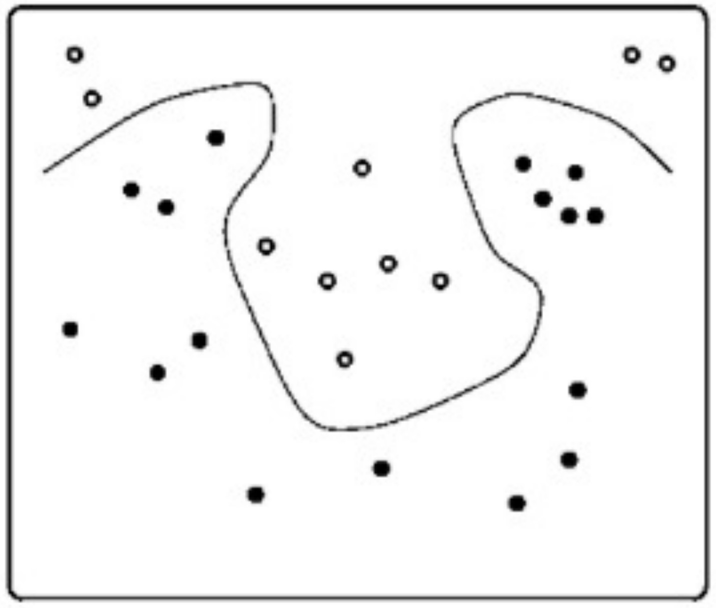


Non-linear:

Kernel SVM

Random Forests

Artificial Neural Networks (ANN)



4) Training & testing

Training data

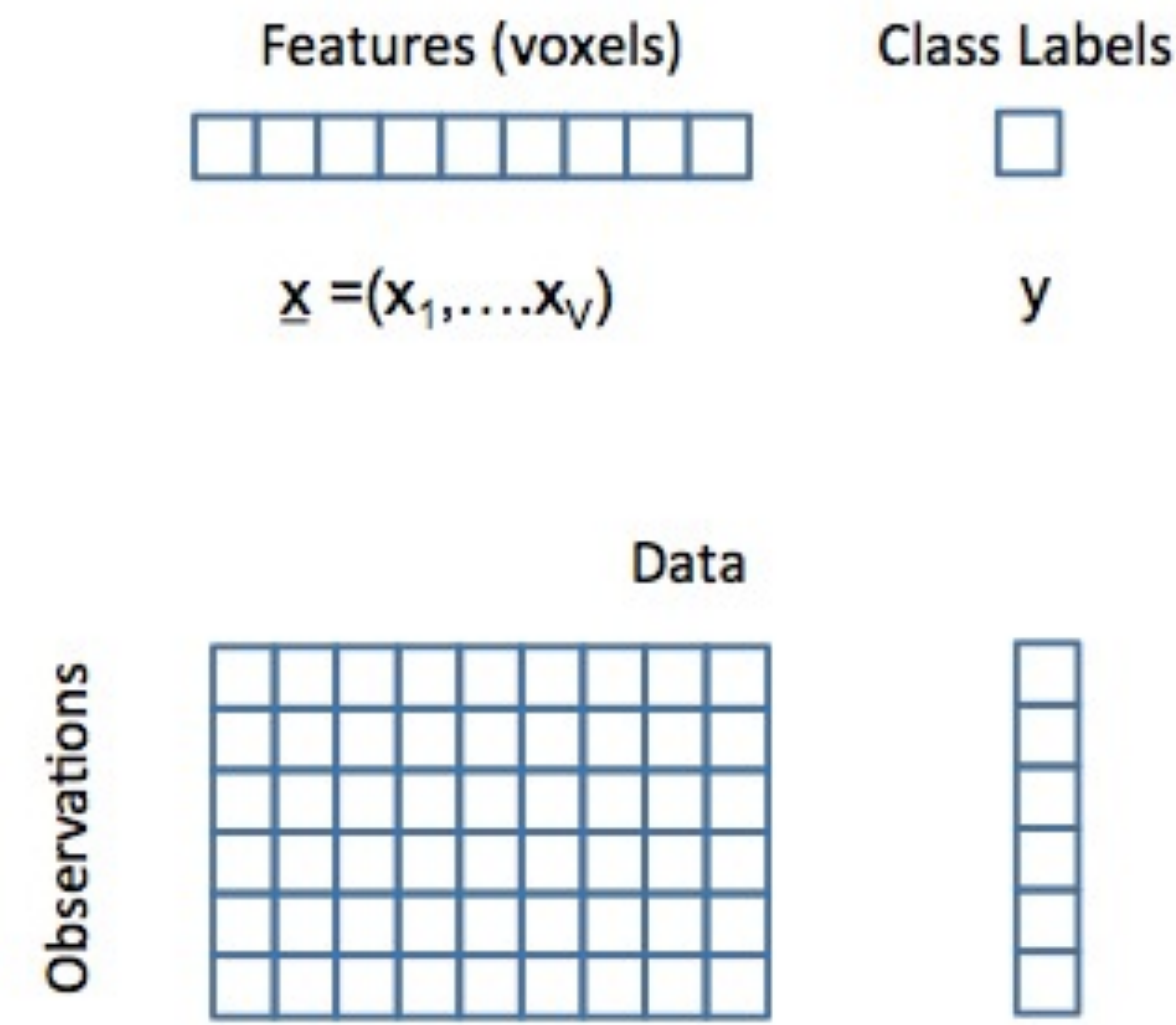
A classifier has a number of parameters that have to be learned

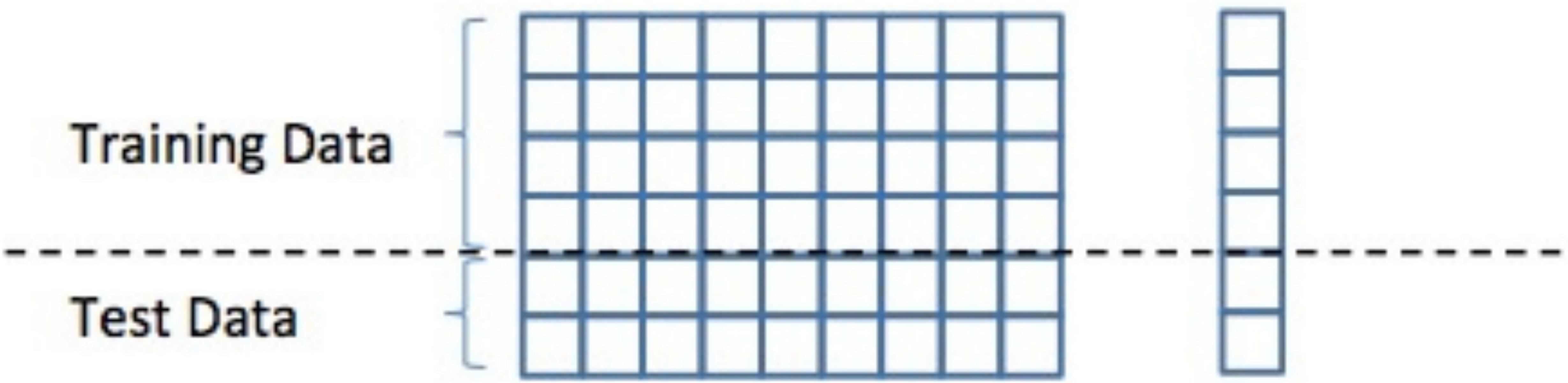
A learned classifiers models the relation between features and class labels in the training data set

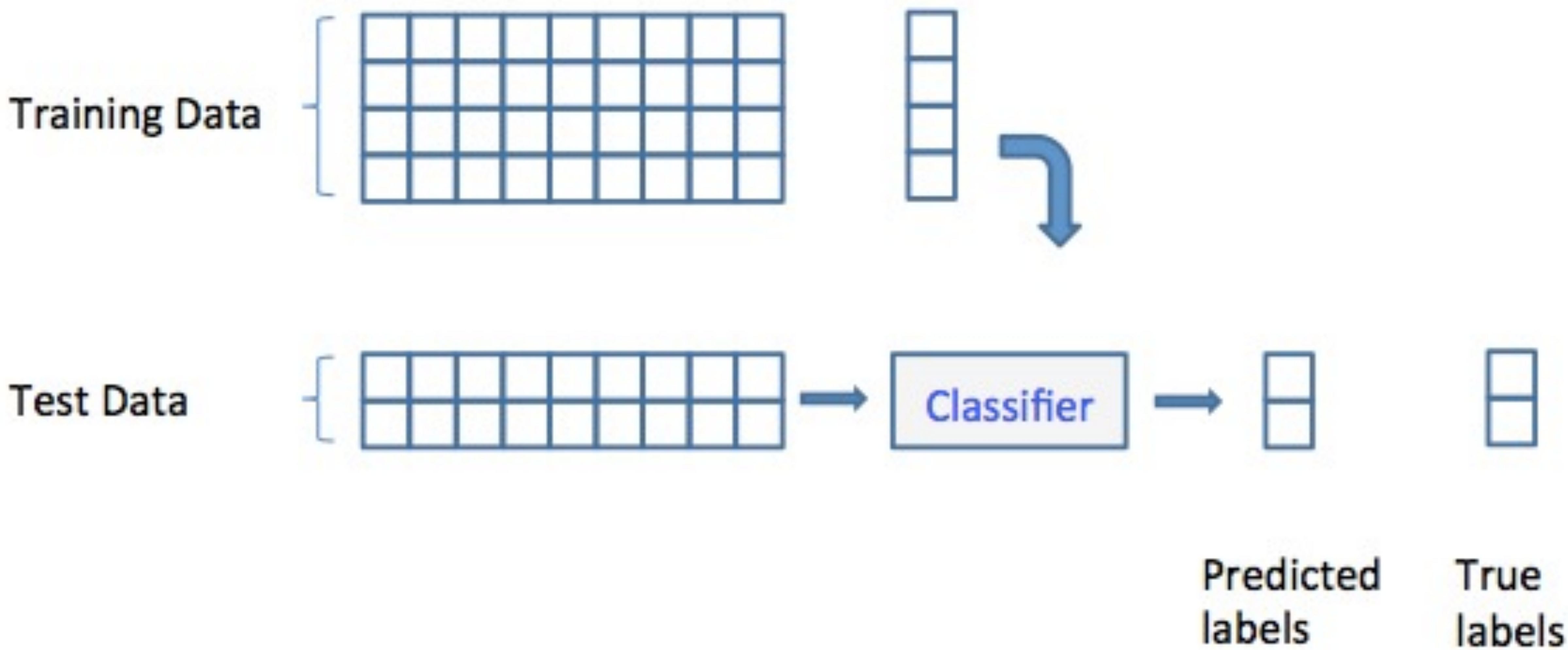
Test data

If the classifiers truly captures the relation between features and labels, it should predict the class label for data it has not seen before

Once trained the classifier is evaluated using an independent set of observations (test data)

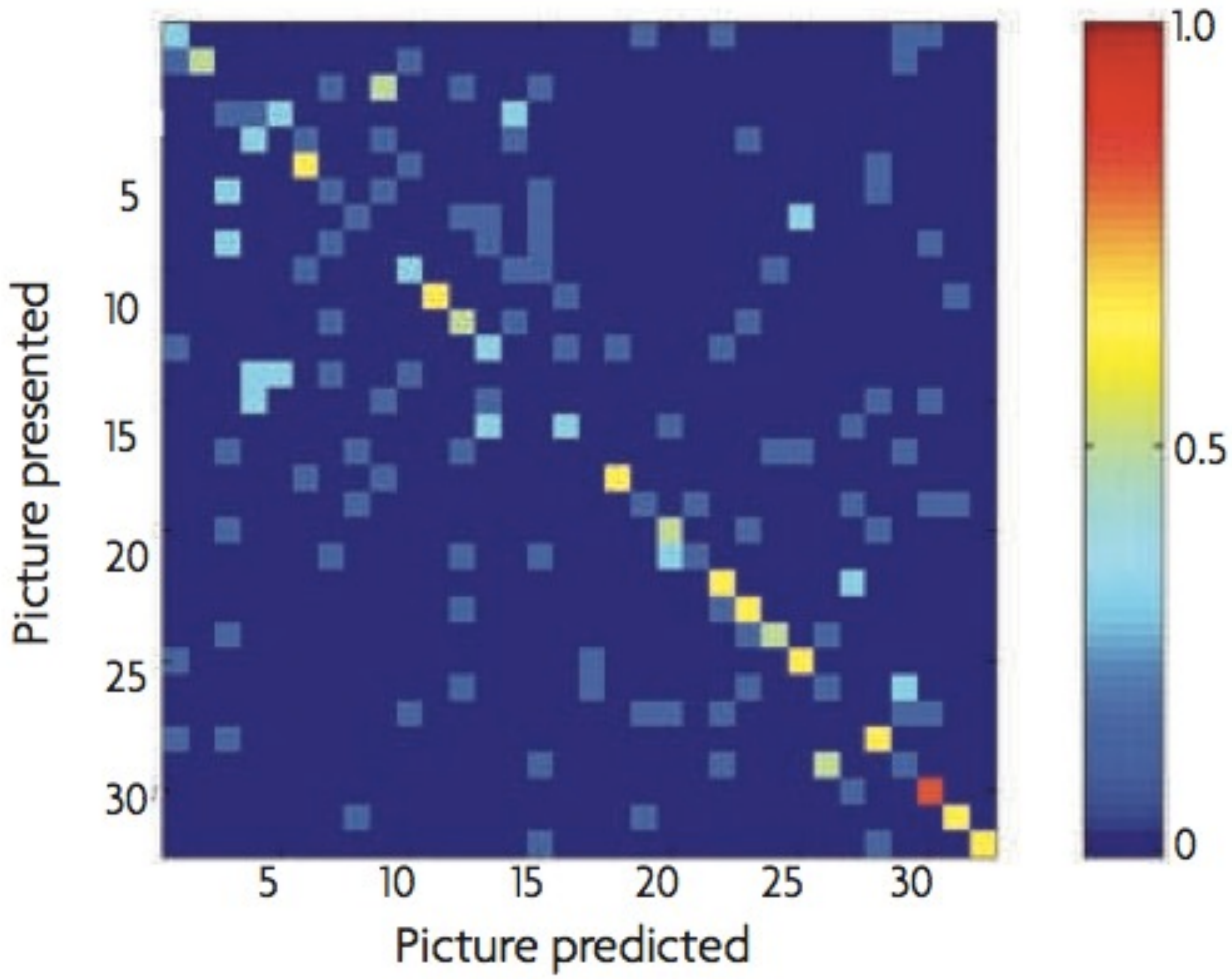

Features (voxels)
Class Labels
$\underline{x} = (x_1, \ldots x_V)$
y
Data
Observations


Training Data
Test Data


Training Data
Test Data
Classifier
Predicted labels
True labels

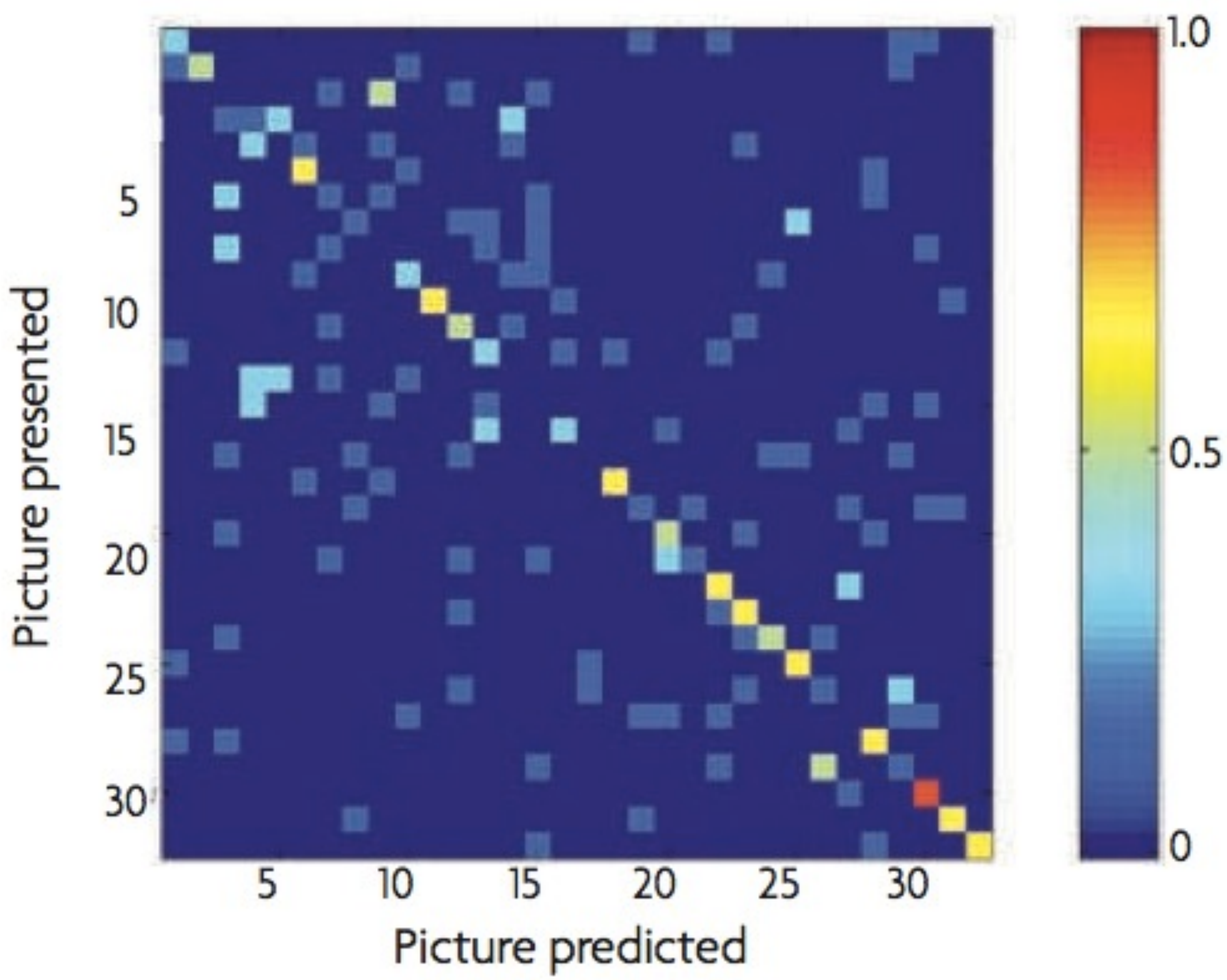
5) Examining results

Confusion matrix



5) Examining results

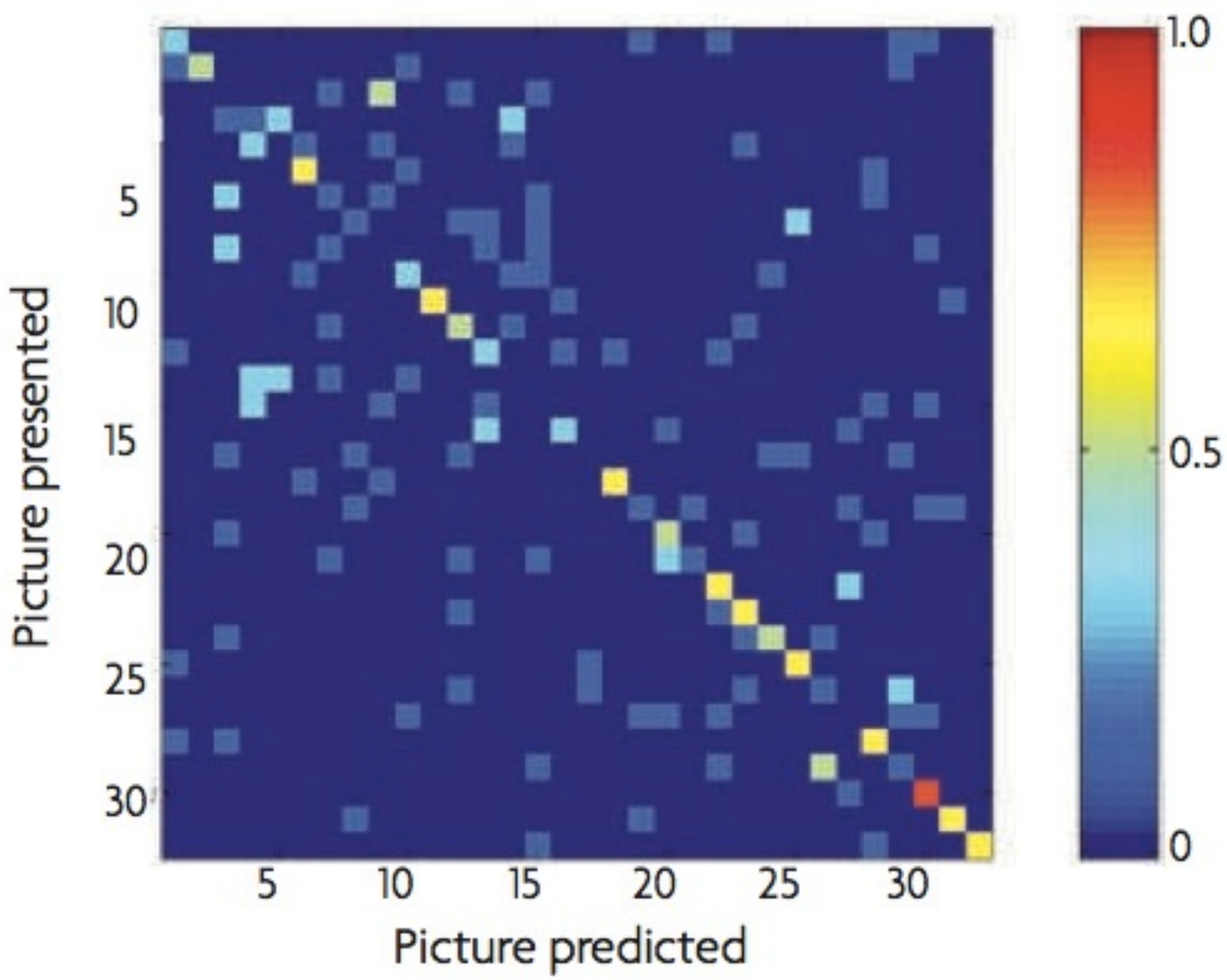
Confusion matrix

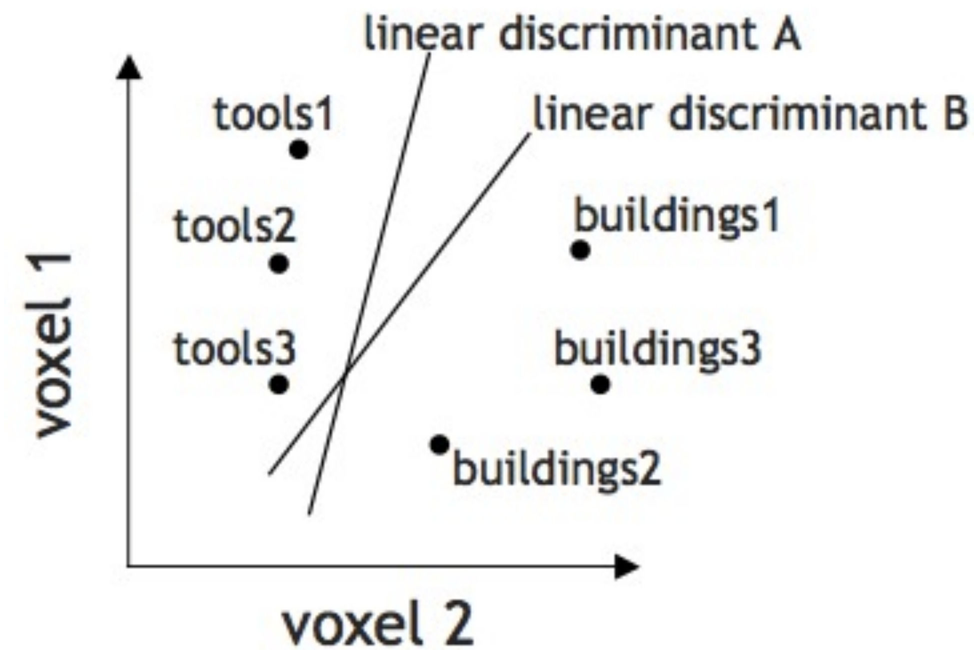


Precision
Recall
Accuracy
F-score

5) Examining results

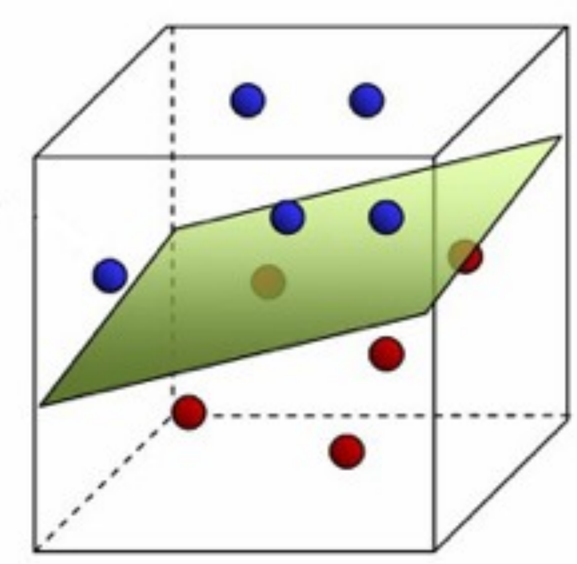
Confusion matrix

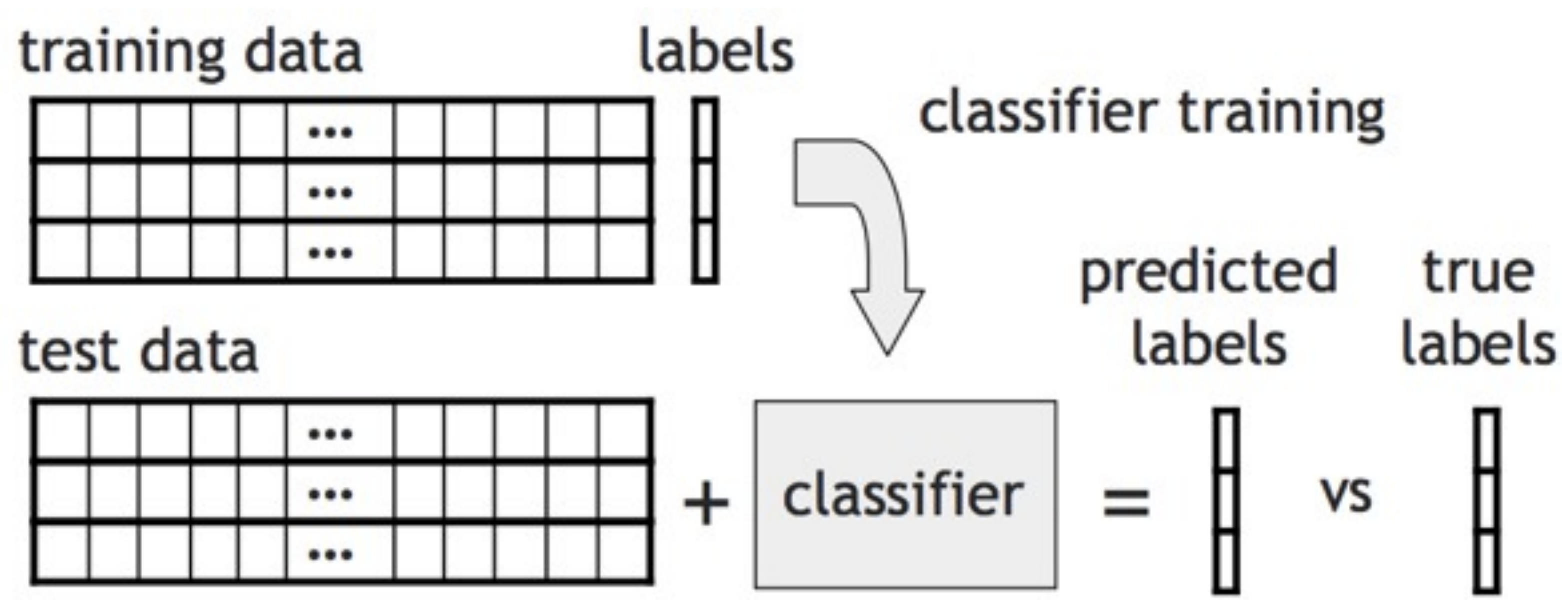




Basic pipeline

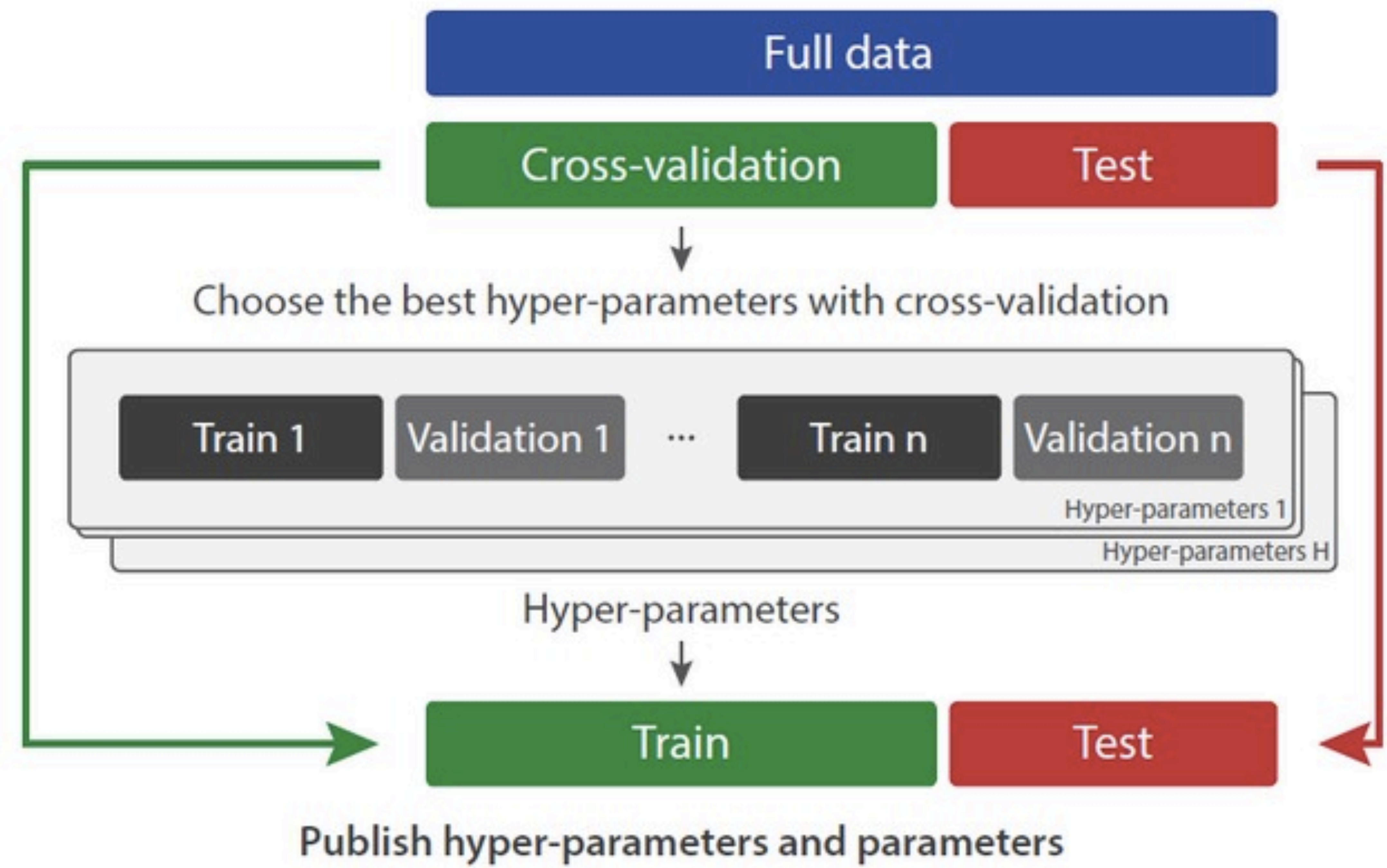
1. Defining features and classes
2. Feature selection
3. Choosing a classifier
4. Training and testing a classifier
5. Examining results

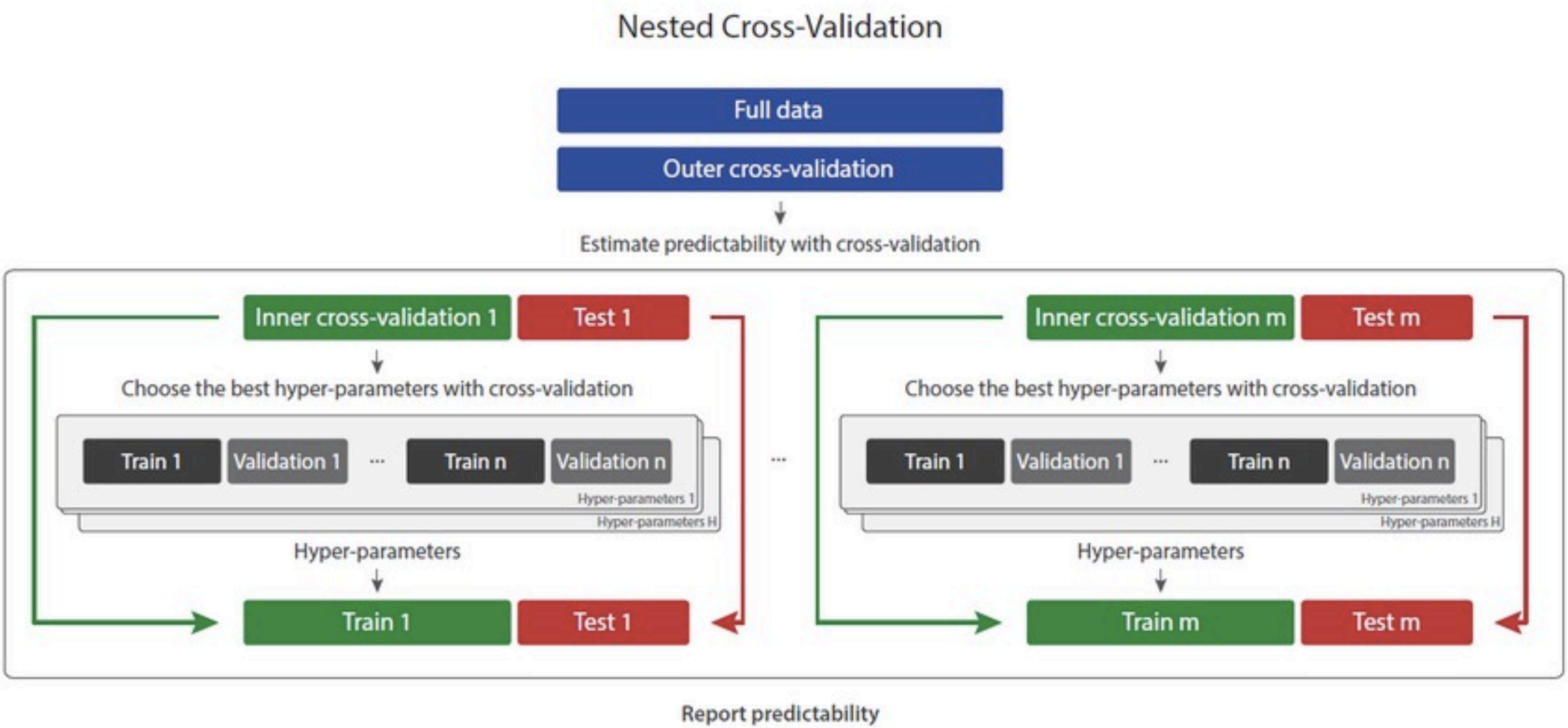




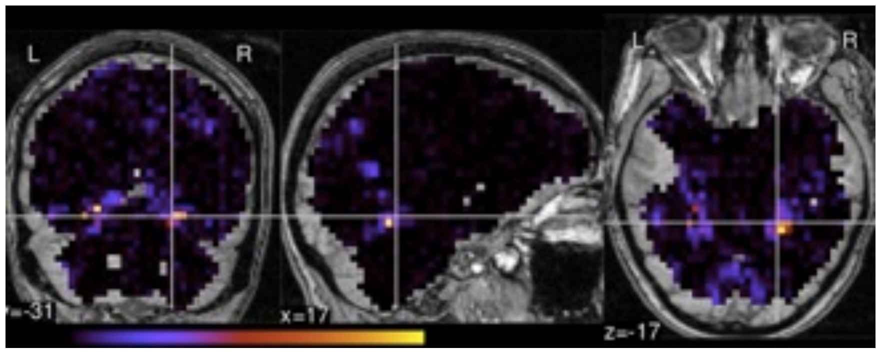
Data = 

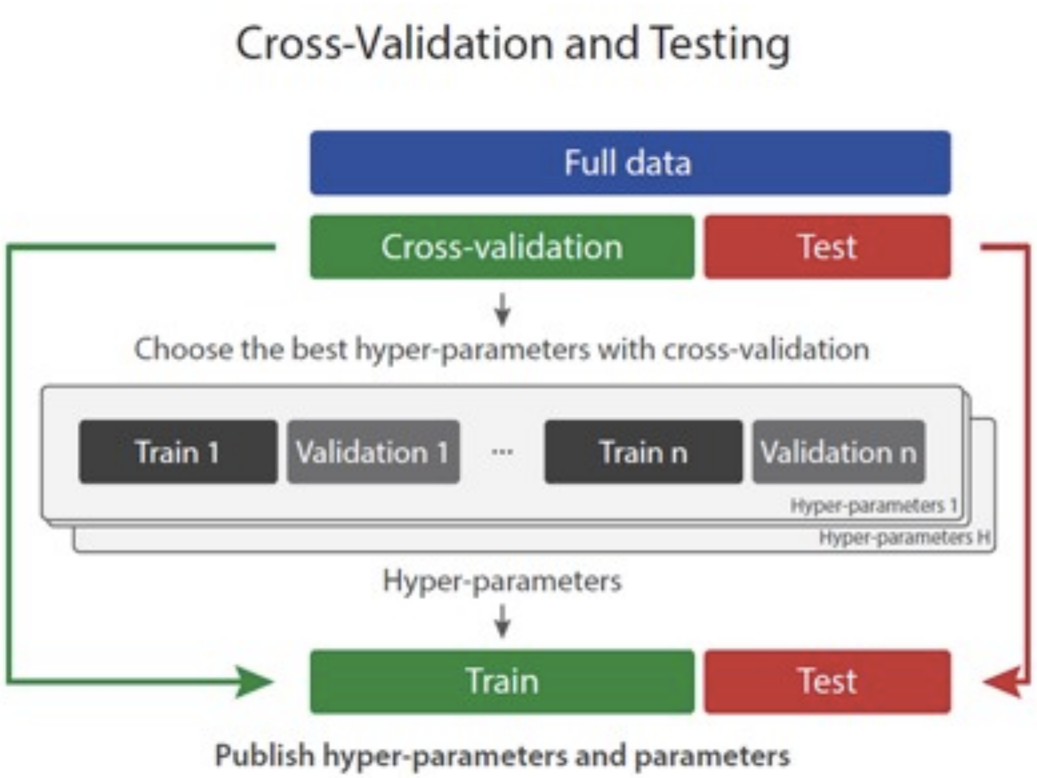
Cross-Validation and Testing

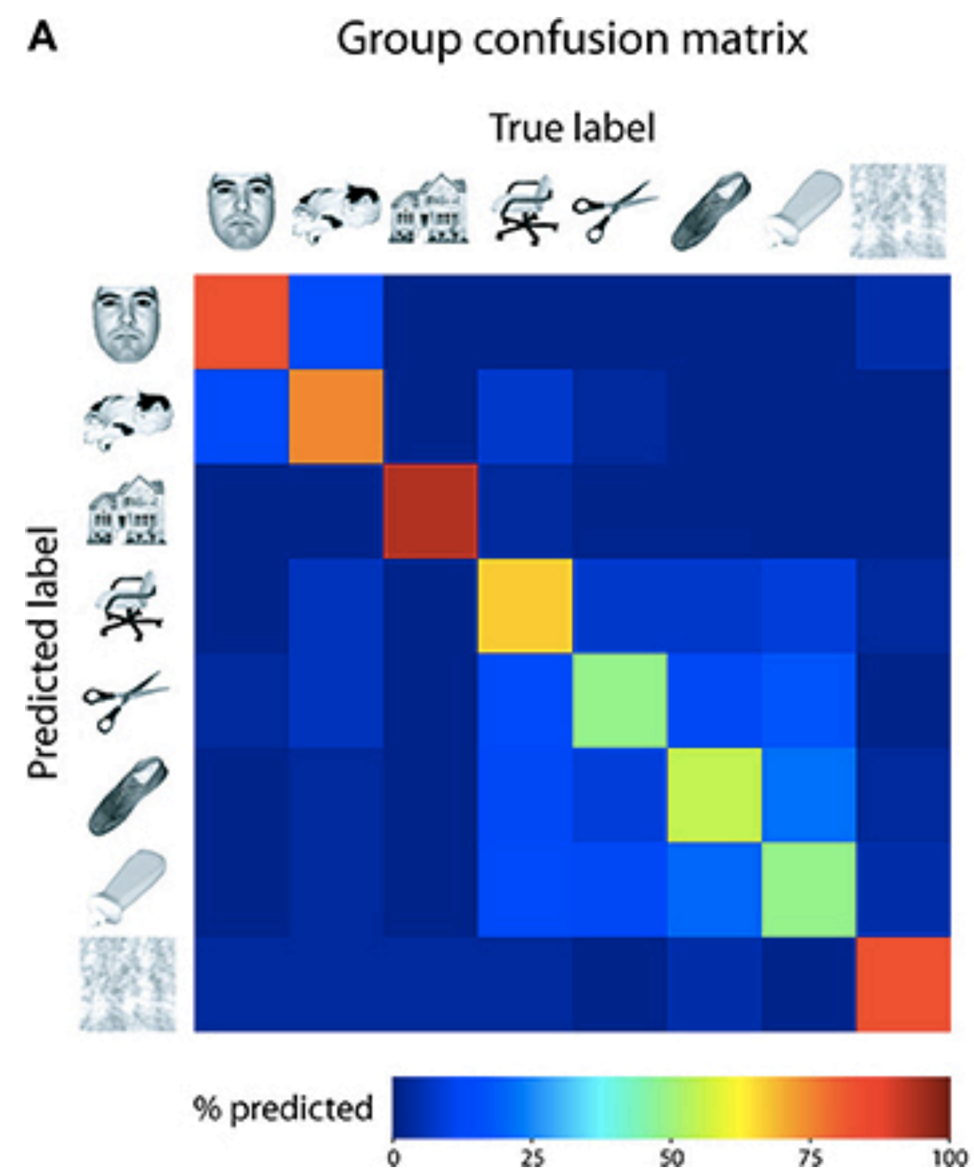



Nested Cross-Validation
Full data
Outer cross-validation
Estimate predictability with cross-validation
Inner cross-validation 1
Test 1
Choose the best hyper-parameters with cross-validation
Train 1
Validation 1
...
Train n
Validation n
Hyper-parameters 1
Hyper-parameters H
Hyper-parameters
Train 1
Test 1
...
Inner cross-validation m
Test m
Choose the best hyper-parameters with cross-validation
Train 1
Validation 1
...
Train n
Validation n
Hyper-parameters 1
Hyper-parameters H
Hyper-parameters
Train m
Test m
Report predictability

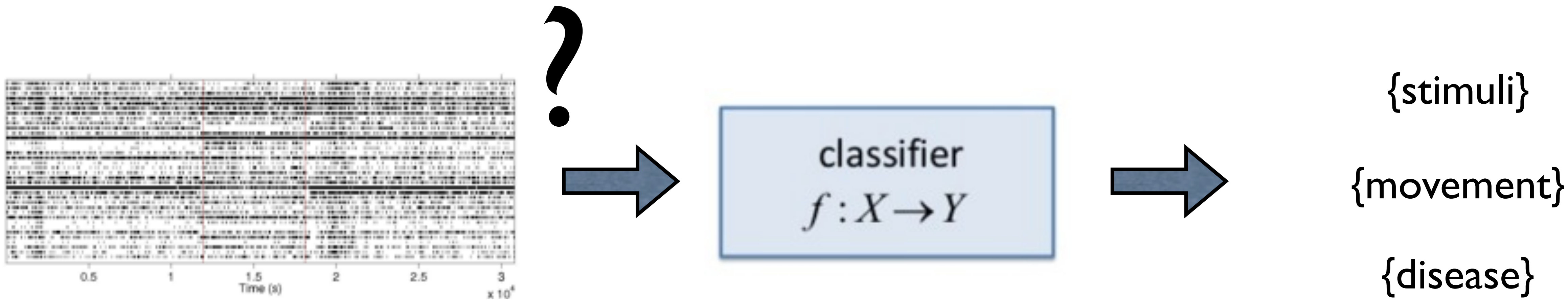
What now?


L
R
x=17
z=-17

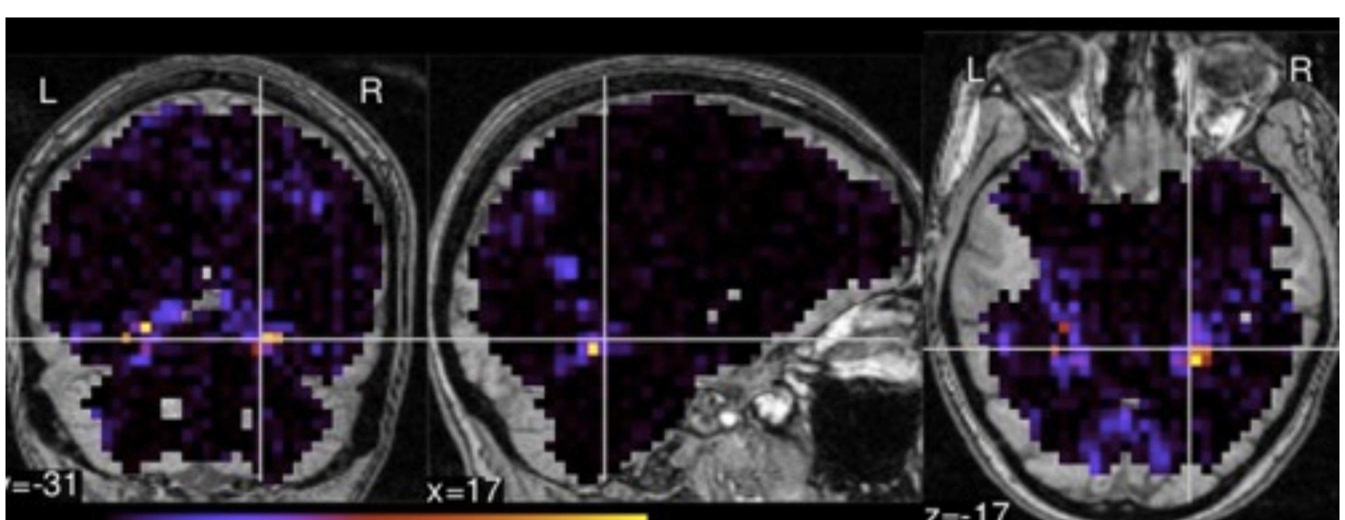

Cross-Validation and Testing
Full data
Cross-validation
Test
Choose the best hyper-parameters with cross-validation
Train 1
Validation 1
...
Train n
Validation n
Hyper-parameters 1
Hyper-parameters H
Hyper-parameters
Train
Test
Publish hyper-parameters and parameters


A
Group confusion matrix
True label
Predicted label
% predicted
0
25
50
75
100

Open the box & look into f!



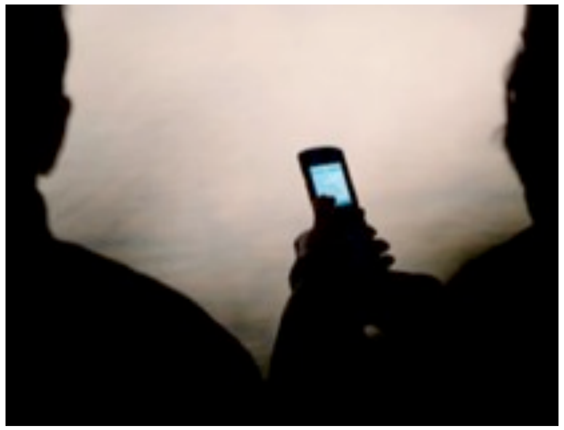
Neuroscience: what, where, when?



Open the box & look into f!

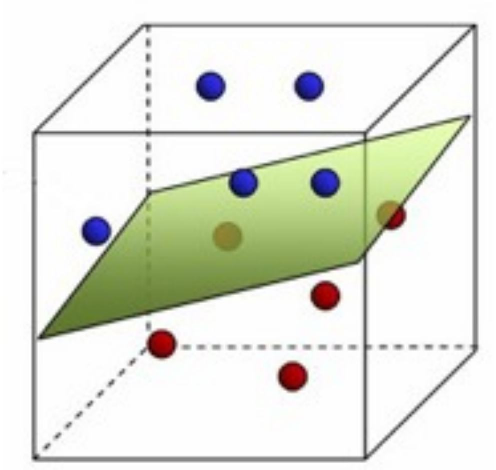
Law: government & banks need to use criteria for decision making that are human readable

Medicine: doctors speak a different language, they would like to know on what is any criteria based

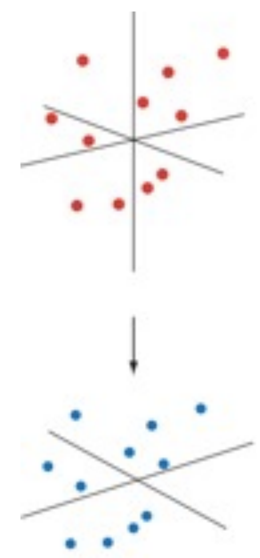


Parkinson tested from phone calls... >95% accuracy

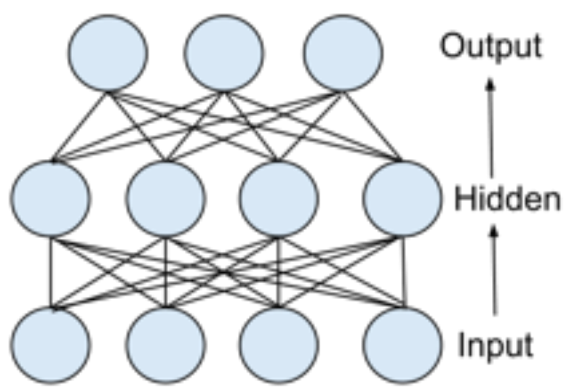
Machine learning



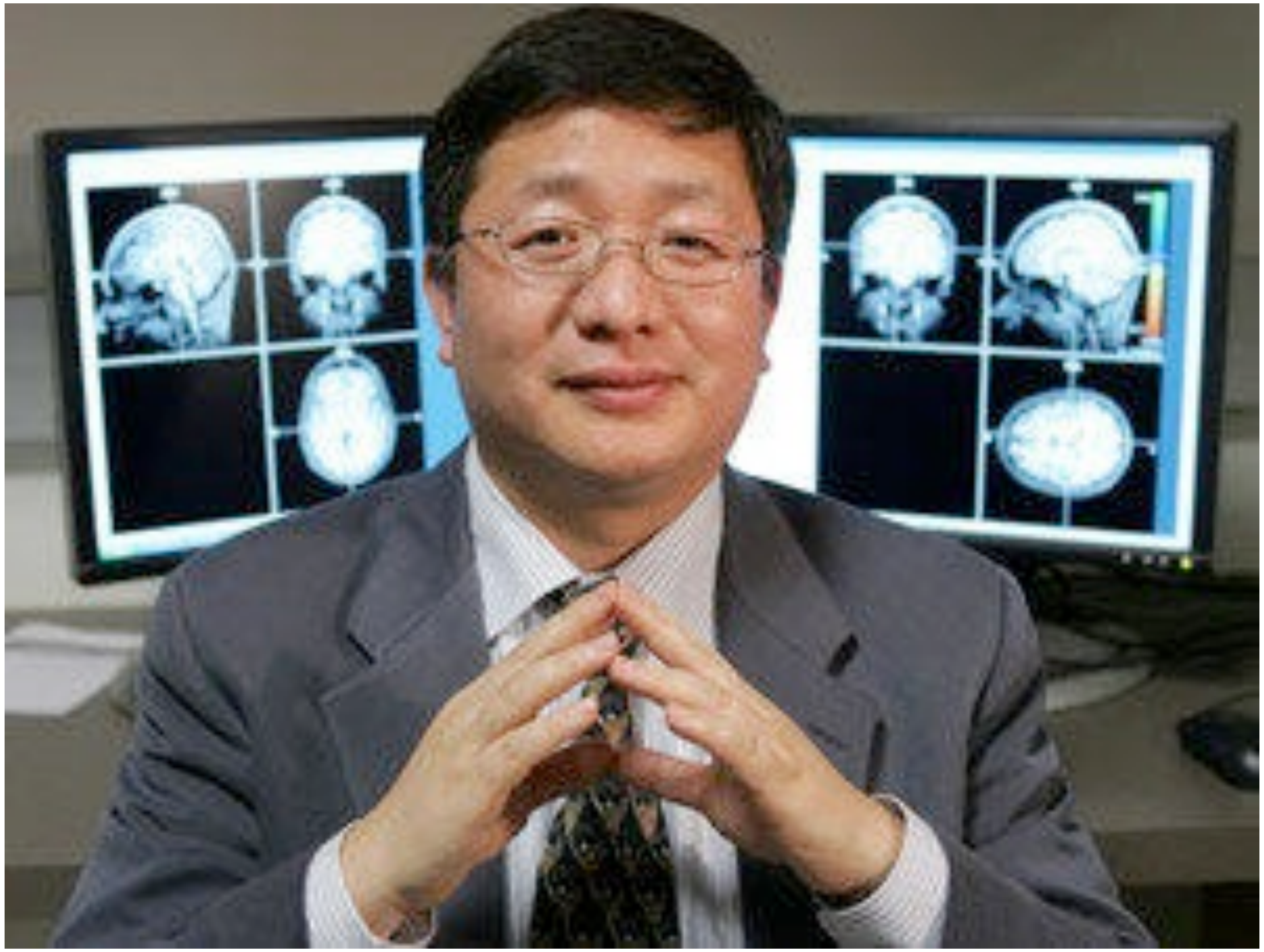
Dimensionality reduction



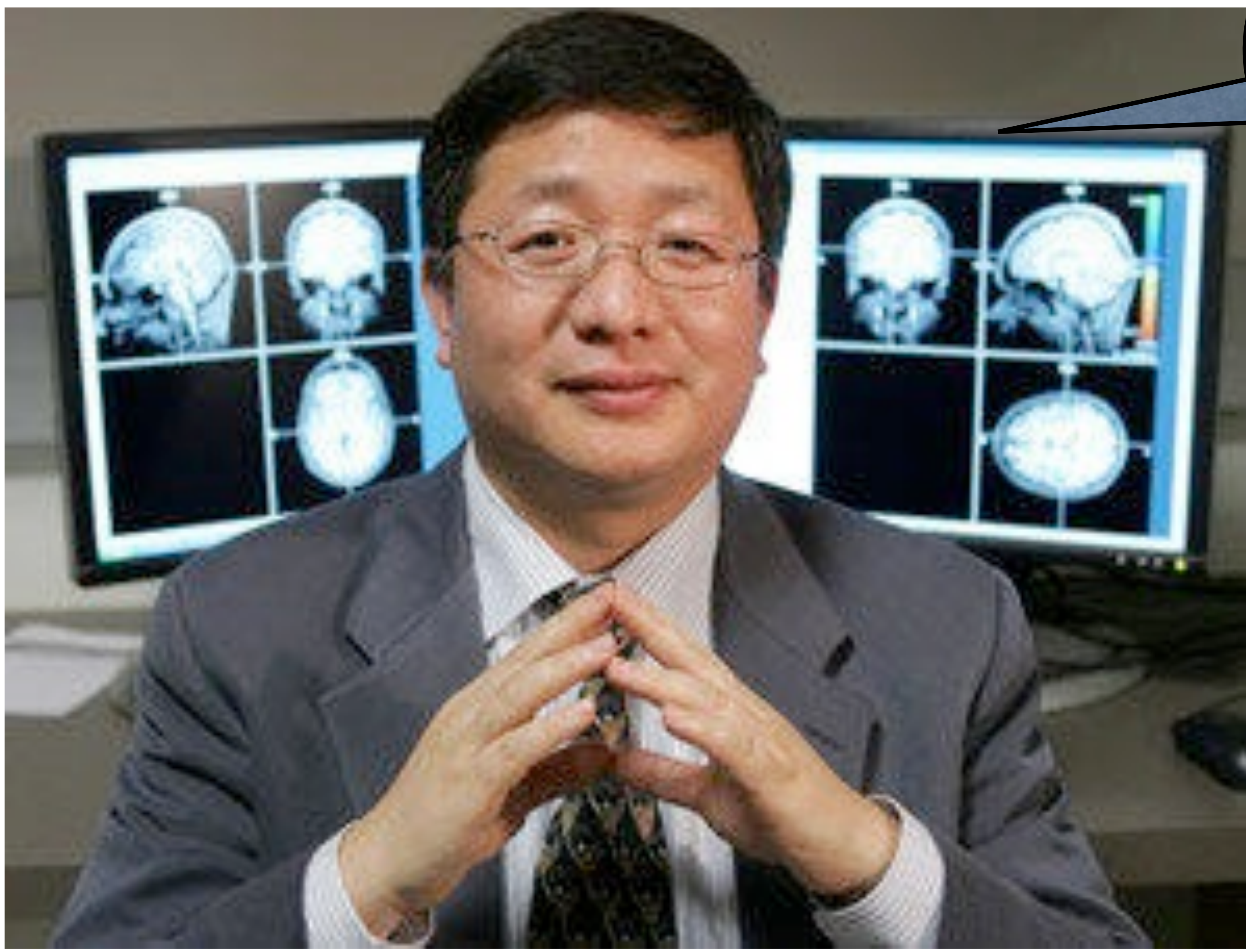
Caveats

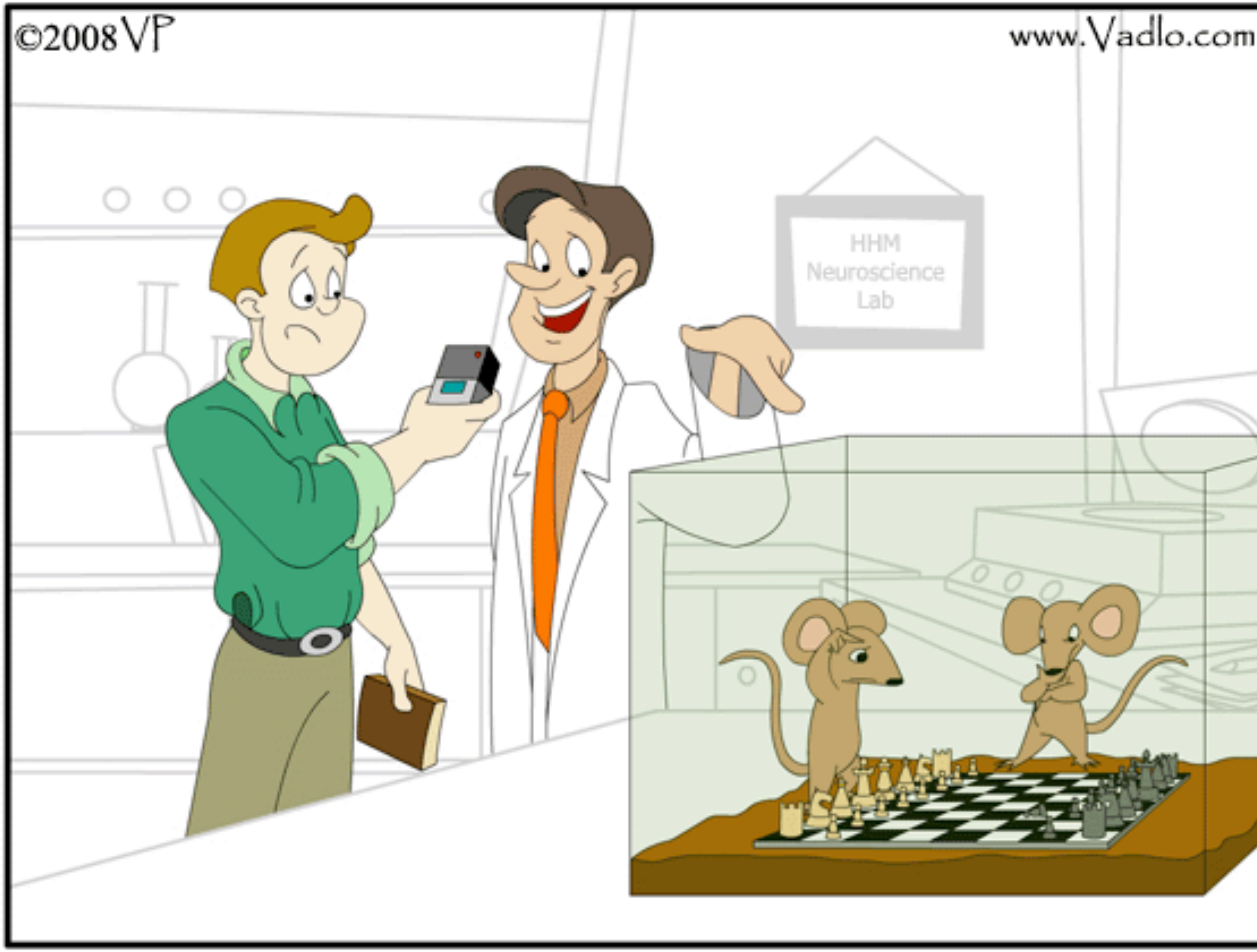


Dream of an experimentalist neuroscientist?



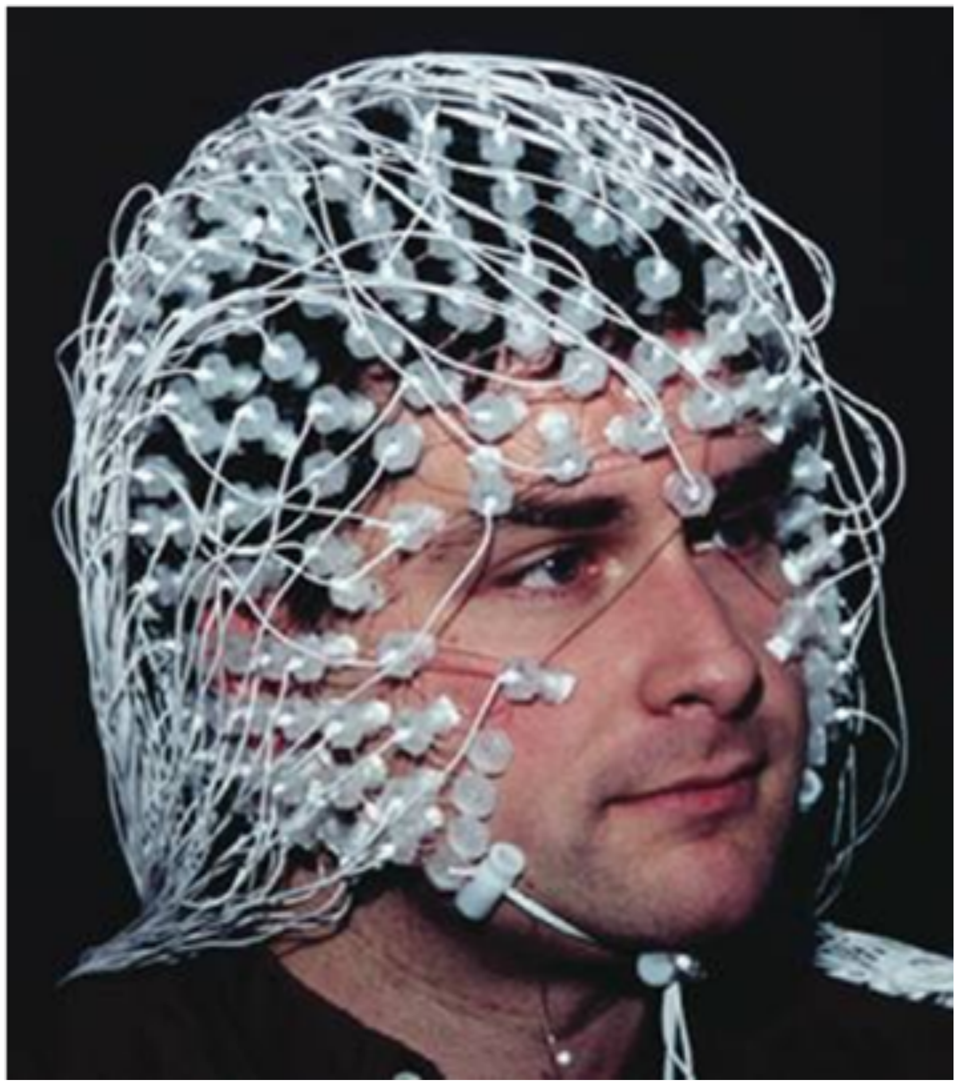
Dream of an experimentalist neuroscientist?





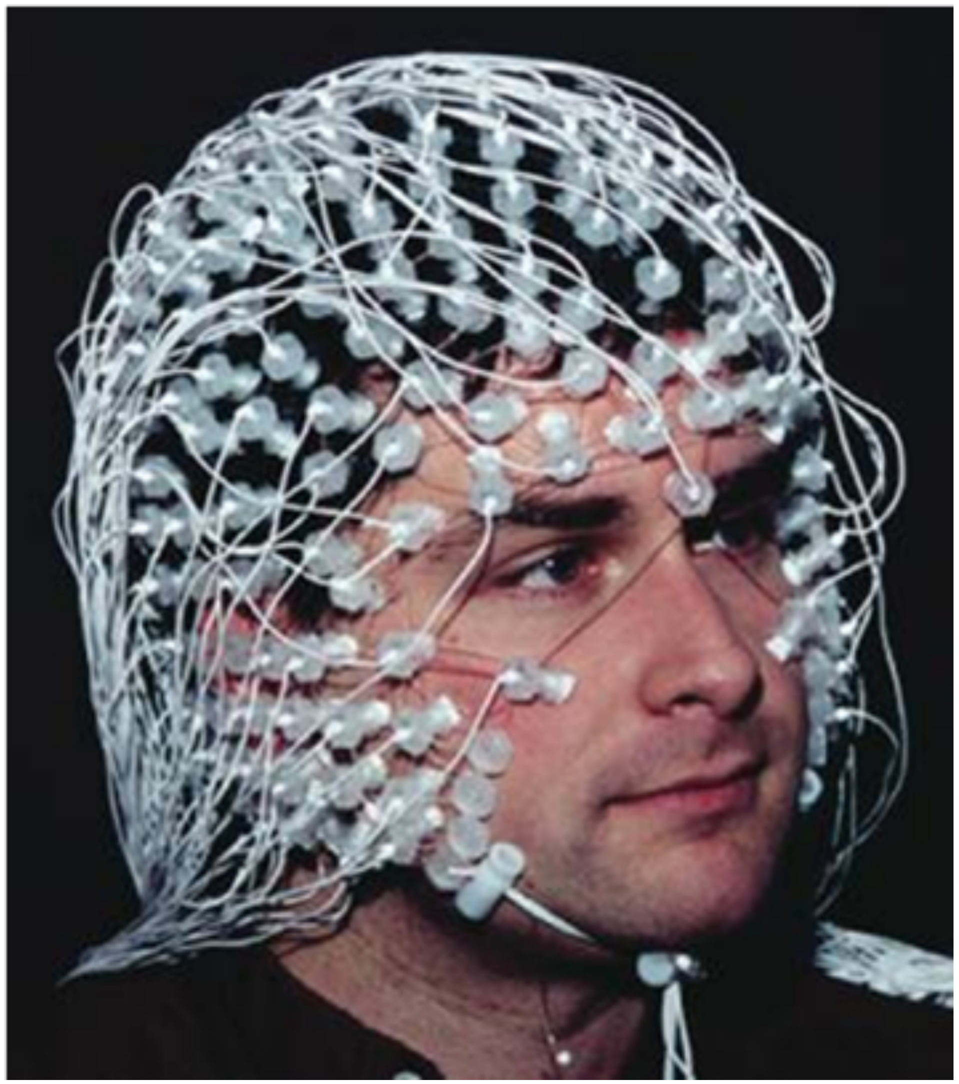
This is our director's pet project.

Slowly getting there...

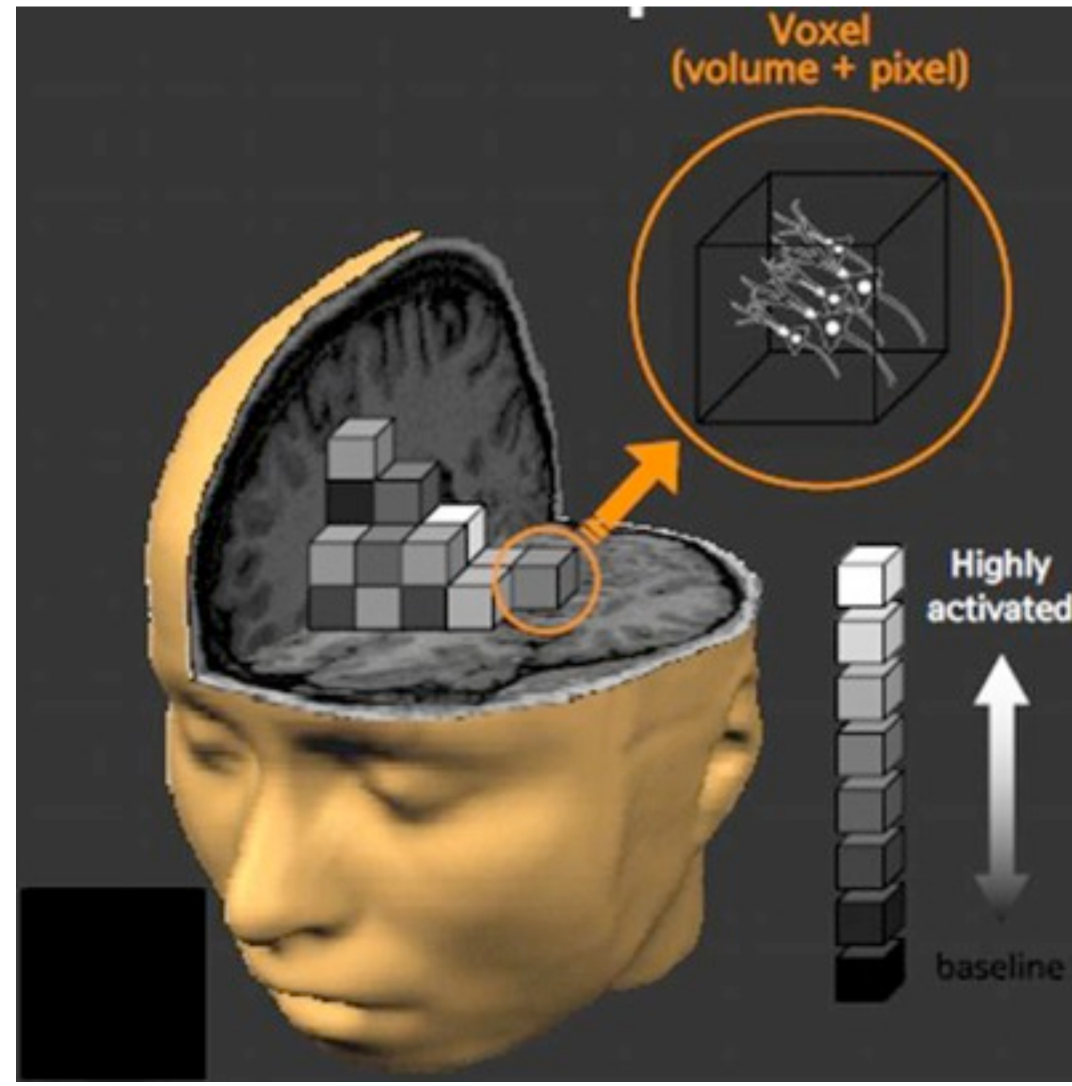


Dense array EEG (256 channels)

Slowly getting there...

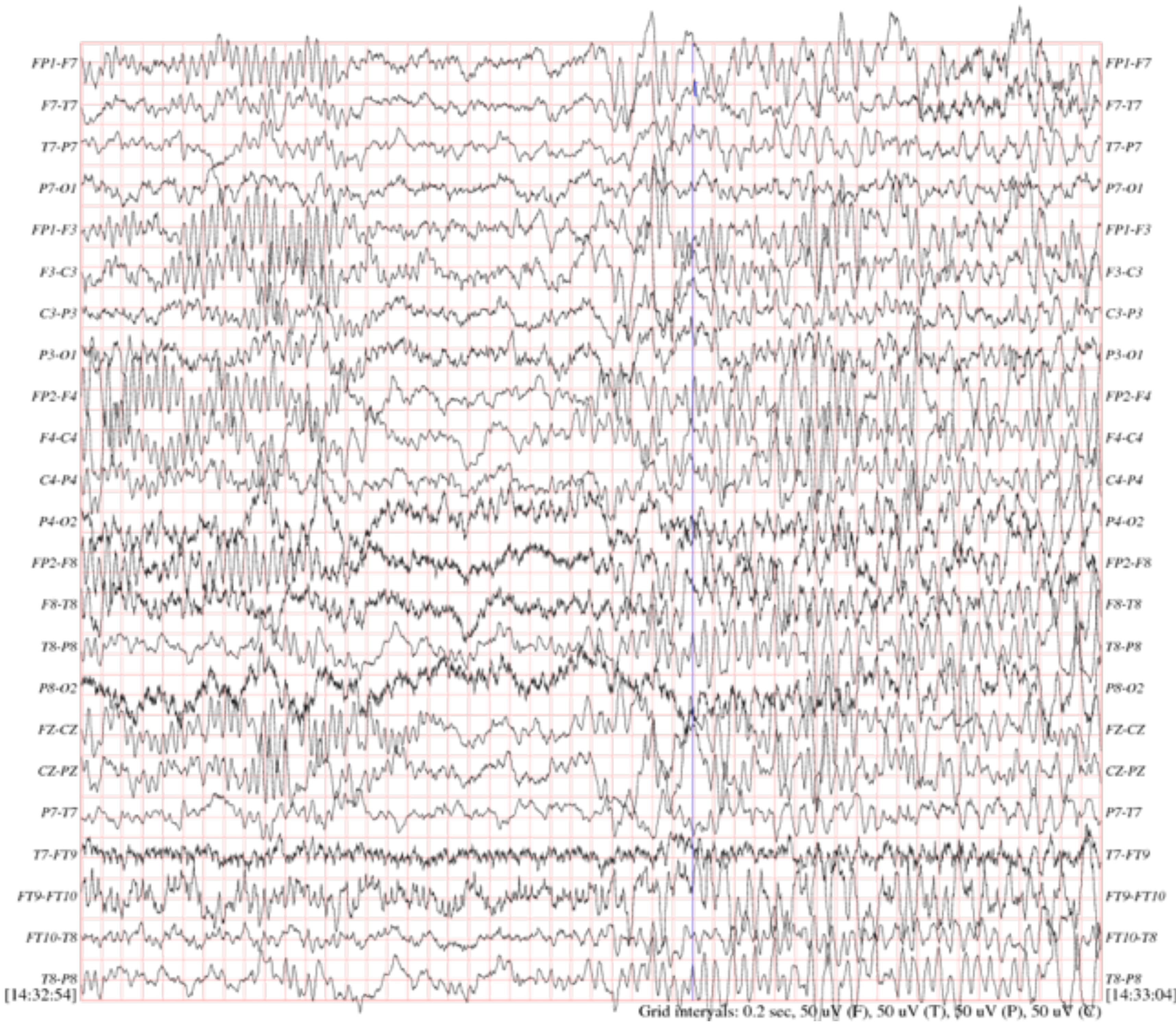


Dense array EEG (256 channels)

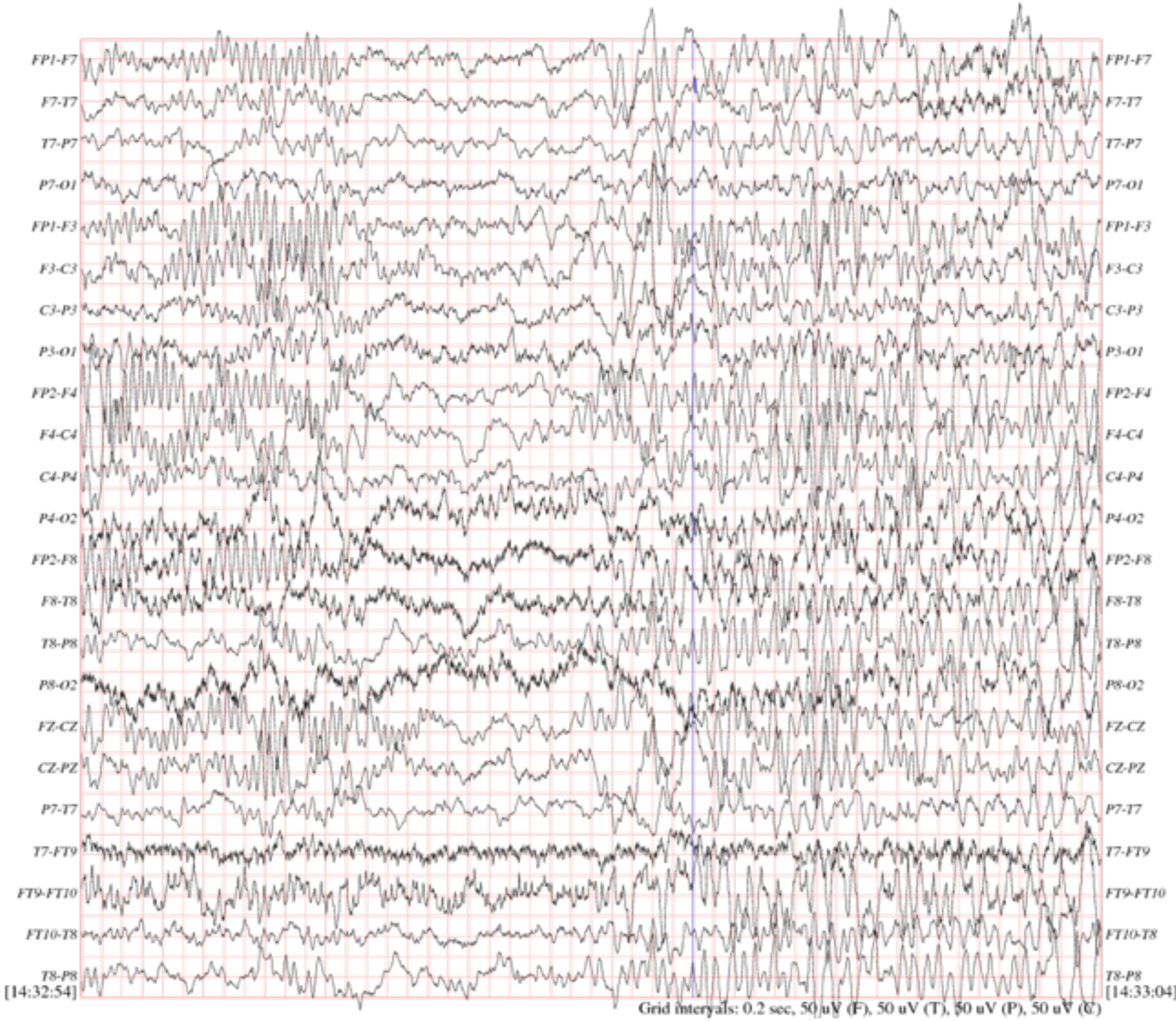


fMRI (130000 voxels)

Nightmare of the analyst



Nightmare of the analyst



How would YOU start exploring a high-dimensional data set?

When less is more

- What is dimensionality reduction?
- The zoo of dimensionality reduction
- Some applications

Dimension reduction

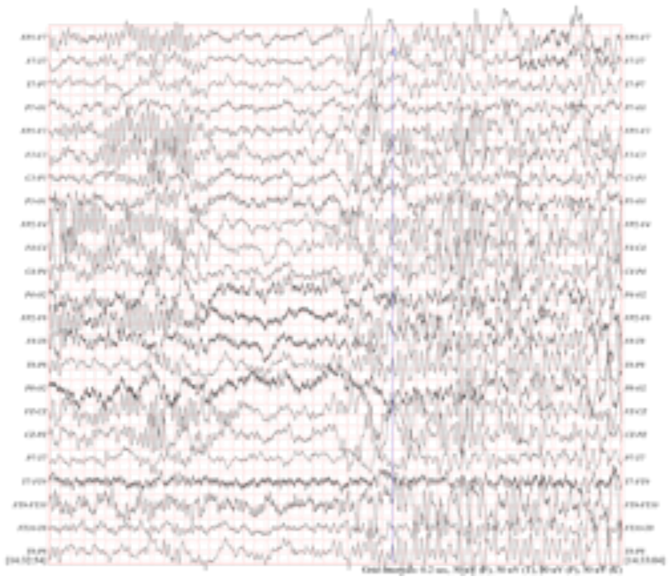
High-dimensional data

Dimension reduction

Low-dimensional data

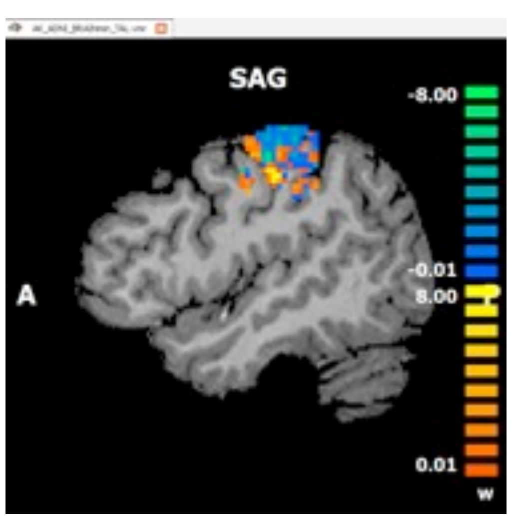
Processing/Display

What about neuronal data?



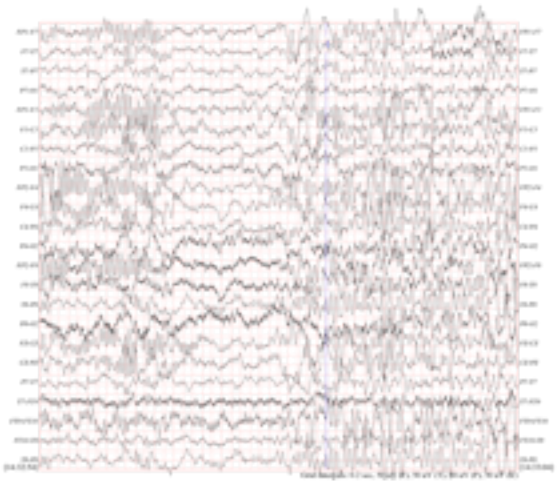
D(EEG) = # channels

D(spike_trains) = # neurons



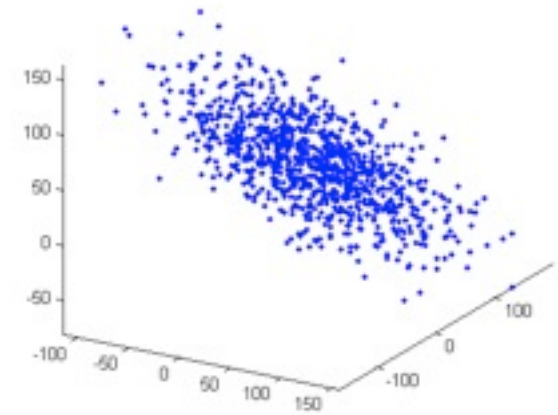
D(fMRI) = # voxels

Geometric view



set of
recordings

set of vectors

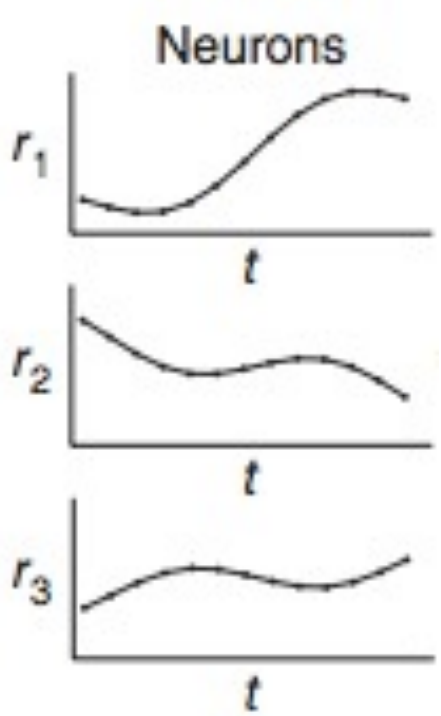


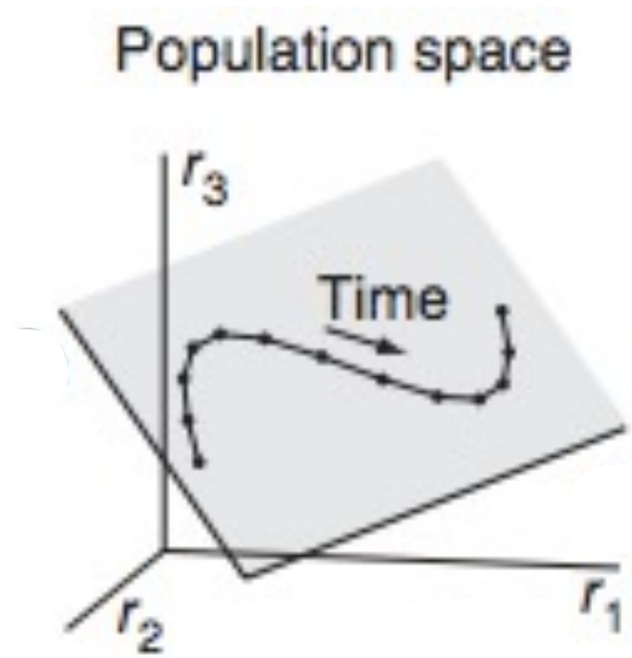
sample = 1 point/vector in a D-dimensional space

data set = cloud of points in a D-dimensional space

What about neuronal data?

In neural recordings one obtains a different point every time!

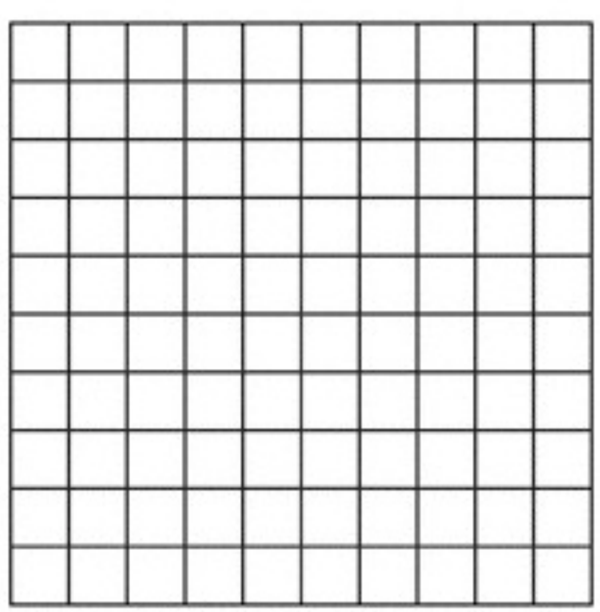


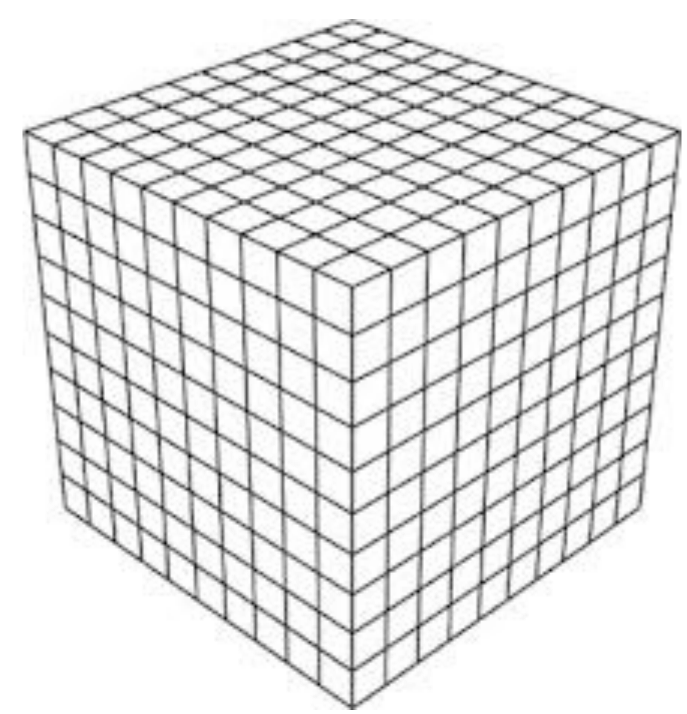


neuronal dynamics = trajectory in a D-dimensional space

What is the problem with high-dimensional data?

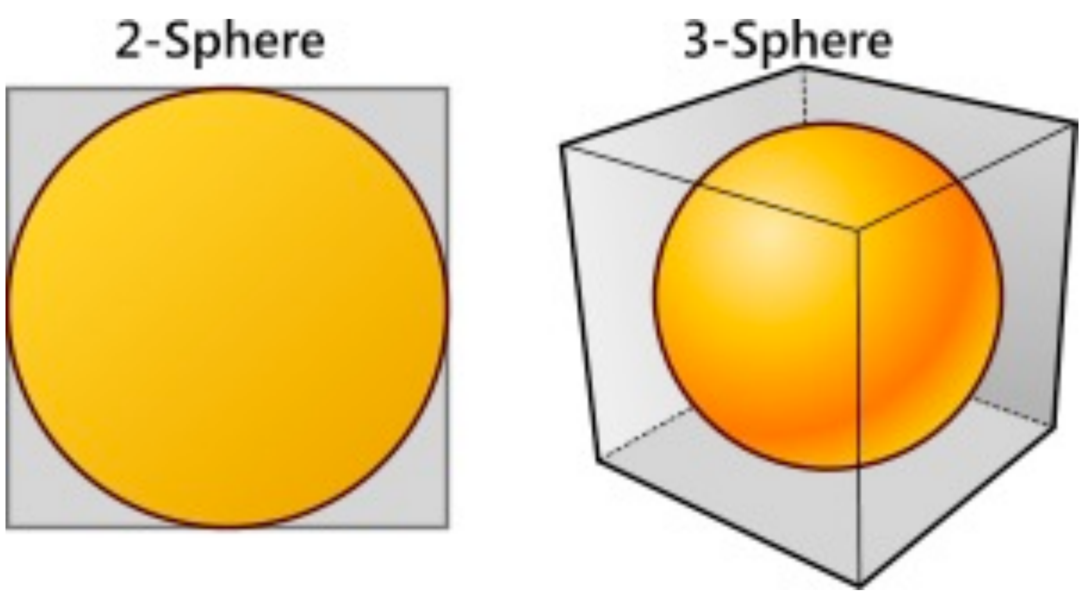
The curse of dimensionality: sample size to estimate a function of several variables grows exponentially with number of variables (D)

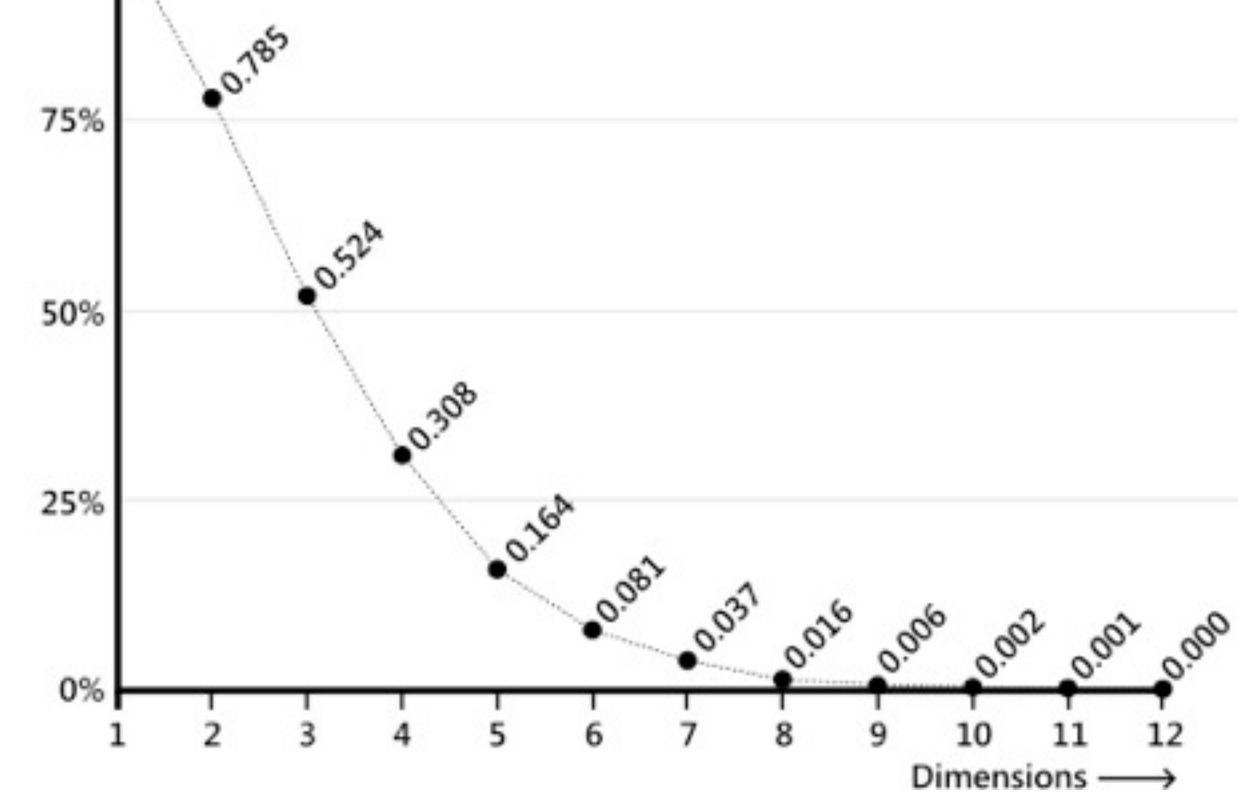




What is the problem with high-dimensional data?

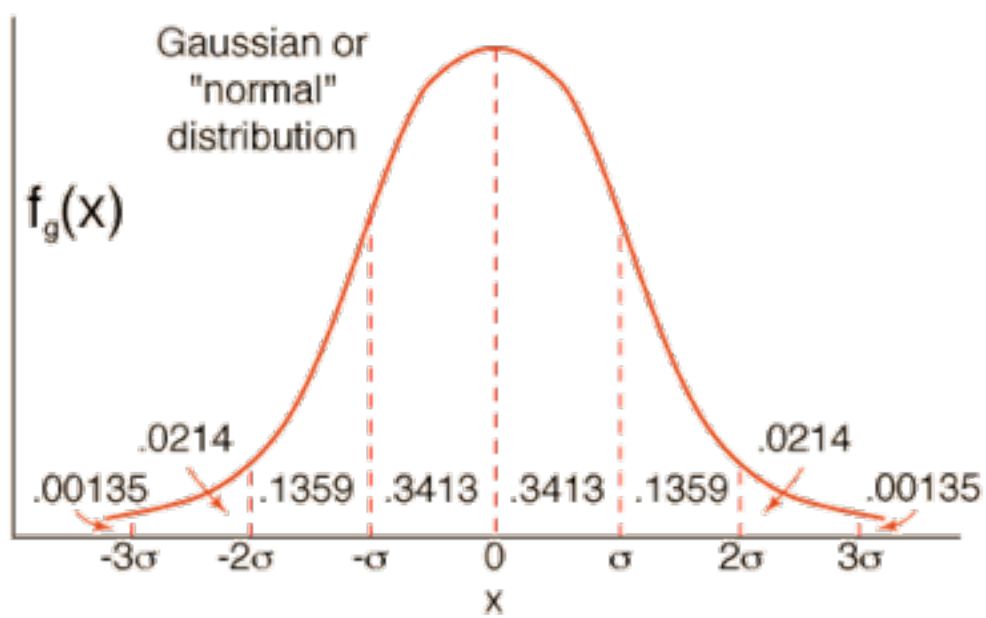
Our intuition fails in high dimensional spaces:
empty space phenomena!





What is the problem with high-dimensional data?

Our intuition fails in high dimensional spaces:
we are sampling from tails!

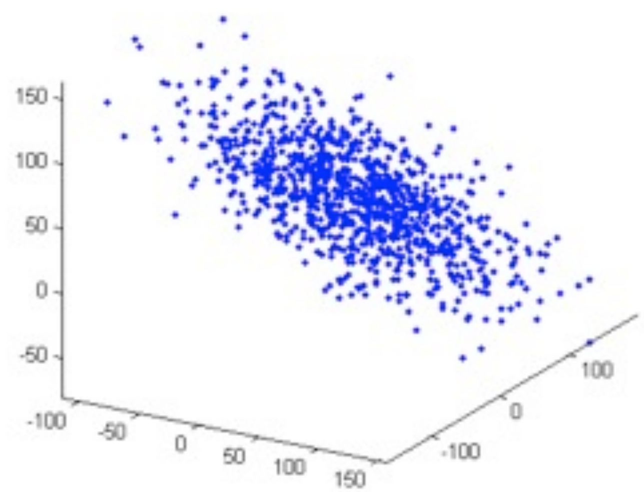


n	1	2	5	10	20	100
Probability	0.04550	0.13534	0.54942	0.94734	0.99995	1.00000

Tail probability of the multivariate Gaussian distributions for different dimensions (P(|x| > 2))

What is the problem with high-dimensional data?

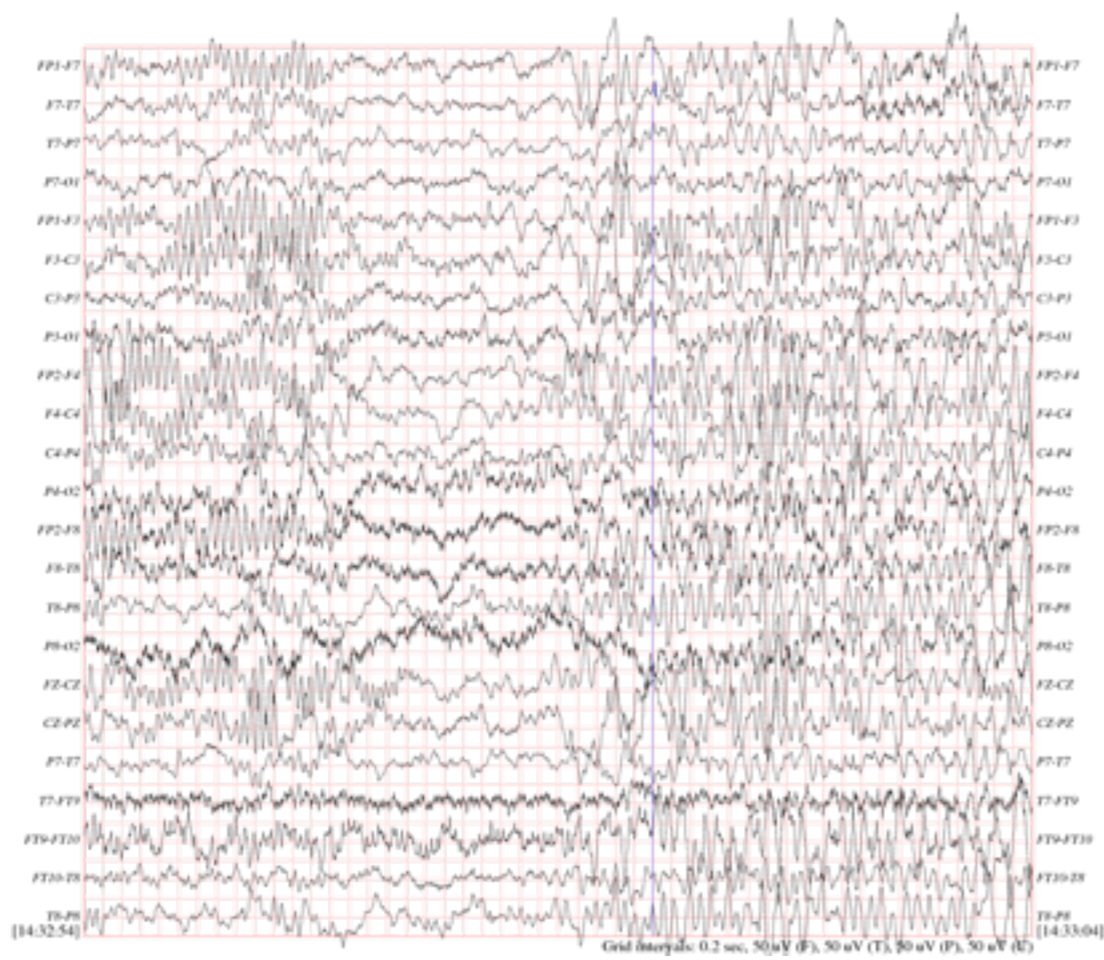
Our intuition fails in high dimensional spaces:
distances have weak discriminative power!



Euclidean distance between random vectors (I.I.D.) in high dimensions is approximately constant (k-NN is risky in very high dim.)

What is the problem with high-dimensional data?

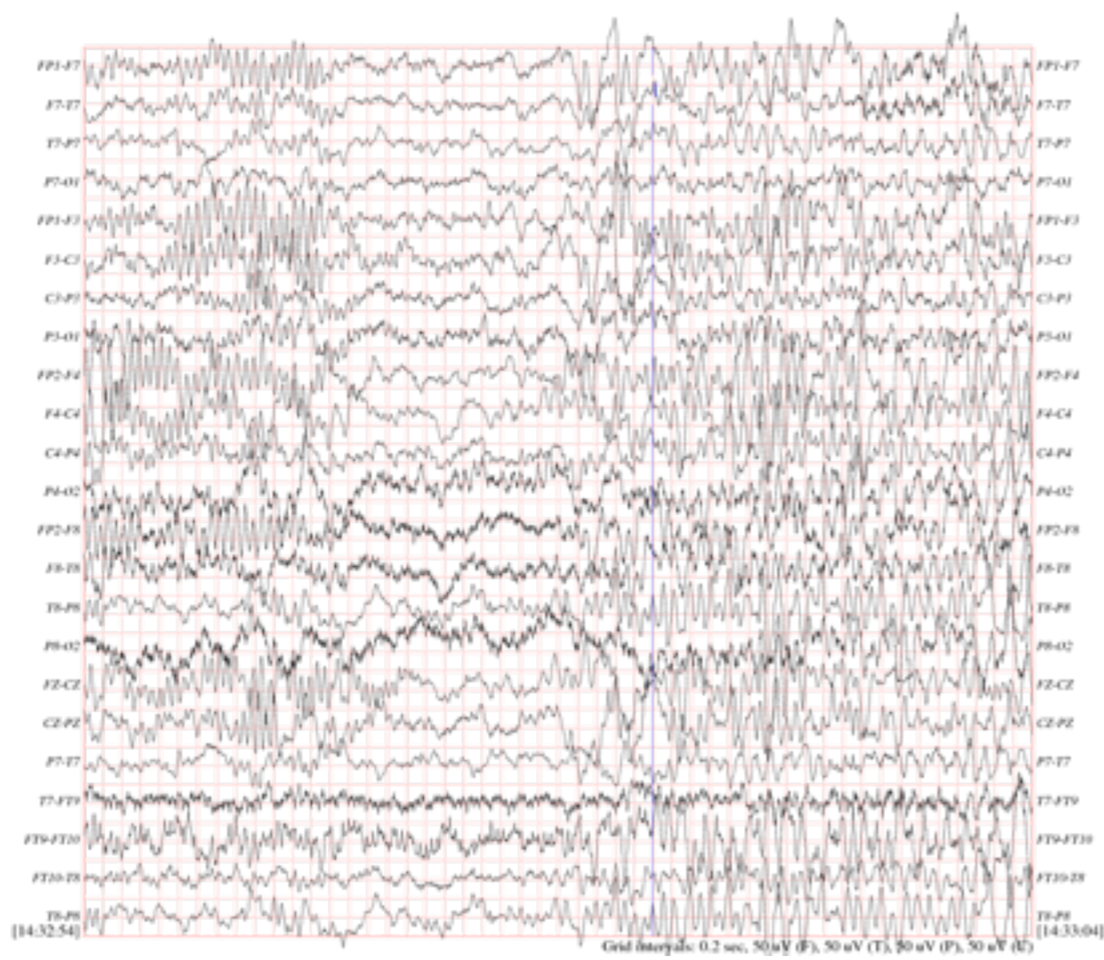
Visualization!



Not very insightful perspective for human processing...

What is the problem with high-dimensional data?

Visualization!



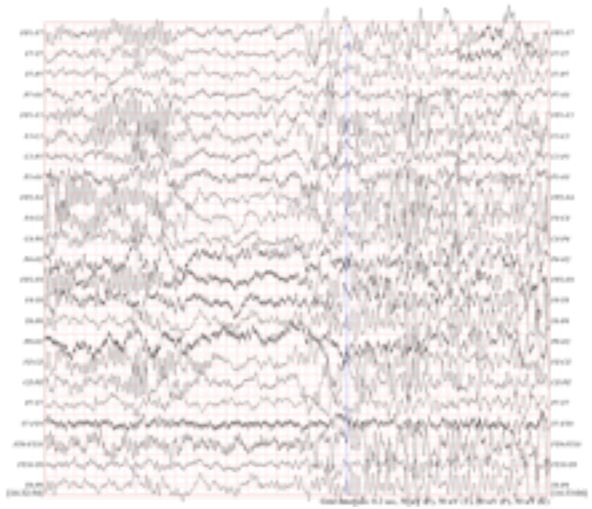
Not very insightful perspective for human processing...

What is the problem with high-dimensional data?

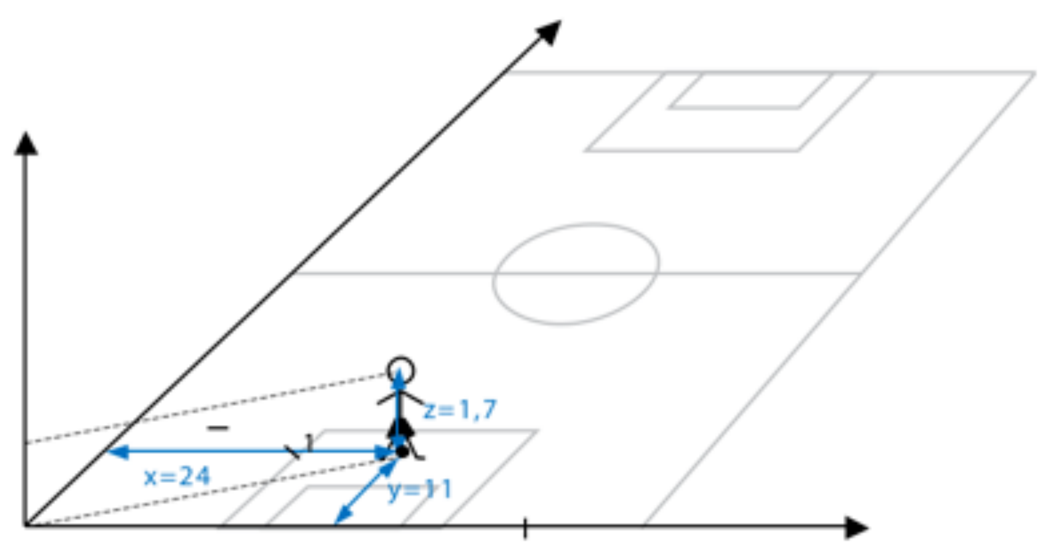
Visualization!

X vs time

Y vs time



X vs Y



Our brains are great pattern analyzers. It makes a huge difference when data can be visualized in 2d or 3d.

The hope

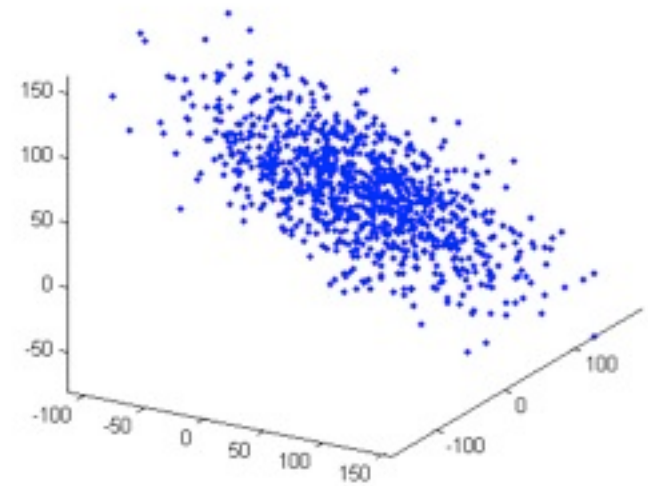
Reducing the dimension of our data in a sensible manner **can make an unfeasible problem** become **feasible**

We might to find **simplifying assumptions** that allow **to reduce the effective dimension** of the data

Geometric structure

How would be that data cloud if dataset would contain...

- noisy neurons?

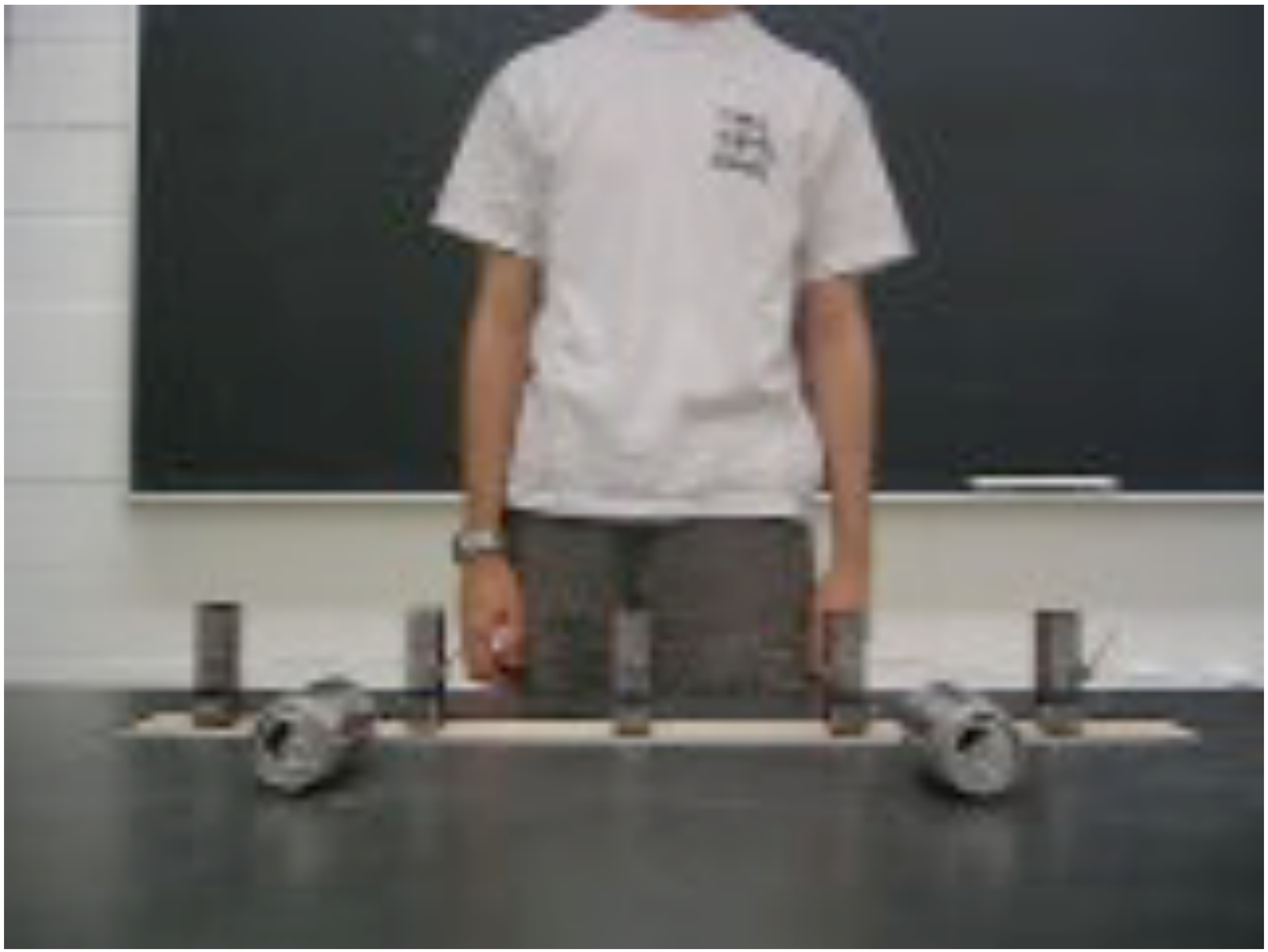


- neurons driven by a common source?

- neurons interacting with each other?

Example from Physics

Example from Physics



dimension reduction

Problem: Given an observed high-dimensional dataset

$$X = \{x_1, \cdots, x_m\} \in R^D$$

find a low-dimensional dataset

$$Y = \{y_1, \cdots, y_m\} \in R^d \qquad (d << D)$$

such that certain tasks of data analysis of the high dimensional data X can be realized on the low dimensional data Y with tolerable error

dimension reduction

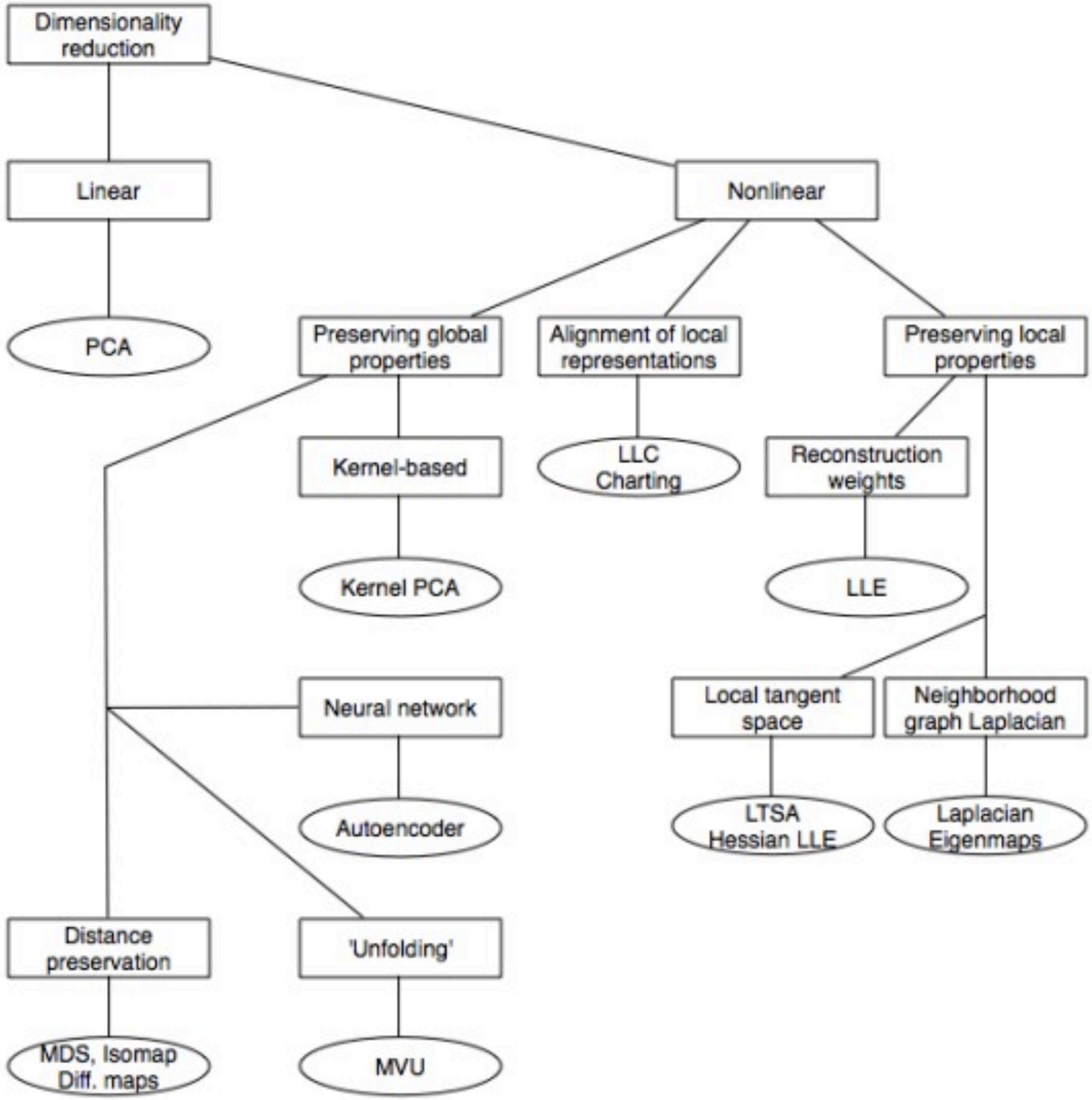
Reward:

- Data visualization
- Extracting key low dimensional features

When less is more

- What is dimensionality reduction?
- The zoo of dimensionality reduction
- Some applications

The zoo


Dimensionality reduction
Linear
Nonlinear
PCA
Preserving global properties
Alignment of local representations
Preserving local properties
Kernel-based
LLC Charting
Reconstruction weights
Kernel PCA
LLE
Neural network
Local tangent space
Neighborhood graph Laplacian
Autoencoder
LTSA Hessian LLE
Laplacian Eigenmaps
Distance preservation
'Unfolding'
MDS, Isomap Diff. maps
MVU

Our division

- Methods that rely on **projections**

- Methods that attempt to **model the manifold** on which data lies

Projective methods

Attempt to find low dimensional projections that extract useful information from the data, by maximizing some objective function

Example: **Projection Pursuit**

How are interesting directions in multidimensional data found?

Search for projections such that the projected data departs from normality

ICA

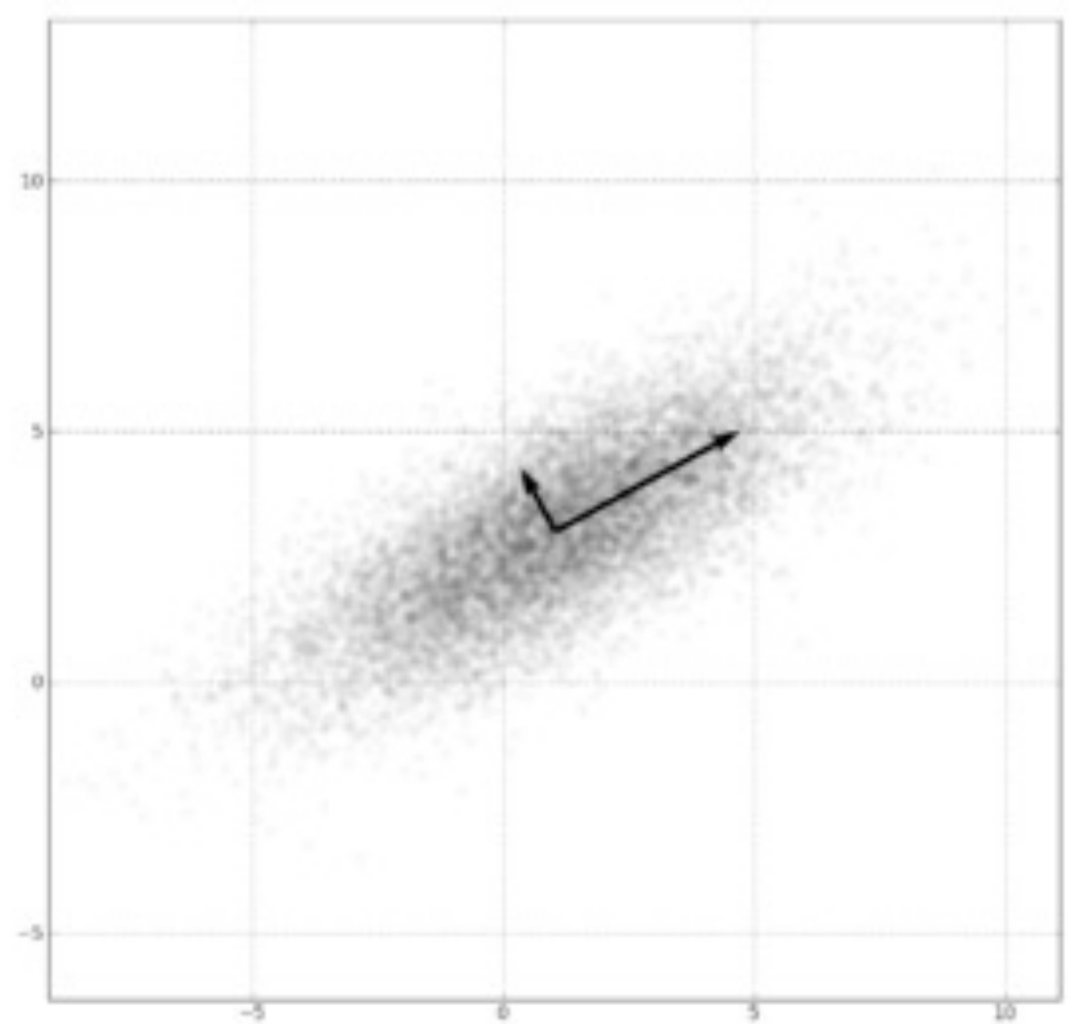
Independent Component Analysis: searches for projections such that the probability distributions of the data along those projections are statistically independent

- Applied for Blind Source Separation. Views data generated by a mixture of unknown latent variables

- Finds maximally non-Gaussian components (minimal mutual information components)

PCA

Principal Component Analysis: searches for projections along which the projected data has maximal variance



PCA

Principal Component Analysis: searches for projections along which the projected data has maximal variance

- If unbeknownst your data in fact lies along a line, PCA gives you the direction you are looking for

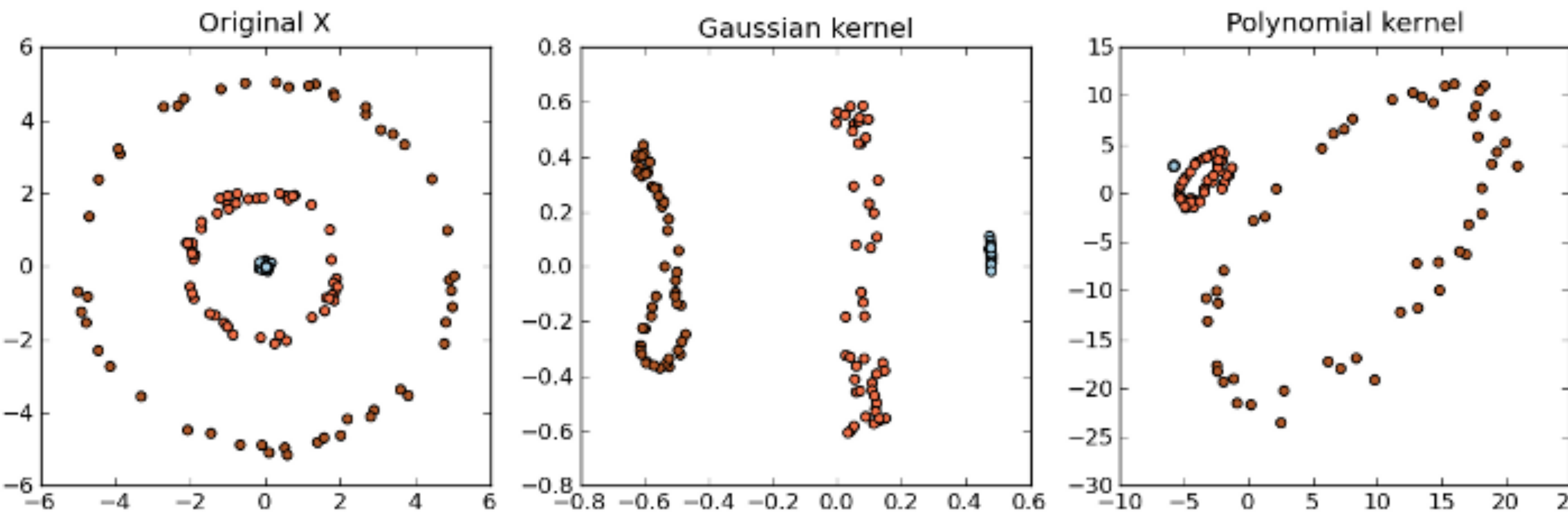
- PCA for feature extraction amounts to project the data to a lower dimensional space

- PCA decorrelates the data

- PCA maximizes mutual information (I(X,Y)) on Gaussian data

KPCA

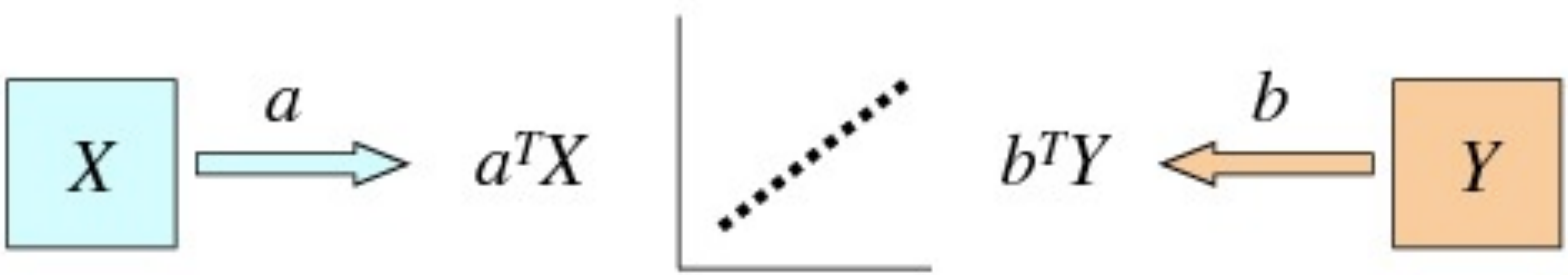
Kernel Principal Component Analysis: applies the kernel trick to create a non-linear version of PCA in sample space by performing PCA on a feature space.



CCA

Canonical Correlation Analysis: suppose we have two paired data sets *X* and *Y* (ex. different views of the same object)

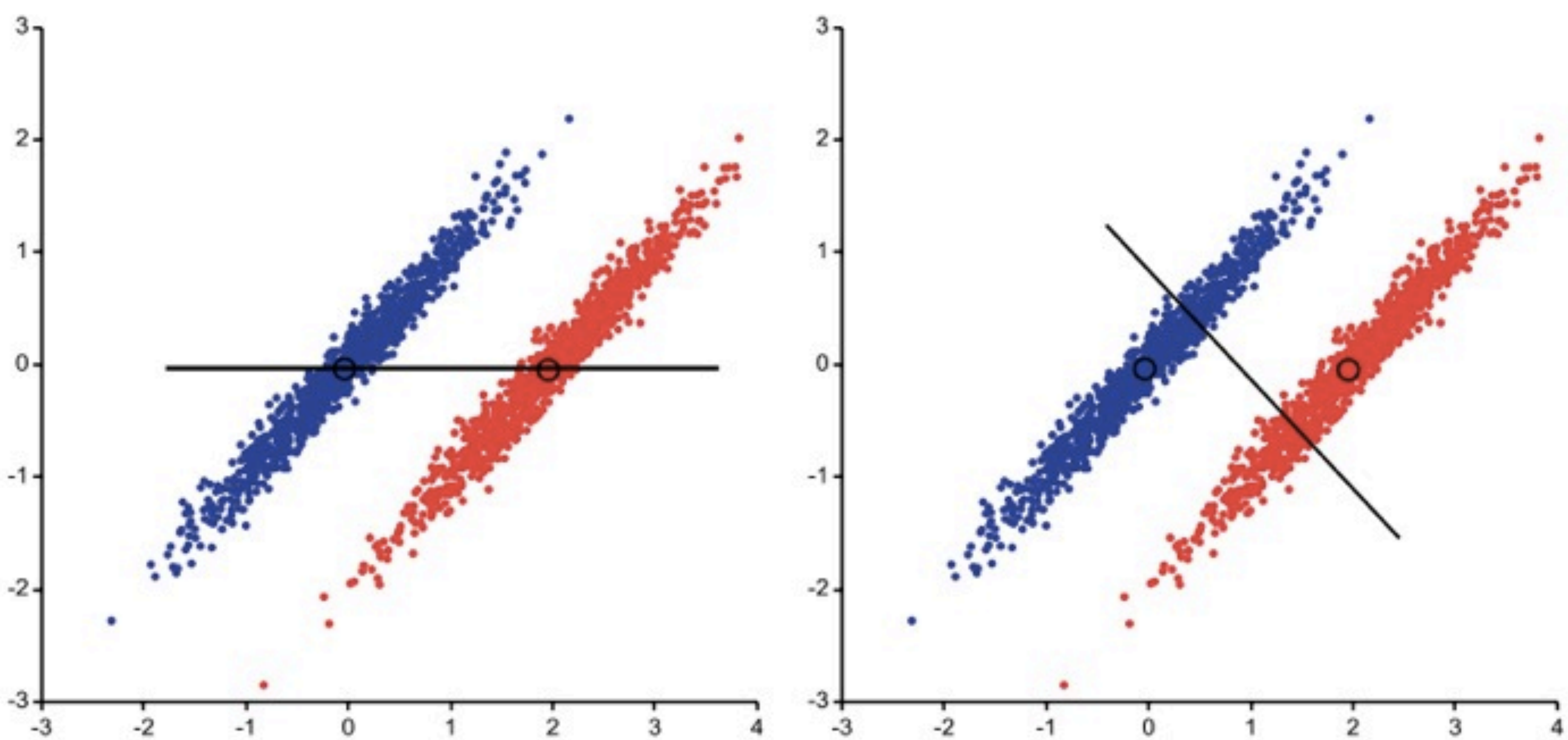
CCA finds paired directions such that the projection of the first data set along the first direction is maximally correlated with the projection of the second data set along the second direction.



LDA

Linear Discrimination Analysis: natural extension of PCA to the case of labeled data

LDA searches projections where two classes are well separated (maximizes inter-class while minimizing intra-class variance)



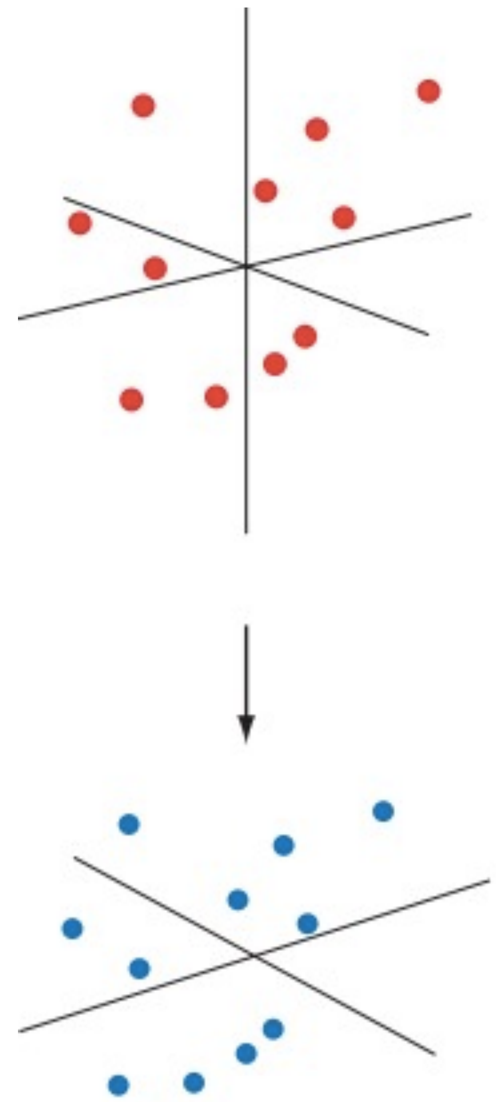
Random projections

Johnson-Lindenstrauss lema:

Lemma 7.1. *Let $\varepsilon > 0$ and integer n be given. Then for all positive integers $k \geqslant k_0 = O(\varepsilon^{-2} \log n)$ and a set $\mathscr{X}$ of n points, which are randomly selected in $\mathbb{R}^D$, there exists a projection $f : \mathbb{R}^D \to \mathbb{R}^k$, which satisfies*

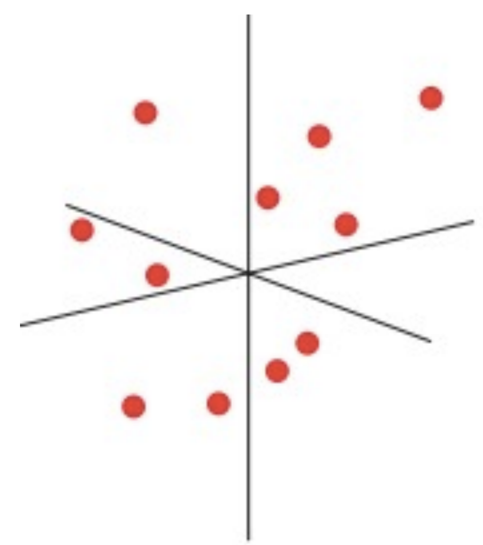
$$(1-\varepsilon)\|\boldsymbol{u}-\boldsymbol{v}\|^2 \leqslant \|f(\boldsymbol{u})-f(\boldsymbol{v})\|^2 \leqslant (1+\varepsilon)\|\boldsymbol{u}-\boldsymbol{v}\|^2, \qquad (7.3)$$

for all $\boldsymbol{u}, \boldsymbol{v} \in \mathscr{X}$.

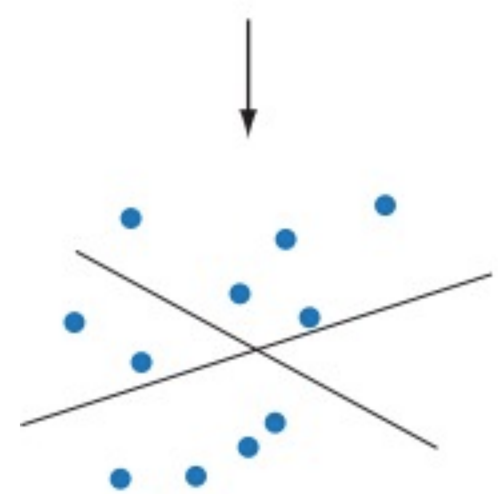


Random projections

Random projections: as opposed to PCA which measures data distortion globally, random projections can be used to guarantee that distances between all pairs are approximately maintained



$$f_r : \mathscr{X} \rightarrow \mathscr{Y} \subset \mathbb{R}^k, \boldsymbol{Y} = \boldsymbol{R}'\boldsymbol{X}$$



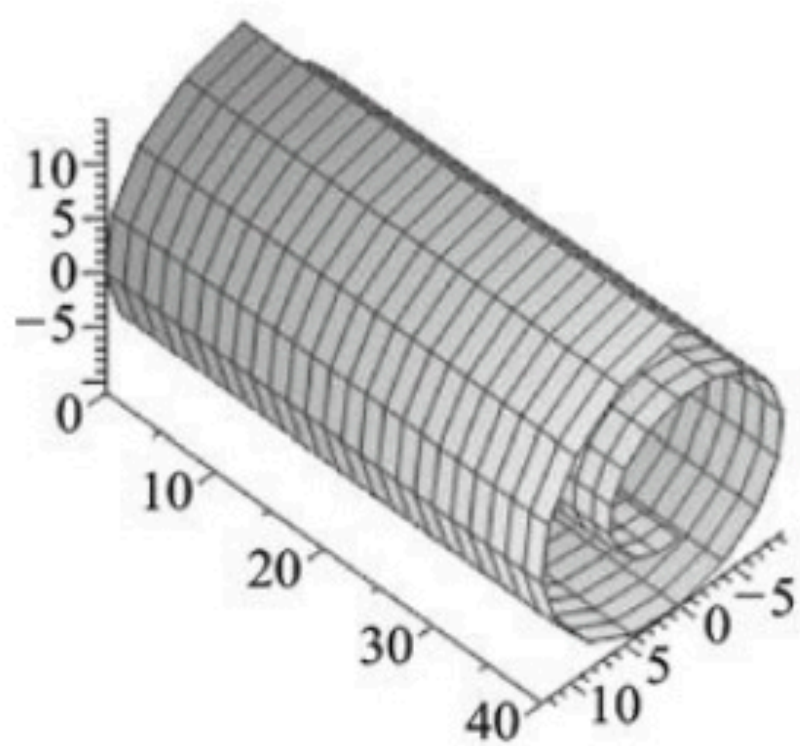
Our division

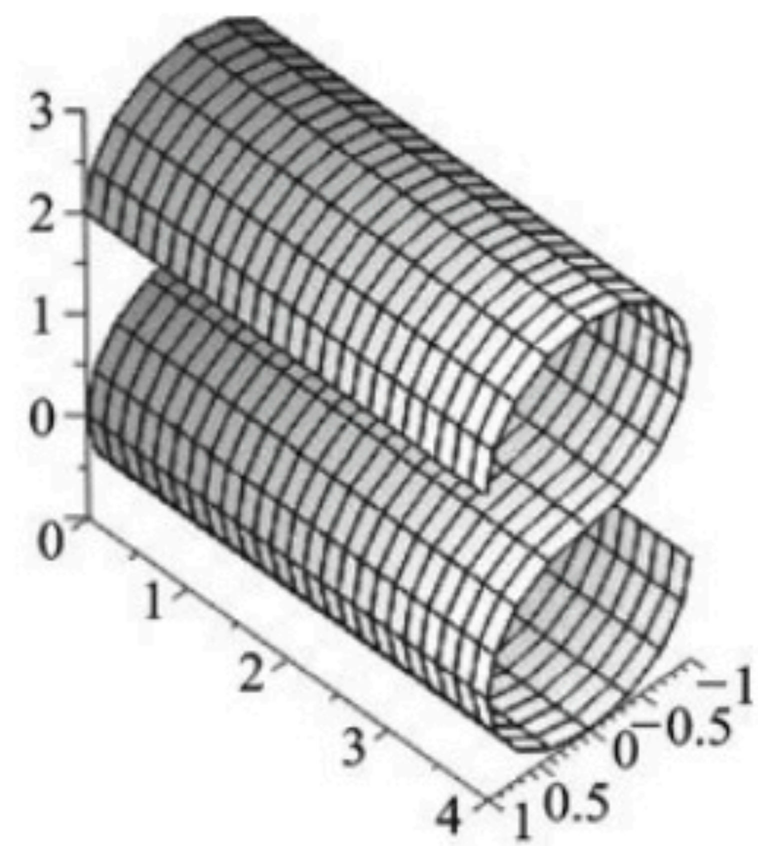
- Methods that rely on **projections**

- Methods that attempt to **model the manifold** on which data lies

Manifold Learning

Attempt to find the underlying geometric structure of data





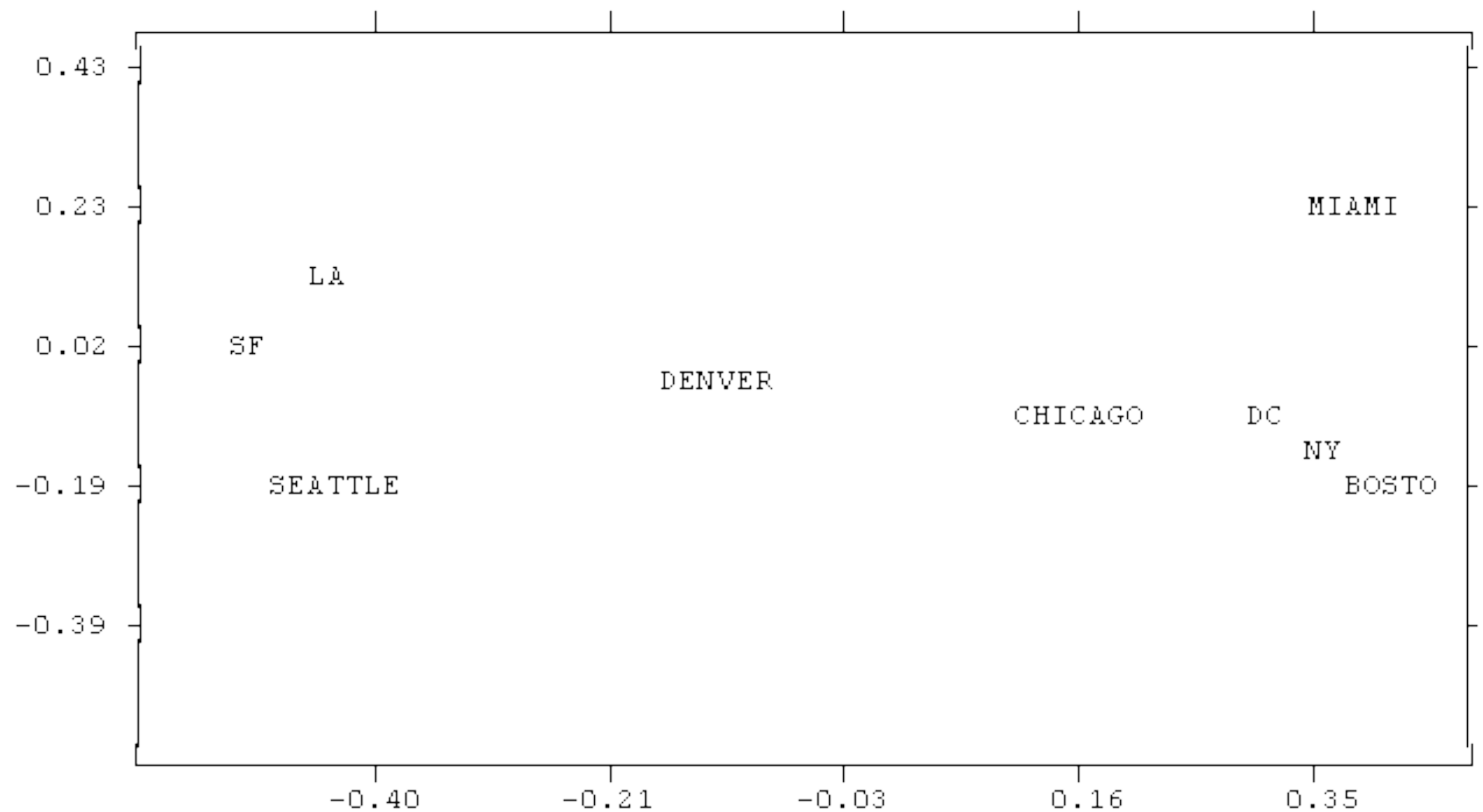
MDS

Multi-dimensional scaling: given a measure of dissimilarity between each pair of data points, MDS searches for a mapping of the dissimilarities to a low dimensional Euclidean space such that dissimilarities become squared distances

- Used in representational similarity analysis
- Input is only a matrix of dissimilarities
- Equivalent to PCA is some cases
- Visualization

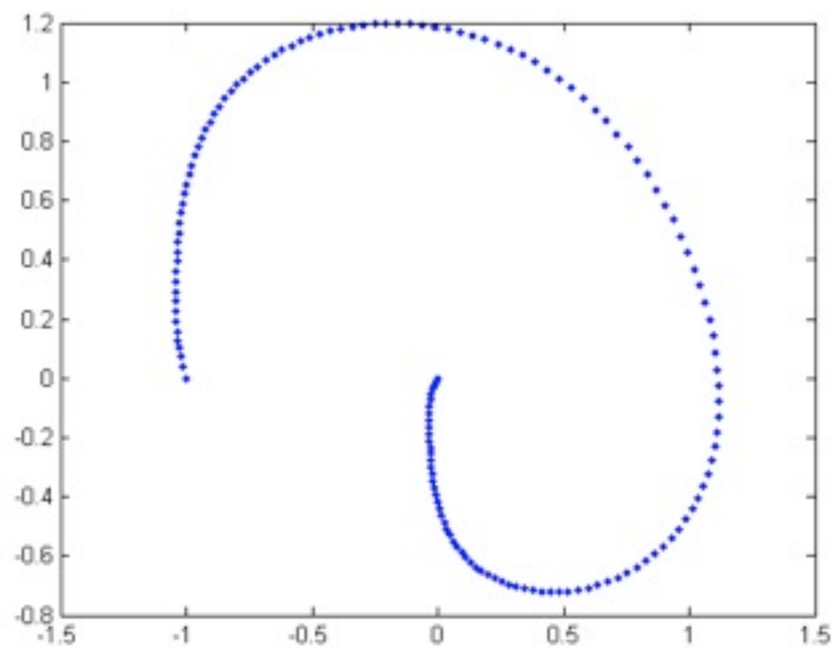
	1 BOST	2 NY	3 DC	4 MIAM	5 CHIC	6 SEAT	7 SF	8 LA	9 DENV
BOSTON	0	206	429	1504	963	2976	3095	2979	1949
NY	206	0	233	1308	802	2815	2934	2786	1771
DC	429	233	0	1075	671	2684	2799	2631	1616
MIAMI	1504	1308	1075	0	1329	3273	3053	2687	2037
CHICAGO	963	802	671	1329	0	2013	2142	2054	996
SEATTLE	2976	2815	2684	3273	2013	0	808	1131	1307
SF	3095	2934	2799	3053	2142	808	0	379	1235
LA	2979	2786	2631	2687	2054	1131	379	0	1059
DENVER	1949	1771	1616	2037	996	1307	1235	1059	0

	1	2	3	4	5	6	7	8	9
	BOST	NY	DC	MIAM	CHIC	SEAT	SF	LA	DENV
BOSTON	0	206	429	1504	963	2976	3095	2979	1949
NY	206	0	233	1308	802	2815	2934	2786	1771
DC	429	233	0	1075	671	2684	2799	2631	1616
MIAMI	1504	1308	1075	0	1329	3273	3053	2687	2037
CHICAGO	963	802	671	1329	0	2013	2142	2054	996
SEATTLE	2976	2815	2684	3273	2013	0	808	1131	1307
SF	3095	2934	2799	3053	2142	808	0	379	1235
LA	2979	2786	2631	2687	2054	1131	379	0	1059
DENVER	1949	1771	1616	2037	996	1307	1235	1059	0



Isomap

Isomap: key assumption is that the quantity of interest, when comparing two points, is the distance along the curve

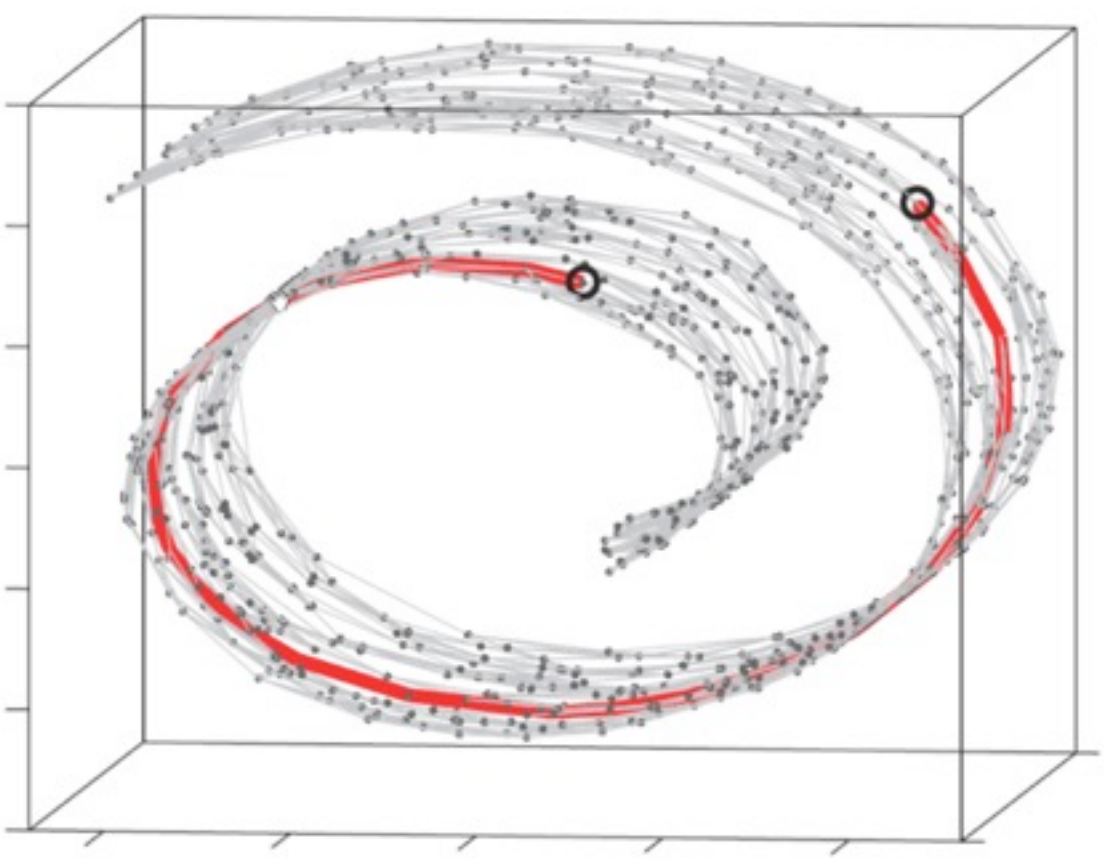


- Example: if data lies on a curve PCA and MDS will fail to discover that data is 1 dim.

- Isomap uses geodesic distance as input dissimilarities in a MDS to find a low dimensional representation (preserving global structure)

Isomap

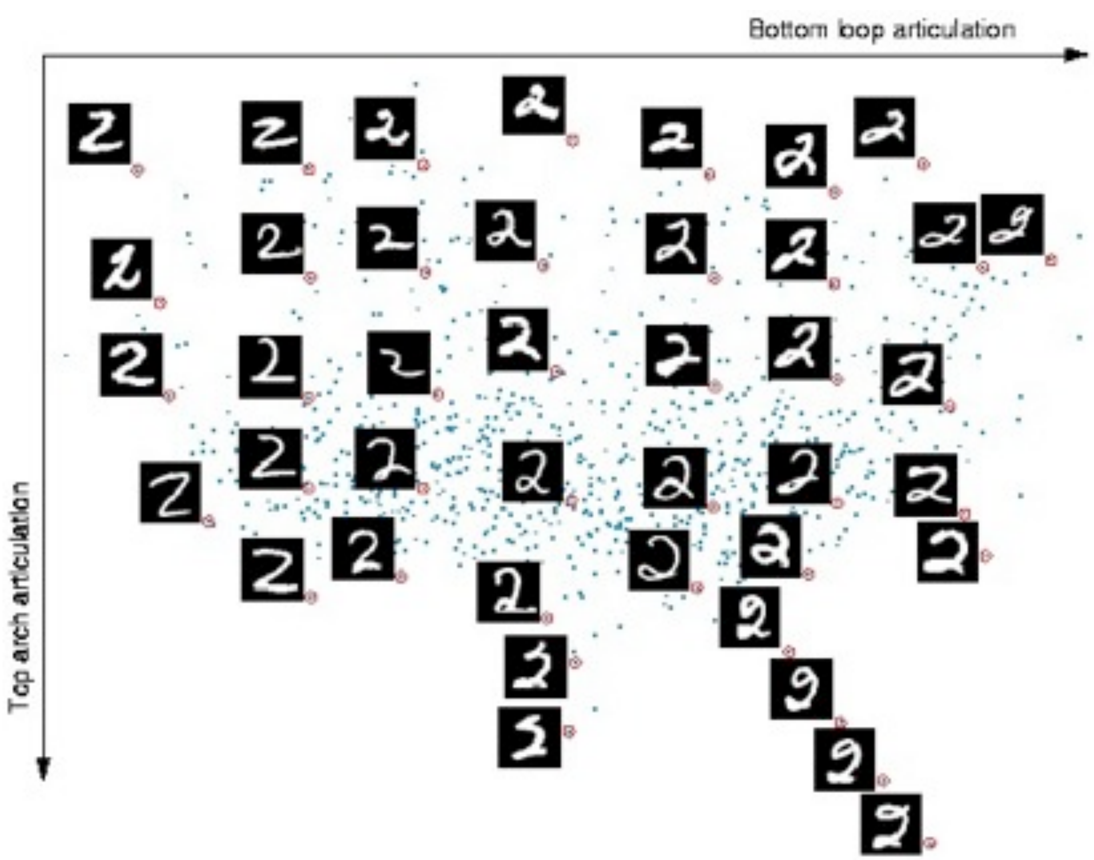
Isomap: key assumption is that the quantity of interest, when comparing two points, is the distance along the curve

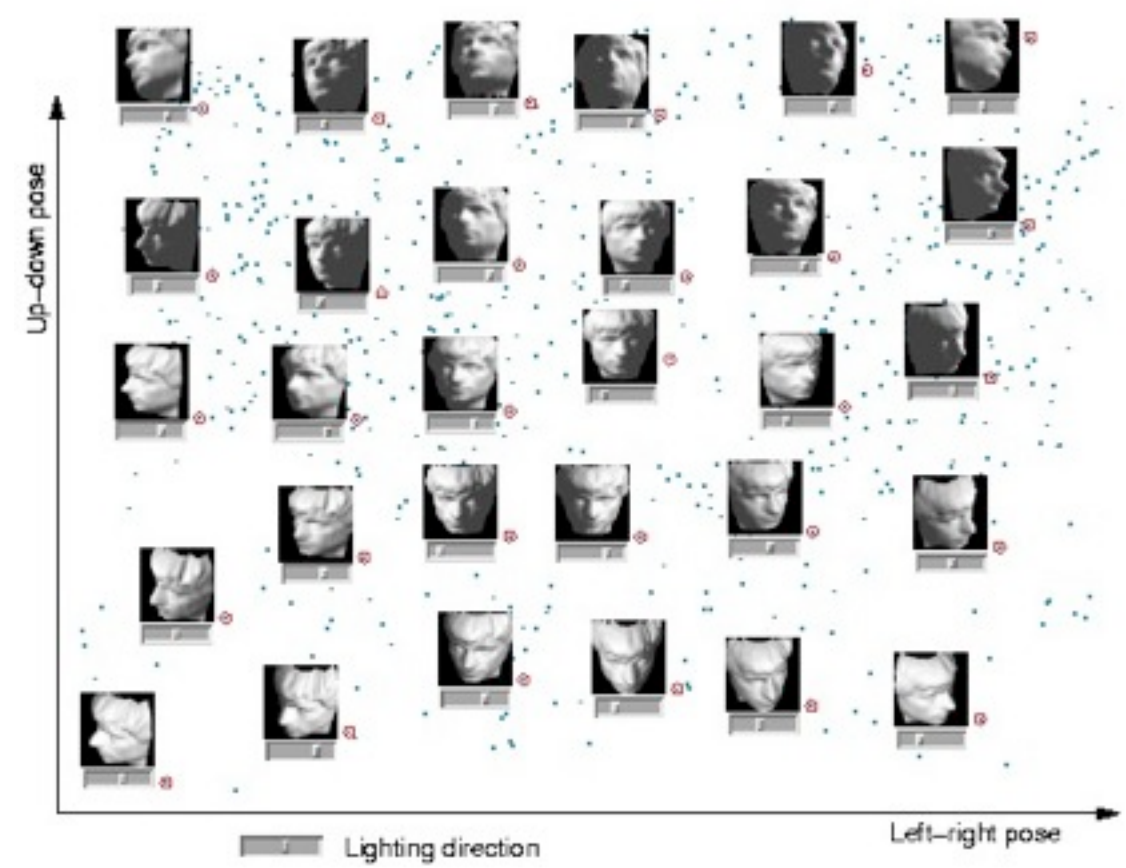


1. Cloud of points to graph
2. Geodesic distance matrix
3. MDS

Isomap

Isomap: unfolds a manifold by keeping the geodesic metric of the original data set





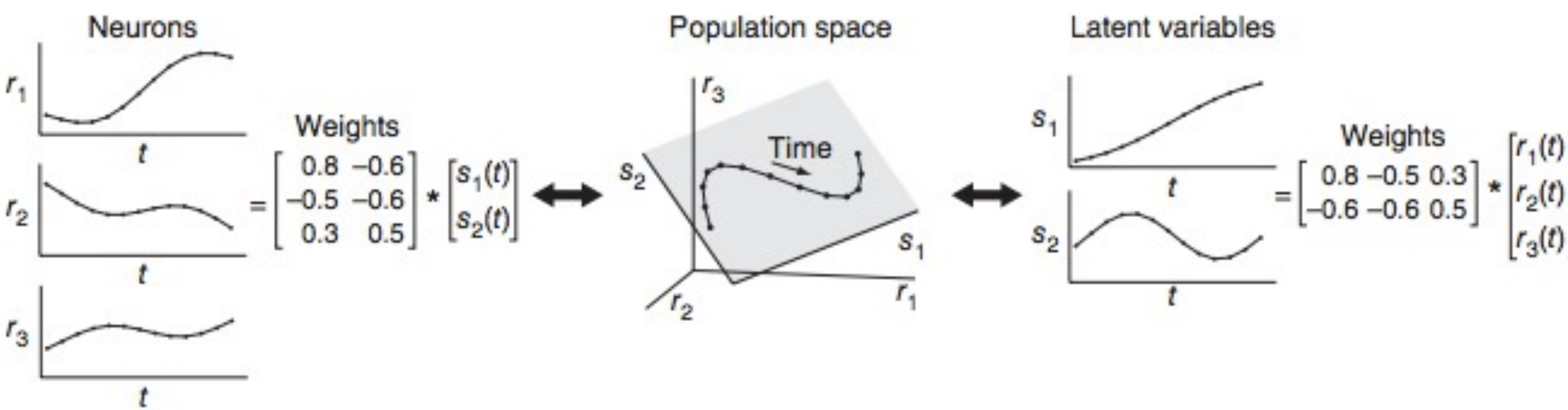
When less is more

- What is dimensionality reduction?
- The zoo of dimensionality reduction
- Some applications

Some applications

Population response: neural code and processing involve the coordination of responses across neurons

• Dimension reduction can help to interpret all variables simultaneously and discover latent variables



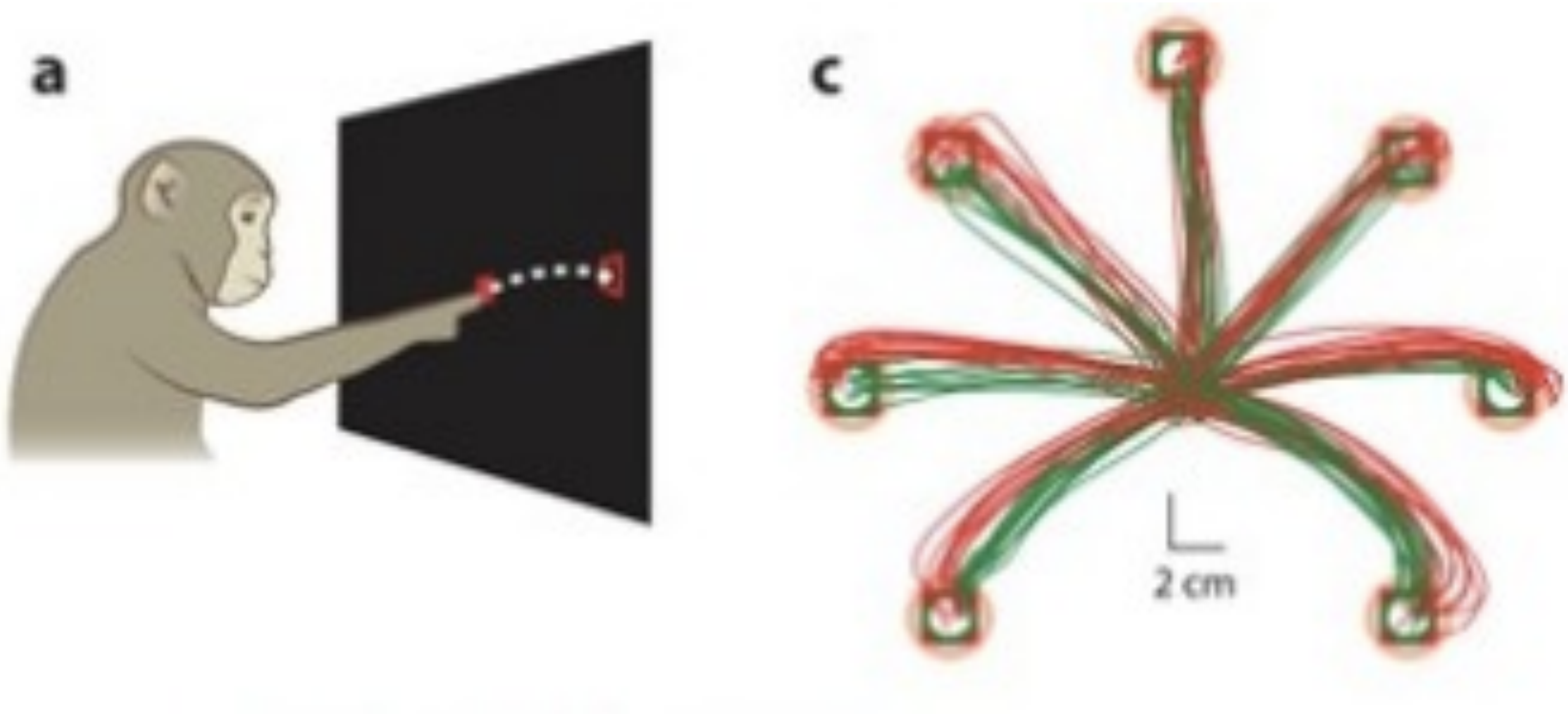
Dimensionality reduction for large-scale neural recordings, Cunningham, J.P., Yu, B

Nature Neuroscience **17**, 1500–1509 (2014)

Some applications

Exploratory analysis: visualization in 2d or 3d has an enormous impact in interpreting and generating hypothesis

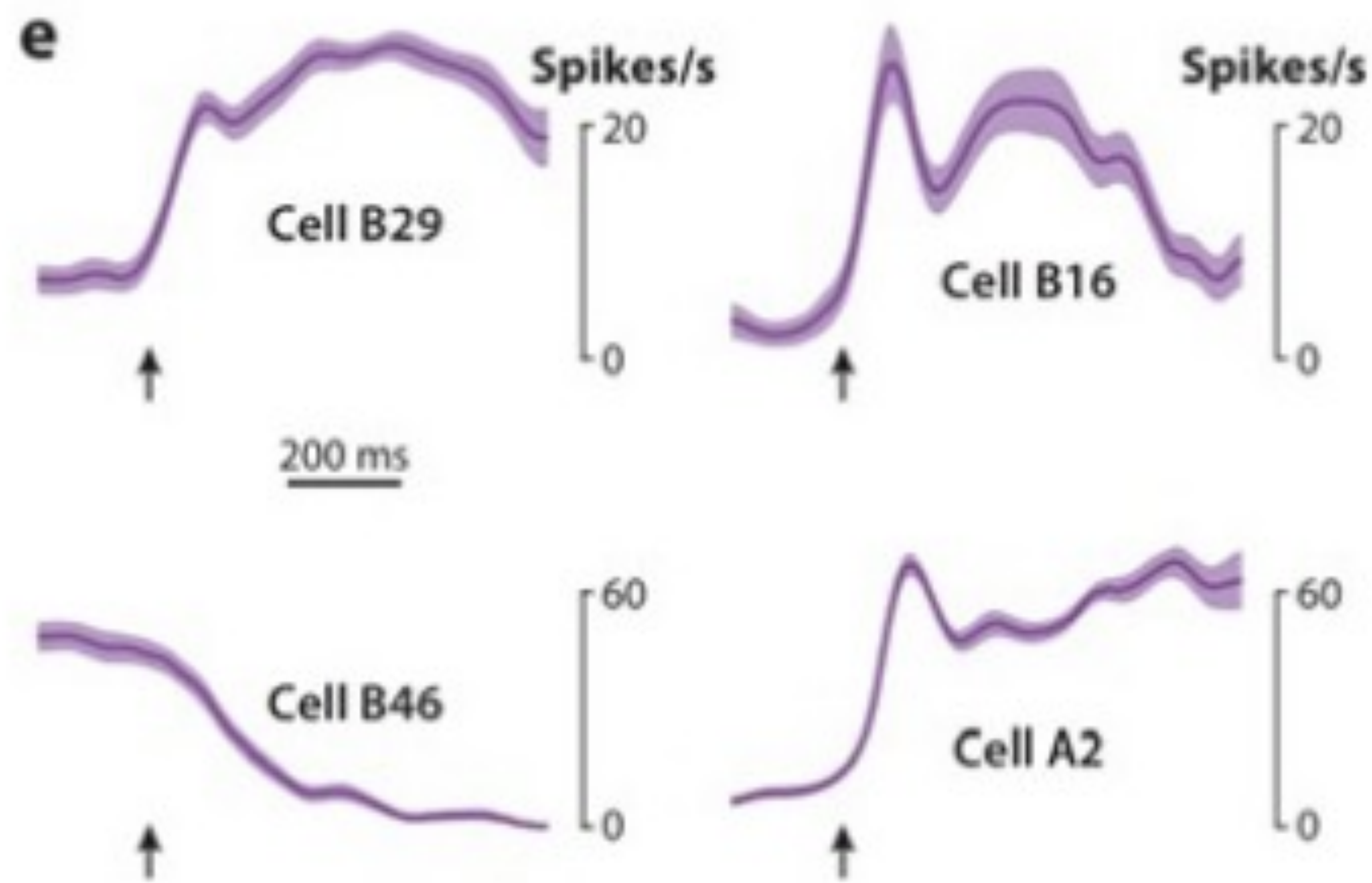
- Example: motor cortex



Some applications

Exploratory analysis: visualization in 2d or 3d has an enormous impact in interpreting and generating hypothesis

- Example: motor cortex

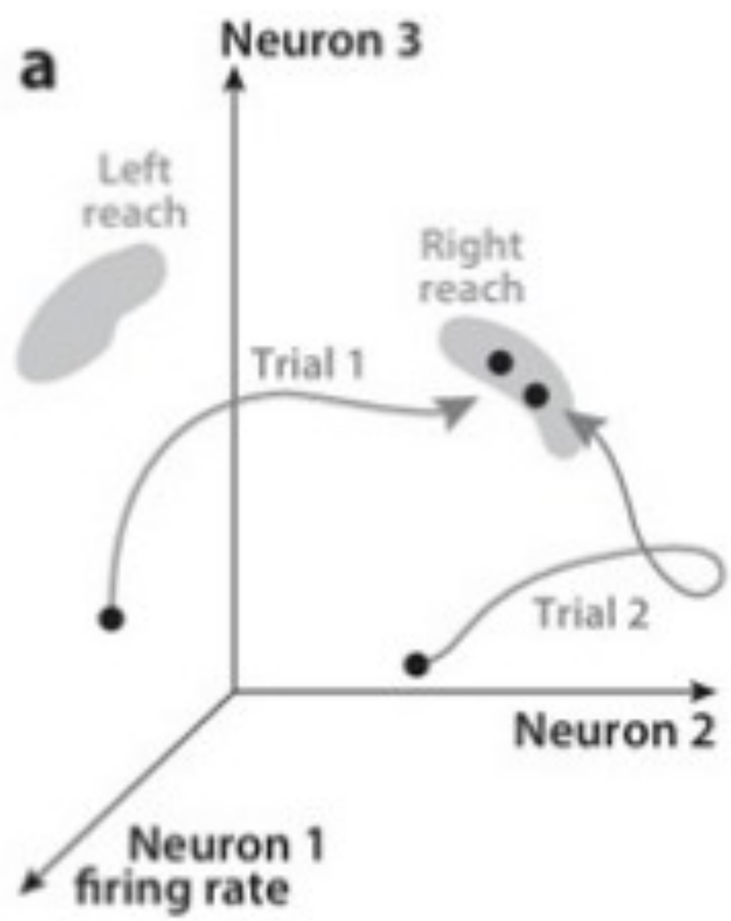


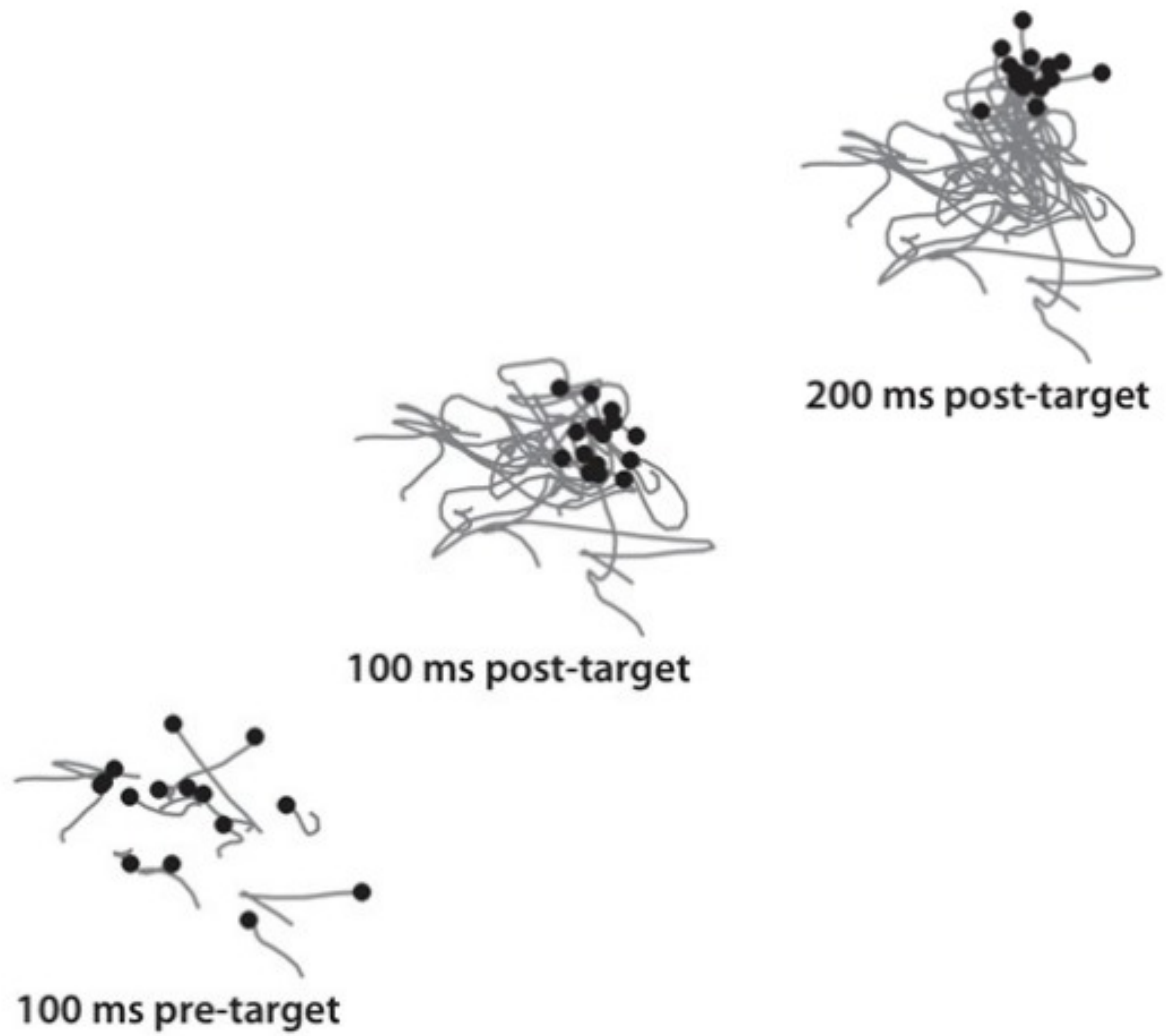
108 neurons

Some applications

Exploratory analysis: visualization in 2d or 3d has an enormous impact in interpreting and generating hypothesis

- Example: motor cortex

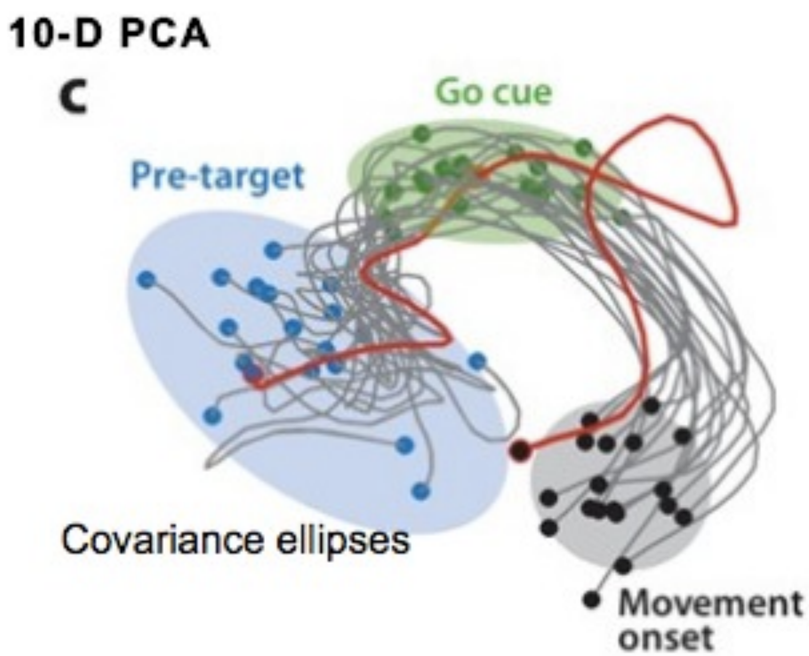


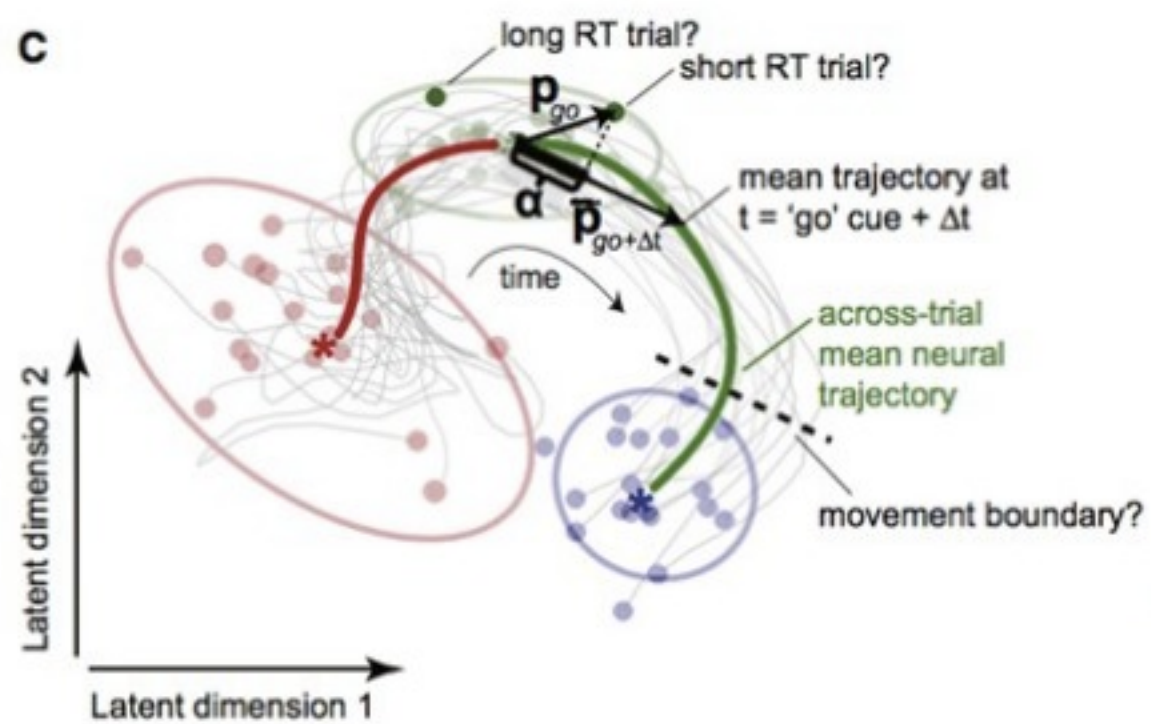


Some applications

Exploratory analysis: visualization in 2d or 3d has an enormous impact in interpreting and generating hypothesis

- Example: motor cortex





Dimension reduction can be useful in Neuroscience but most data is still analyzed only classically

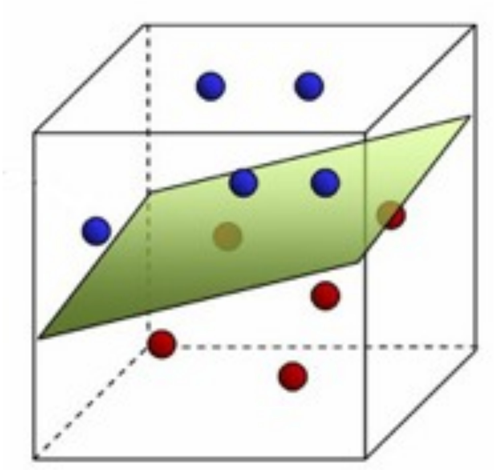
- A lot of work to be done by just applying the techniques we have seen today to interesting data sets (Visualization)

- Compressed sensing (faster fMRI scans)

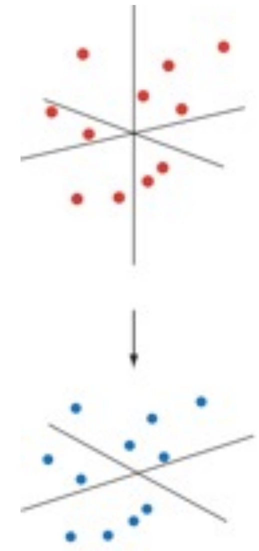
Take home messages

- Sometimes less is more
- Both the brain and the analyst would like to implement dimension reduction
- There are a lot of data sets waiting

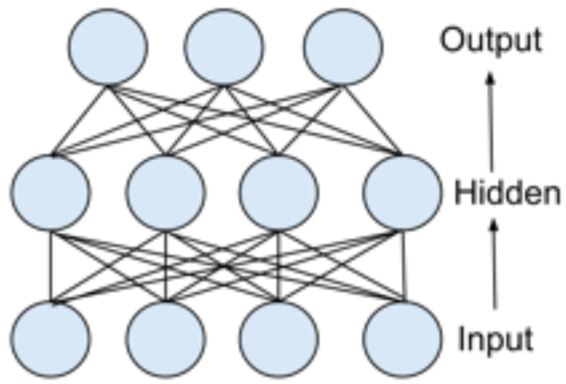
Machine learning



Dimensionality reduction



Caveats



"The brain is a very generous organ, it offers enough data to fit almost every theory"

- Nikos Logothetis

“If you torture the data enough it will confess to anything”

- Ronald Coase

confirmation bias

20 years of learning about vision: Questions answered, questions unanswered, and questions not yet asked

Bruno A. Olshausen (2010)

confirmation bias

20 years of learning about vision: Questions answered, questions unanswered, and questions not yet asked

Bruno A. Olshausen (2010)

too early for hypothesis driven research in systems neuroscience?

confirmation bias

20 years of learning about vision: Questions answered, questions unanswered, and questions not yet asked

Bruno A. Olshausen (2010)

too early for hypothesis driven research in systems neuroscience?

- Story telling findings are compelling

confirmation bias

20 years of learning about vision: Questions answered, questions unanswered, and questions not yet asked

Bruno A. Olshausen (2010)

too early for hypothesis driven research in systems neuroscience?

- Story telling findings are compelling
- A "must" to publish in glamorous journals

confirmation bias

20 years of learning about vision: Questions answered, questions unanswered, and questions not yet asked

Bruno A. Olshausen (2010)

too early for hypothesis driven research in systems neuroscience?

- Story telling findings are compelling
- A "must" to publish in glamorous journals
- Most of our hypothesis are probably wrong

confirmation bias

20 years of learning about vision: Questions answered, questions unanswered, and questions not yet asked

Bruno A. Olshausen (2010)

too early for hypothesis driven research in systems neuroscience?

- Story telling findings are compelling
- A "must" to publish in glamorous journals
- Most of our hypothesis are probably wrong
- Negative results disappear out of view

confirmation bias

20 years of learning about vision: Questions answered, questions unanswered, and questions not yet asked

Bruno A. Olshausen (2010)

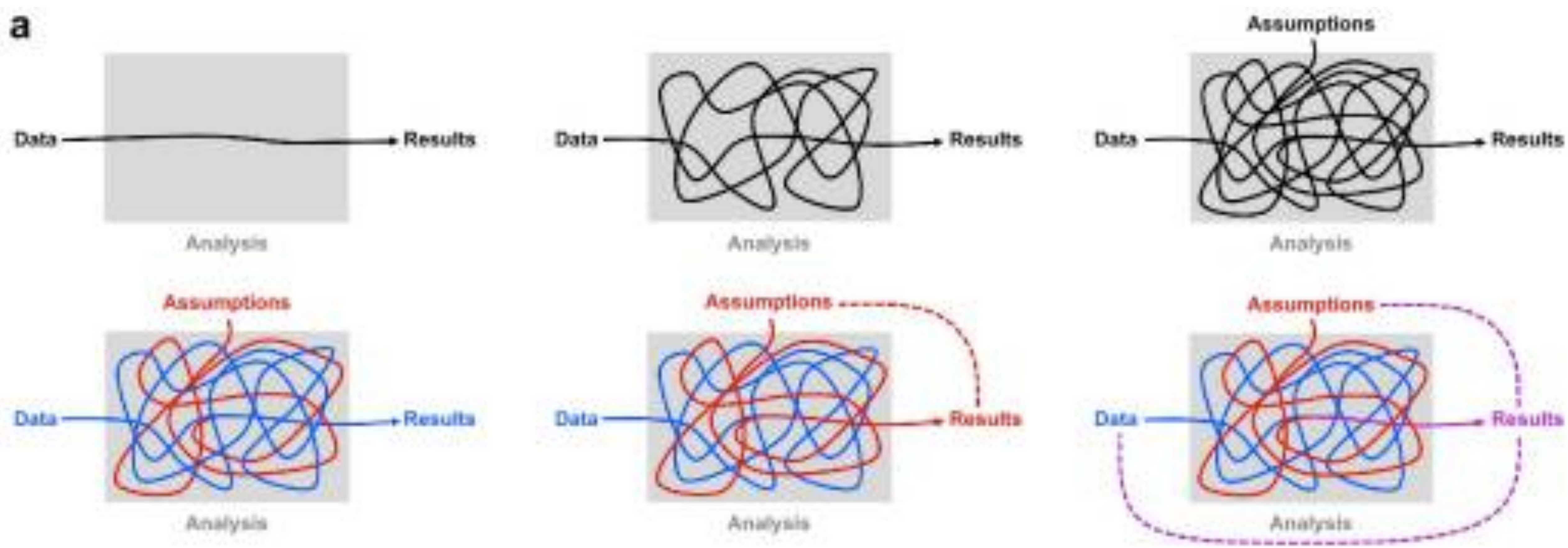
too early for hypothesis driven research in systems neuroscience?

- Story telling findings are compelling
- A “must” to publish in glamorous journals
- Most of our hypothesis are probably wrong
- Negative results disappear out of view

Huge multiple comparison problem not accounted for!

Analysis is what lies between your data and results and it is very easy to contaminate

Circular analysis in systems neuroscience: the dangers of double dipping.
Kriegeskorte, N. et al., Nature Neuroscience 2009, **12,** p 535.



confirmation bias

Circular analysis in systems neuroscience: the dangers of double dipping
Nikolaus Kriegeskorte et al. (2009)

(Nature, Science, Nature Neuroscience, Neuron, Journal of Neuroscience) in 2008.
134 fMRI papers, 42% (57 papers) contained at least one nonindependent selective analysis

confirmation bias

Selection criteria depends on results

analysis (sensitivity)
channels (neurons, regions, animals)
parameters (uncertainty, scanning)
statistics (+/- conservative)

confirmation bias

Selection criteria depends on results

analysis (sensitivity)
channels (neurons, regions, animals)
parameters (uncertainty, scanning)
statistics (+/- conservative)

Computational models can fit almost anything!

confirmation bias

Selection criteria depends on results

analysis (sensitivity)
channels (neurons, regions, animals)
parameters (uncertainty, scanning)
statistics (+/- conservative)

Computational models can fit almost anything!

connectivity
currents
input

confirmation bias

Selection criteria depends on results

analysis (sensitivity)
channels (neurons, regions, animals)
parameters (uncertainty, scanning)
statistics (+/- conservative)

Computational models can fit almost anything!

connectivity
currents
input

$\longrightarrow$

amplitude
frequency
phase relation

confirmation bias

Selection criteria depends on results

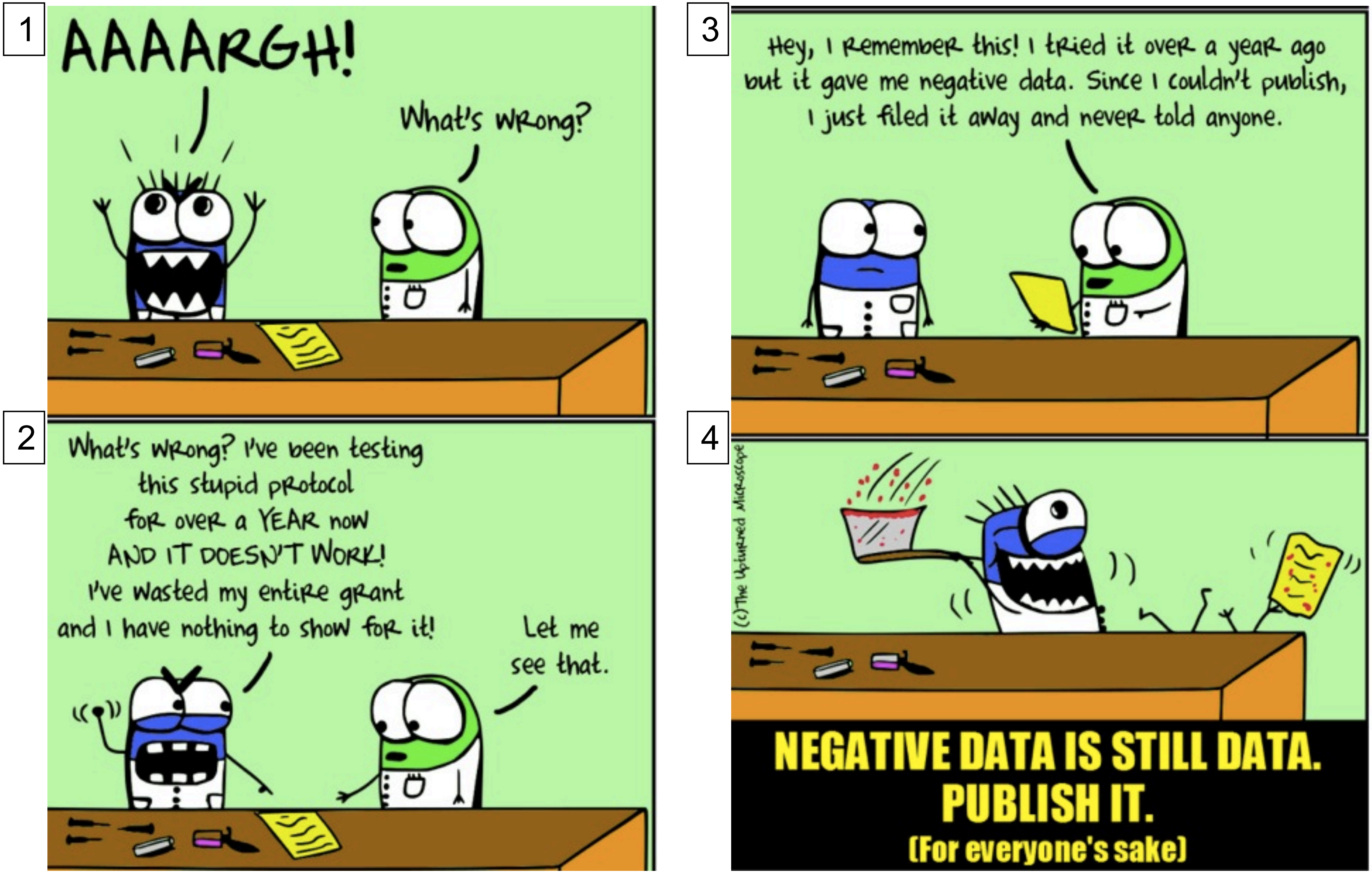
analysis (sensitivity)
channels (neurons, regions, animals)
parameters (uncertainty, scanning)
statistics (+/- conservative)

Computational models can fit almost anything!

connectivity
currents
input

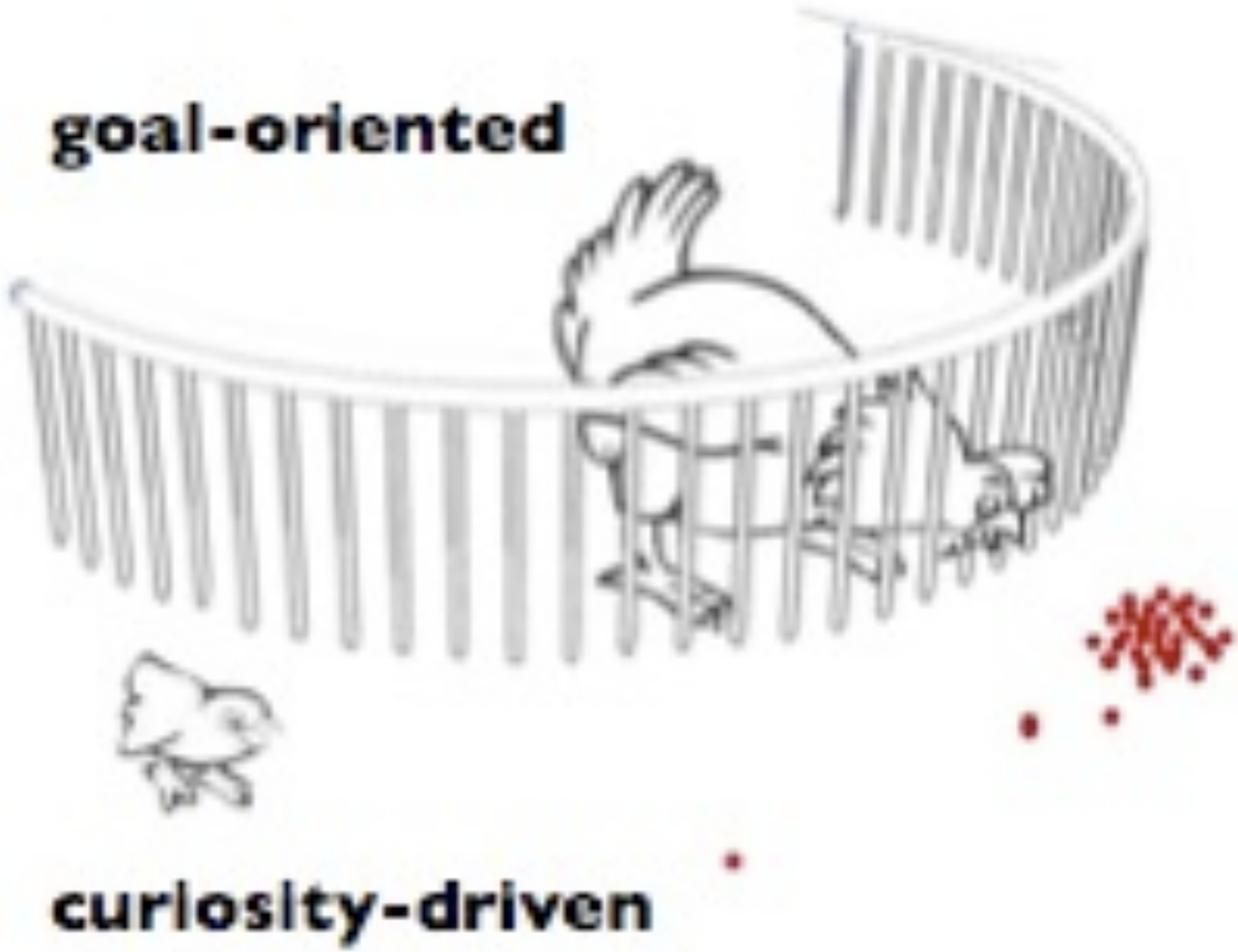
amplitude
frequency
phase relation

calibration


1
AAAARGH!
What's wrong?
2
What's wrong? I've been testing this stupid protocol for over a YEAR now AND IT DOESN'T WORK! I've wasted my entire grant and I have nothing to show for it!
Let me see that.
3
Hey, I remember this! I tried it over a year ago but it gave me negative data. Since I couldn't publish, I just filed it away and never told anyone.
4
(c)The Upturned Microscope
NEGATIVE DATA IS STILL DATA.
PUBLISH IT.
(For everyone's sake)

Registered Report

Go and read this book!

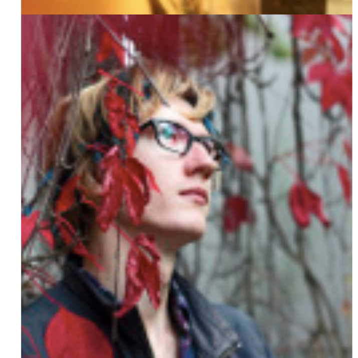


Source: T.W. Hänsch, Nobel lecture


Jaan


Toomas


Liisa


Julius


Sander


Dorian

COMPUTATIONAL NEUROSCIENCE
Research group at the University of Tartu

Thanks!

TARTU ÜLIKOOL · UNIVERSITAS TARTUENSIS · 1632

Luiz Lana

Michael Wibral