Machine Learning and NeuroEngineering

机器学习与神经工程

Lecture 2 – Introduction to Machine Learning

Quanying Liu （刘泉影）

SUSTech, BME department

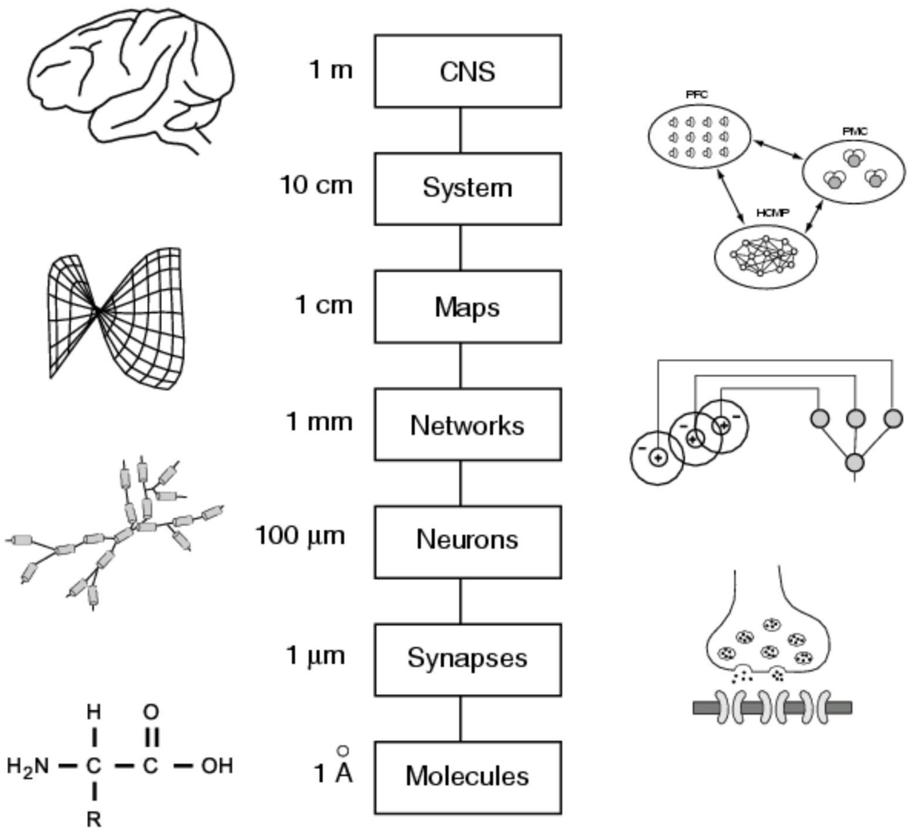
Email: liuqy@sustech.edu.cn

Lecture 1 Recap

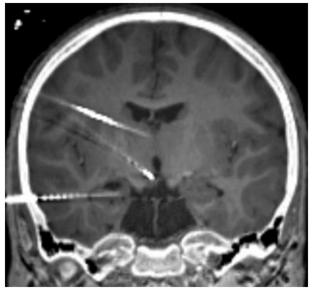
1. The purpose of models is not to fit the data but to sharpen the questions.
2. Some basic neuroimaging techniques
3. Levels: molecular, synapse, neurons, networks, maps, system, CNS

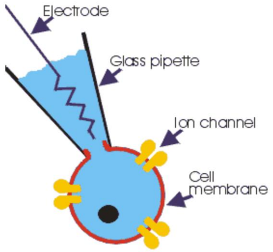
Electrophysiological and Neuroimaging techniques
help to see what is happening in the brain

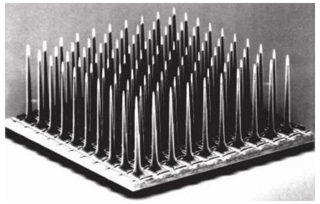
Different levels

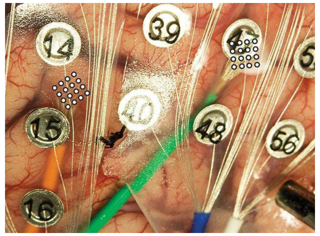


- MRI, fMRI
- EEG, MEG
- SEEG
- ECoG
- Local field potential (extracellular)
- Utah Array
- Patch clamp (intracellular recording)





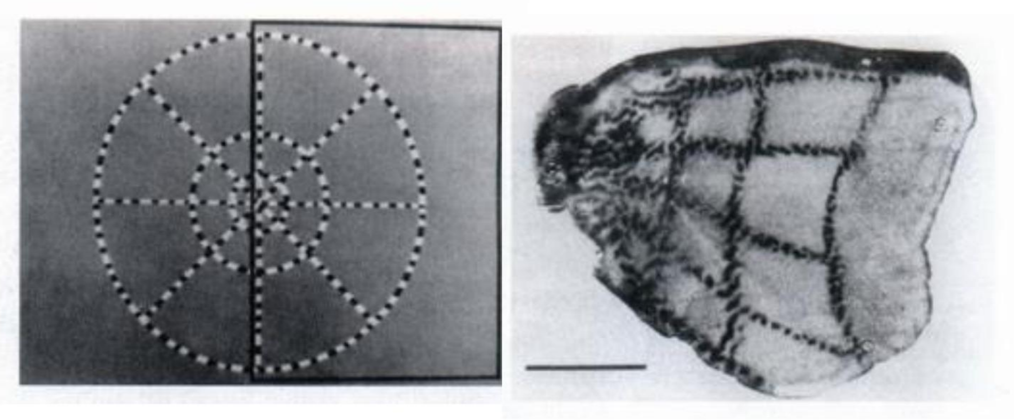




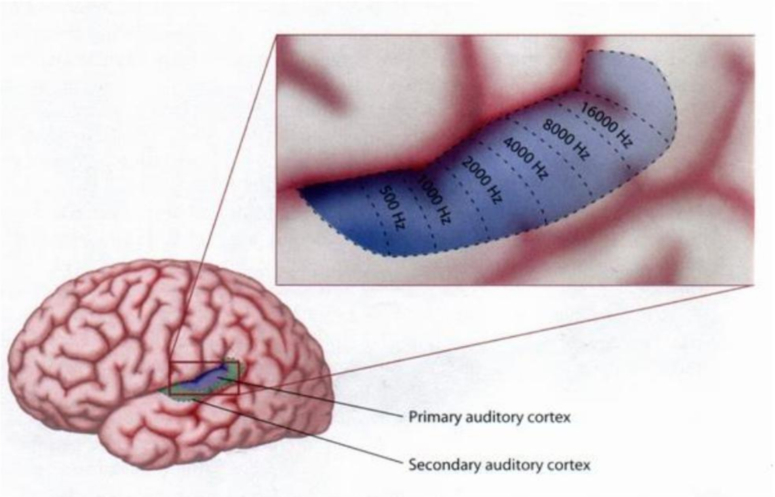
Lecture 1 Recap

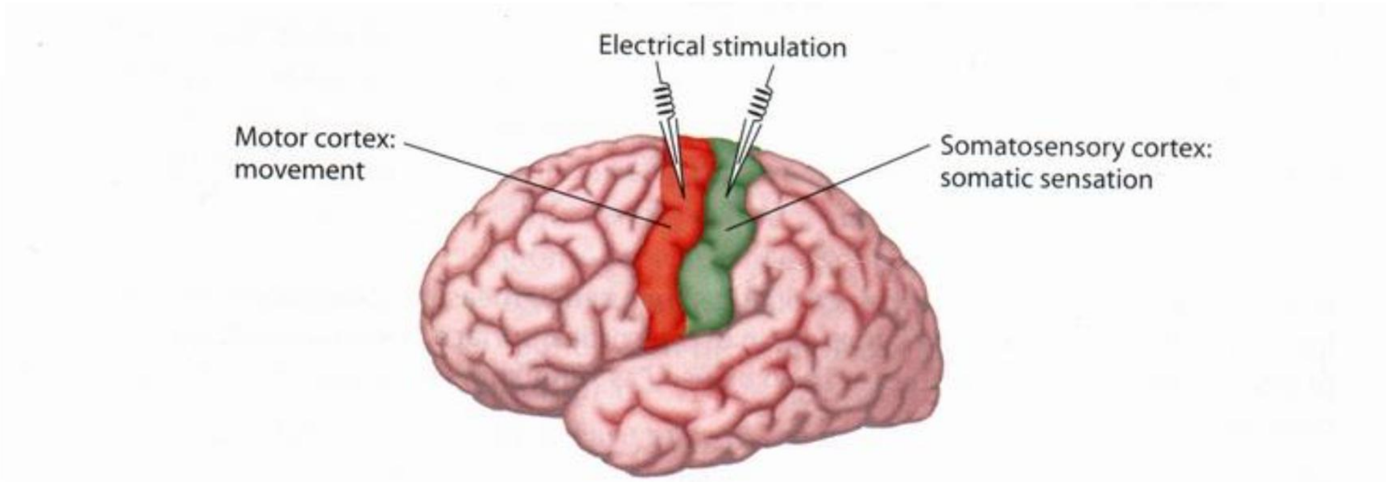
1. The purpose of models is not to fit the data but to sharpen the questions.
2. Some basic neuroimaging techniques
3. Levels: molecular, synapse, neurons, networks, maps, system, CNS
4. Neuron, action potential, Hodgkin-Huxley model
5. Dorsal 'where' pathway and ventral 'what' pathway
6. Topographic Maps at cortex

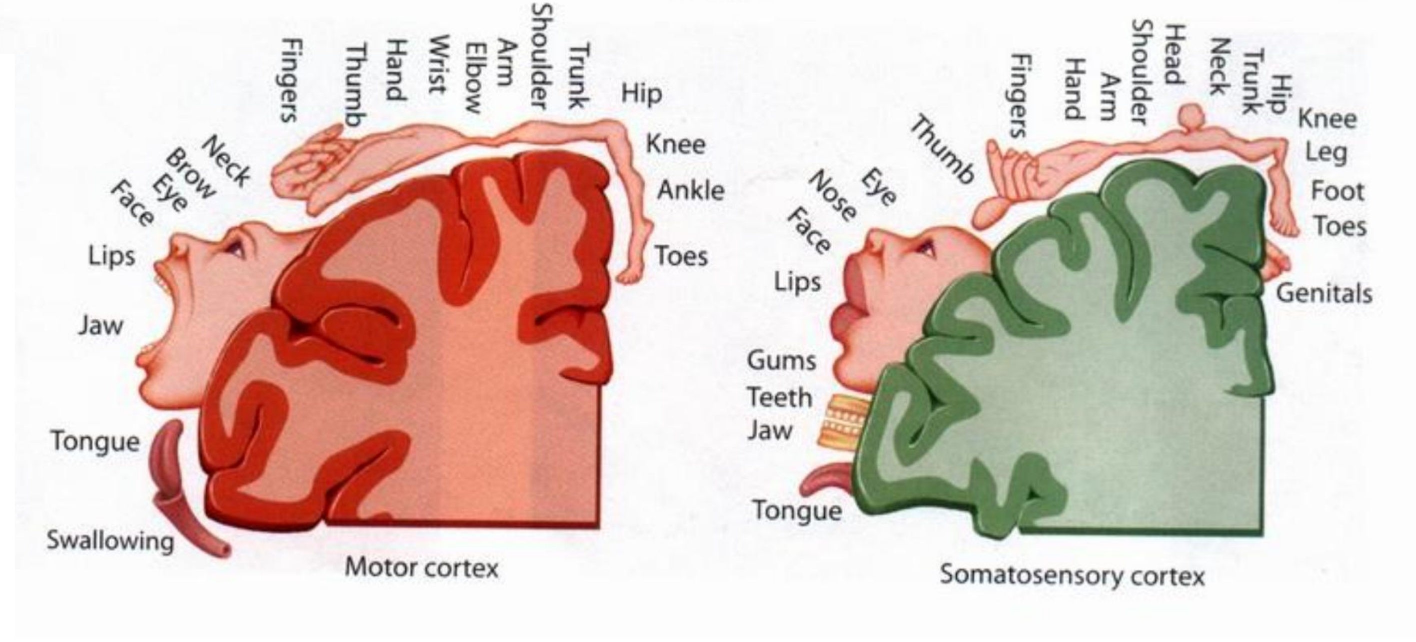
Retinotopic map



Tonotopic map







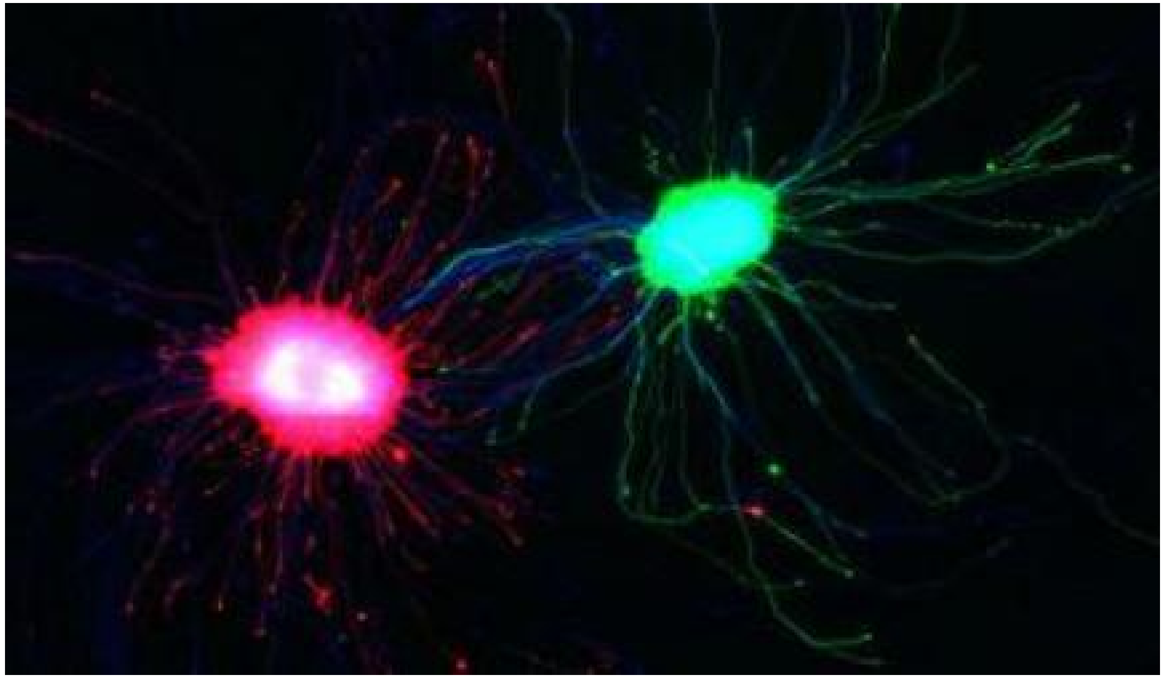
Lecture 1 Recap

1. The purpose of models is not to fit the data but to sharpen the questions.
2. Some basic neuroimaging techniques
3. Levels: molecular, synapse, neurons, networks, maps, system, CNS
4. Neuron, action potential, Hodgkin-Huxley model
5. Dorsal ‘where’ pathway and ventral ‘what’ pathway
6. Topographic Maps at cortex
7. Marco organization of human brain (brain regions and their function)

Some fun facts of brain

True or False?

We have over 80 billion cells in our brain.



True!

We only use around 10% of our brain.

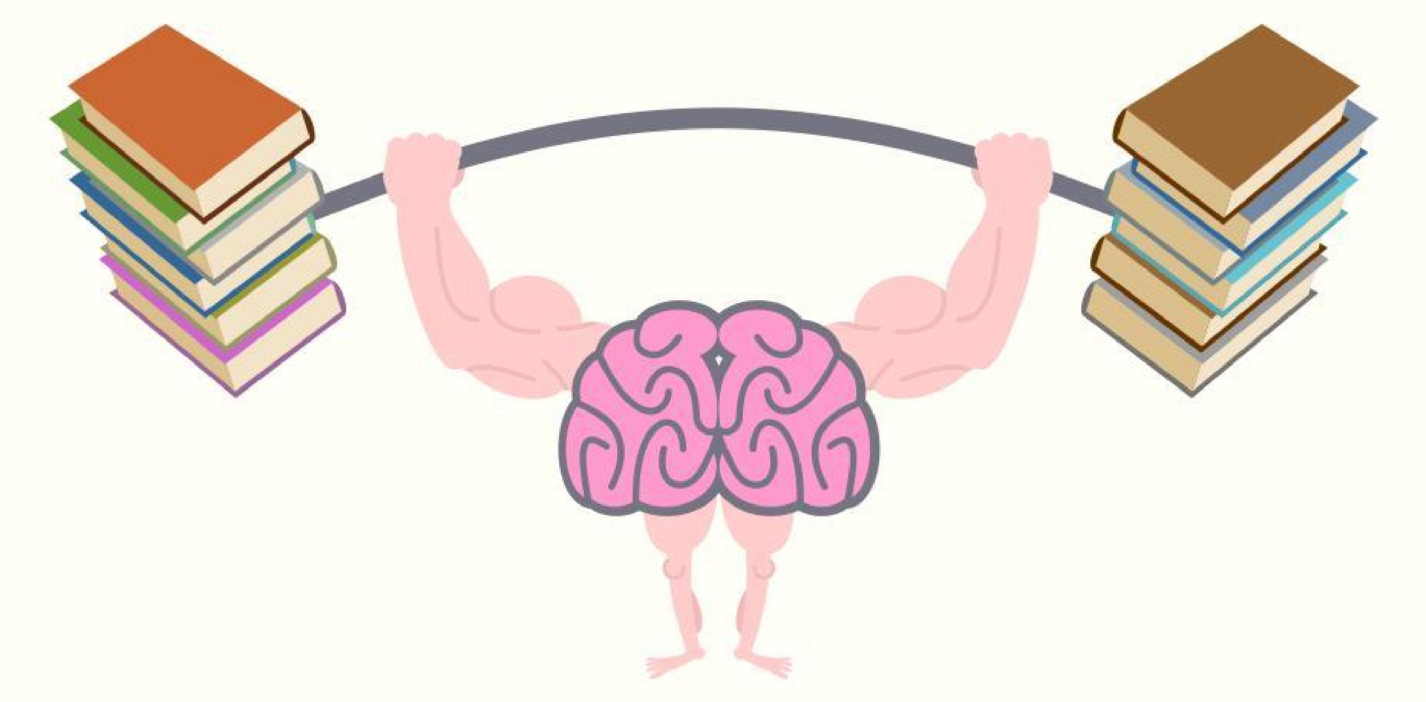
False!

Messages from our brain to our body can travel as fast as 400 km per hour.

True!

Once we reach adult age, our brain's structure does <u>not</u> change anymore.

False!



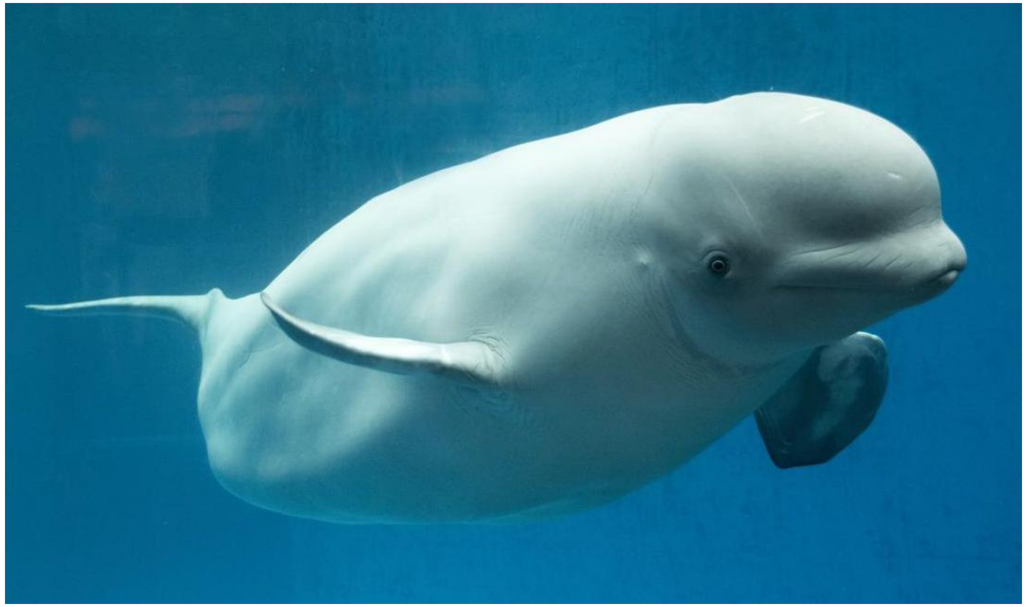
A bigger human brain is a smarter brain.

False!

A *Spermaceti* whale has the largest brain in the world.

True!

A *While* whale does not sleep.



False!

Unihemispheric sleep

The energy used by the brain similar as a 25 watts bulb.

True!

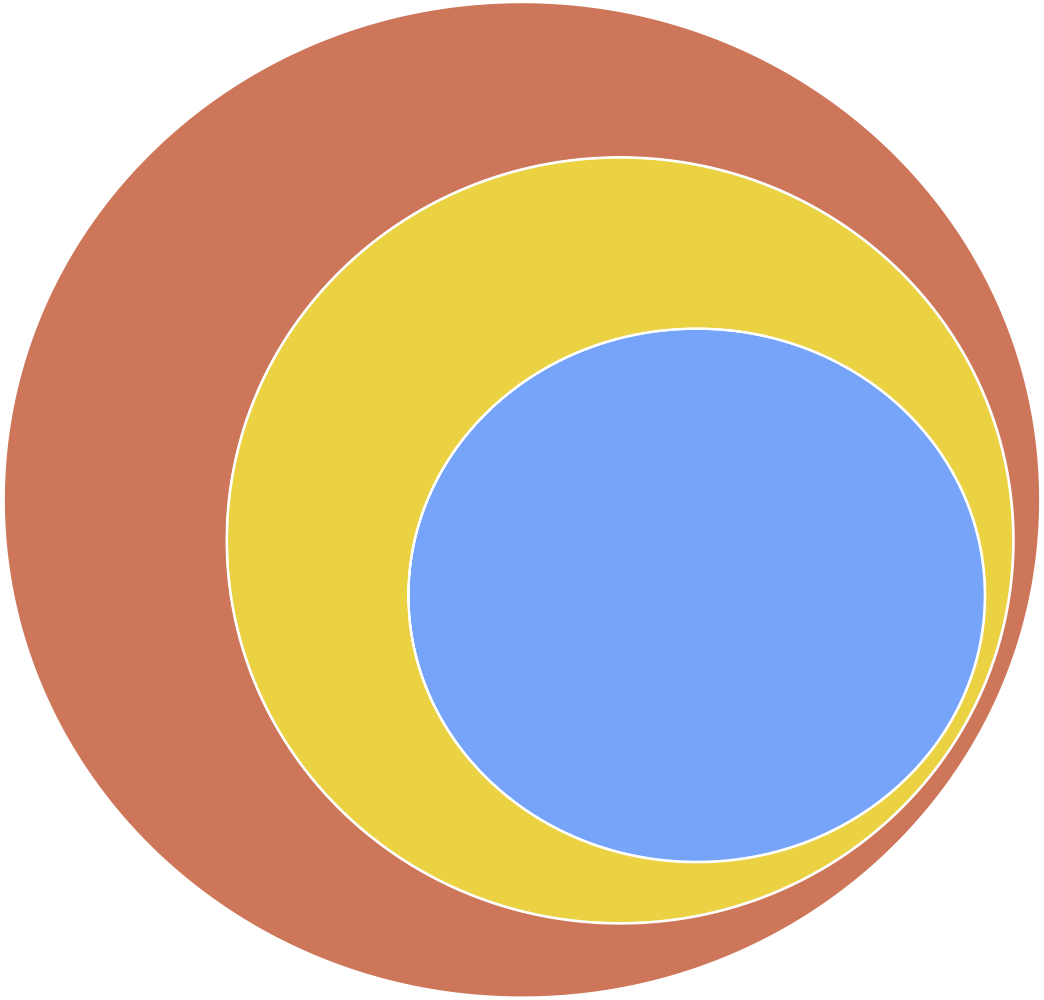
Some people have brains that are more right-sided (artistic) and others more left-sided (logical).

False!

Lecture 2

1. Artificial intelligence > Machine learning > Deep learning
2. Supervised & Unsupervised & Semi-supervised learning

3. Why do we need model? (The 4 points)
4. Generalized Context Model
5. Scope and Falsifiability
6. Summary



Artificial Intelligence

grand project to build non-human intelligence

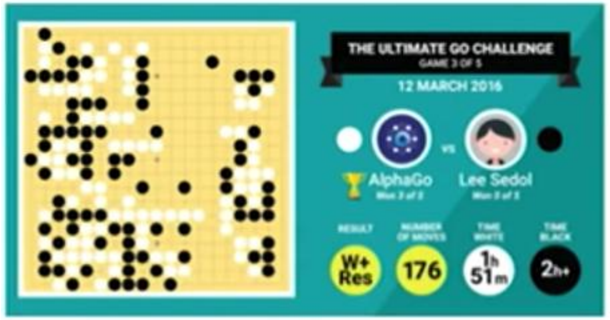
Machine Learning

machines that learn to be smarter

Deep Learning

particular kind of machine learning

Is Machine Learning Everywhere?



AlphaGo

Recommendation systems

Autonomous Vehicles

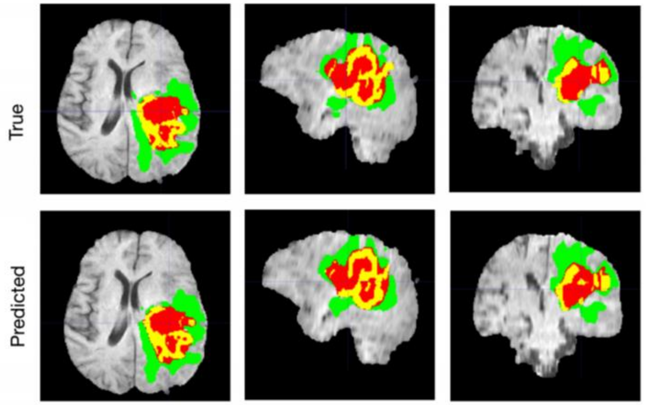
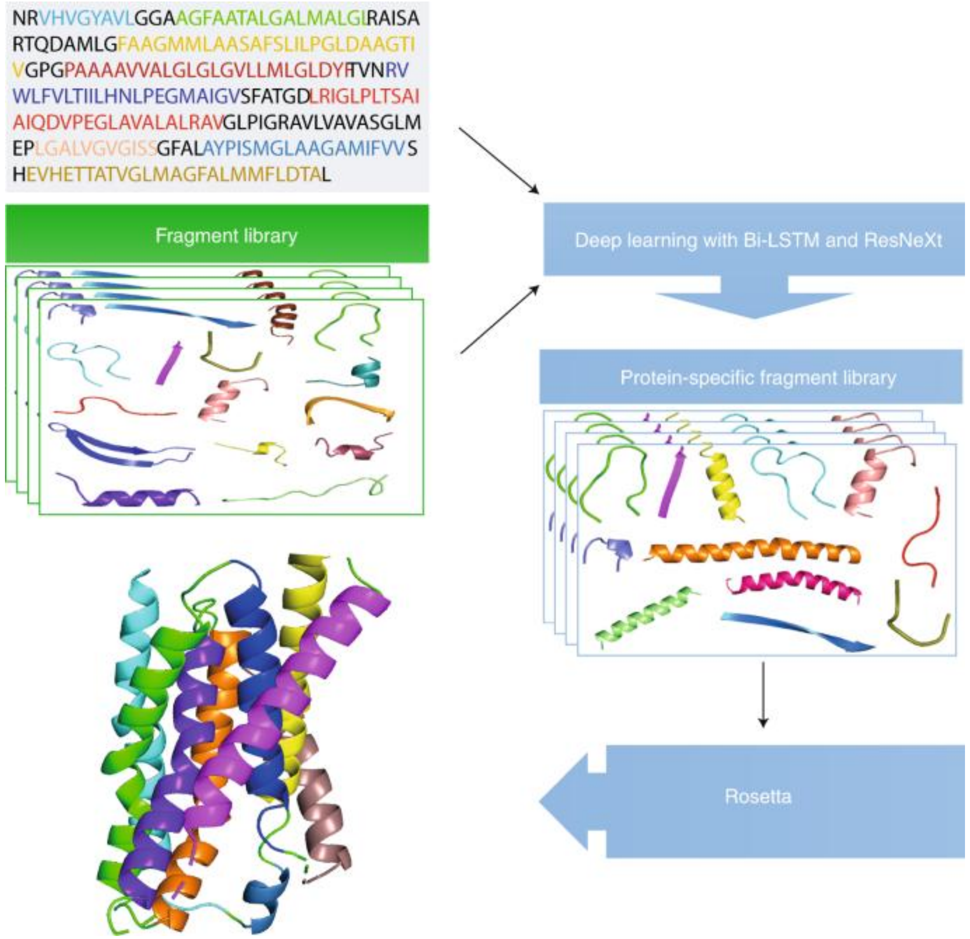


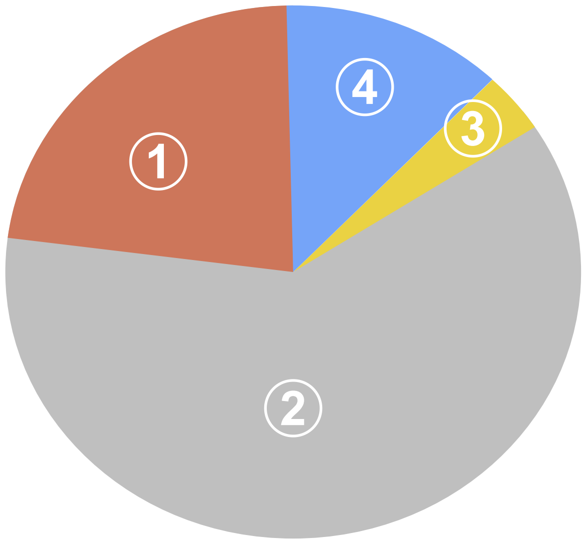
Image processing



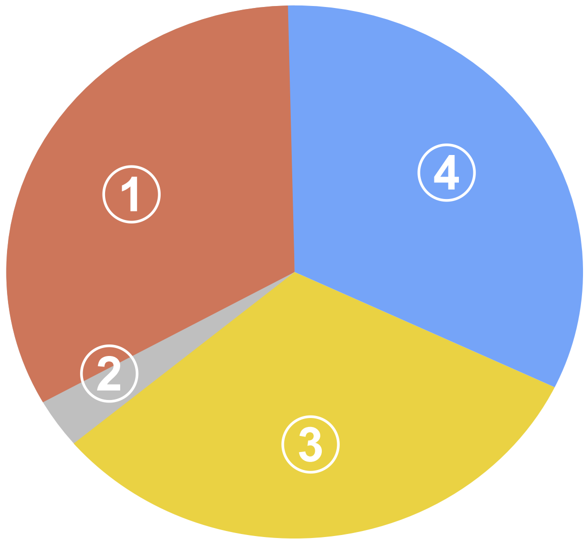
AlphaFold

Rebalances Effort Invested to Model

Traditional Machine Learning

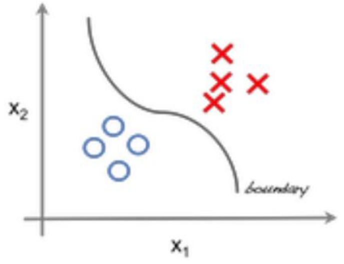


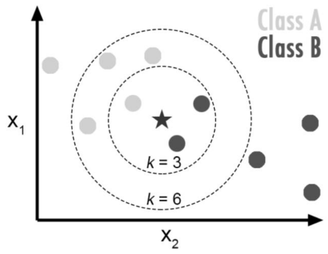
Deep Learning



1. Data preparation
2. Features engineering
3. Model architecture
4. Numerical optimization

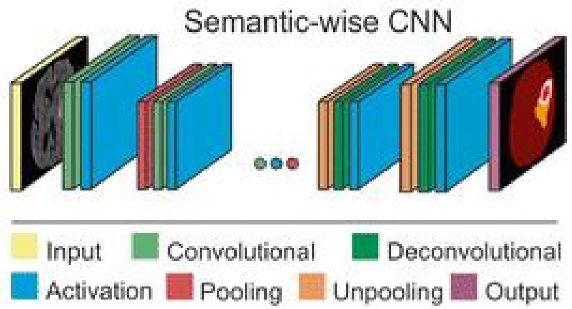
How much data is ”enough”?

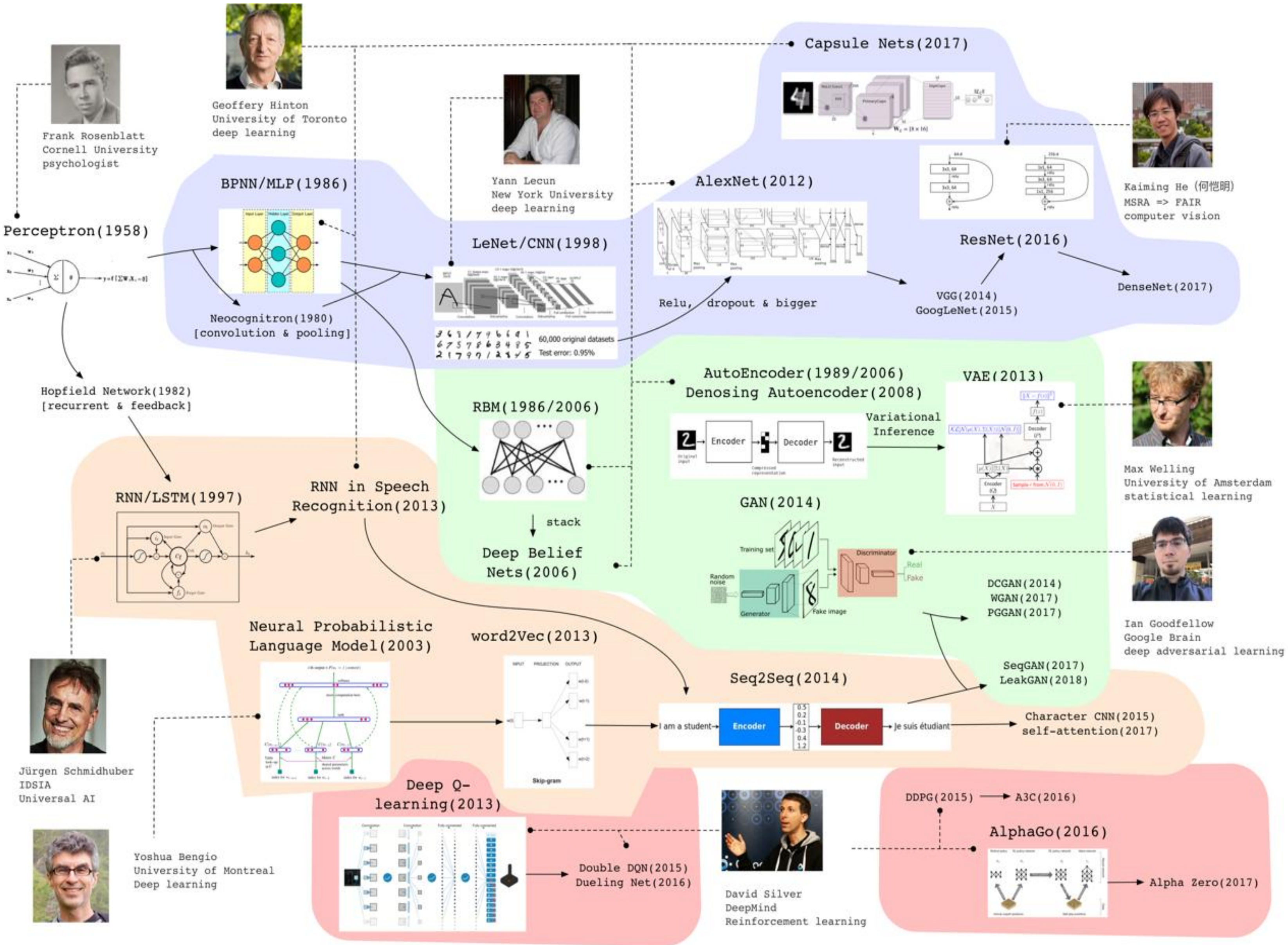




More is better.

General rule: **>5k** per class





Computer Vision and CNN

1979, ***Neocognitron*** by Fukushima
1986, ***Backpropogation MLP*** by Hinton
1998, ***LeNet-5*** by LeCun
2012, ***AlexNet*** by Hinton
2016, ***ResNet*** by 何恺明

Generative Models

1986-2006, ***RBM*** by Hinton
1989/2006, ***AutoEncoder*** by Hinton
2014, ***GAN*** by Goodfellow

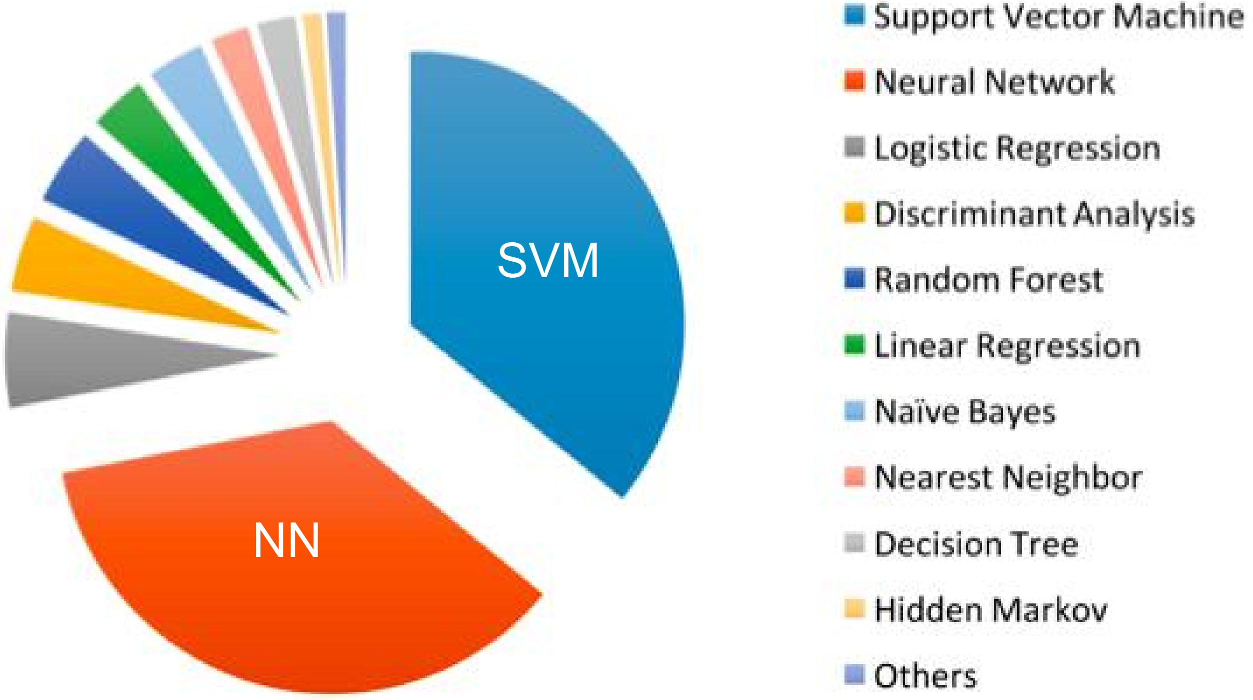
Sequence Models

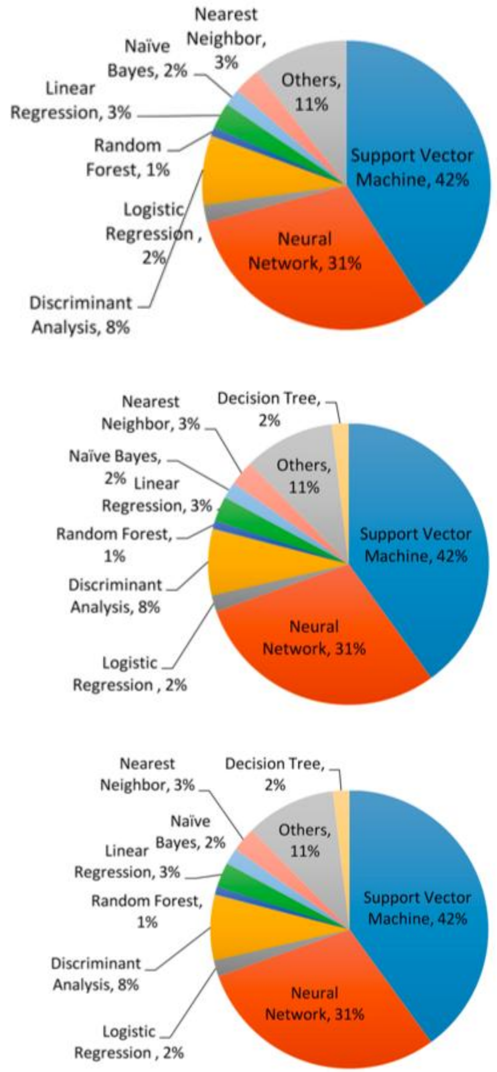
1982, ***Hopfield Network*** by Hopfield
1997, ***LSTM*** by Schmidhuber
2013, ***RNN*** by Hinton

Reinforcement Learning

2013, ***Deep Q-learning*** by Silver
2016, ***AlphaGo*** by Google

The top 10 machine learning algorithms in medical literature





Imaging data

Genetic data

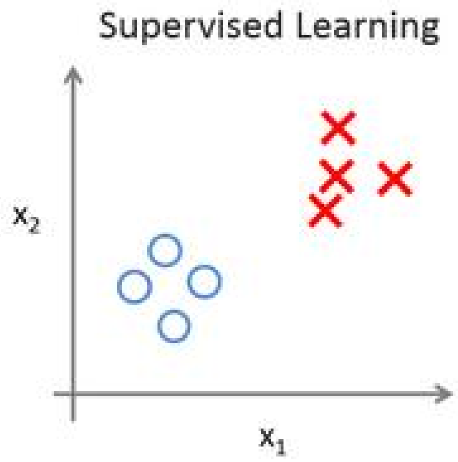
Electrophysiological data

Jiang F, et al. Artificial intelligence in healthcare: past, present and future

Machine learning

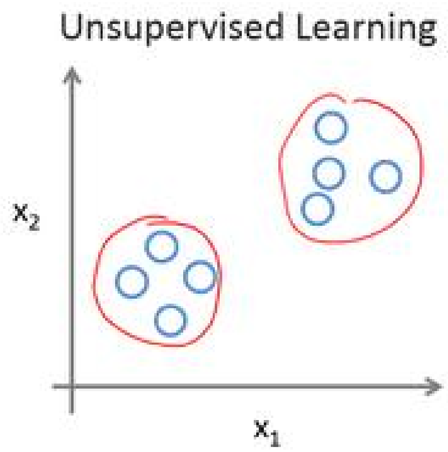
Supervised Learning

Given (data, label) pairs
e.g. classification, regression



Unsupervised Learning

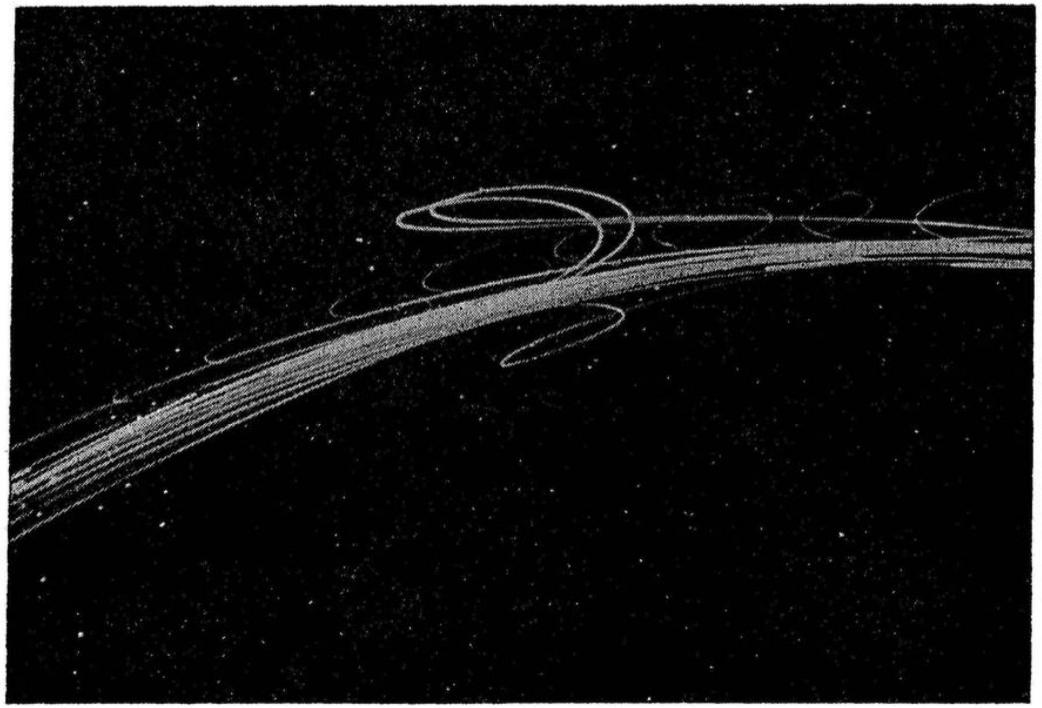
Given data, **no** labels
e.g. clustering, PCA



Semi-supervised Learning

Some labeled data, and some unlabeled data
Transductive learning: unlabeled data is the testing data
Inductive learning: unlabeled data is not the testing data

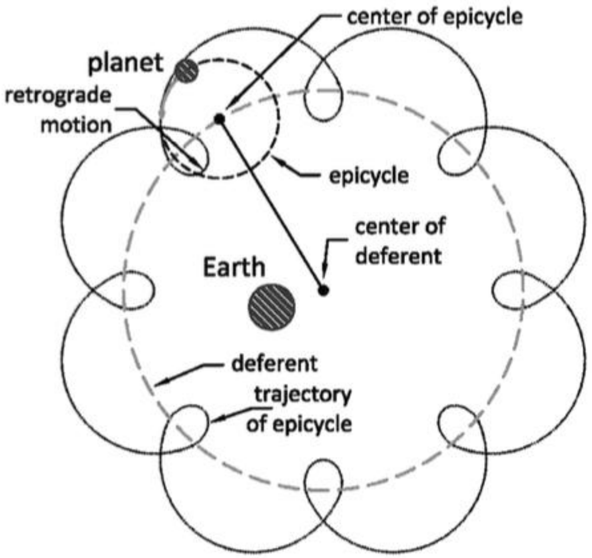
The orbit of the Mars



Ptolemy

Geocentric model

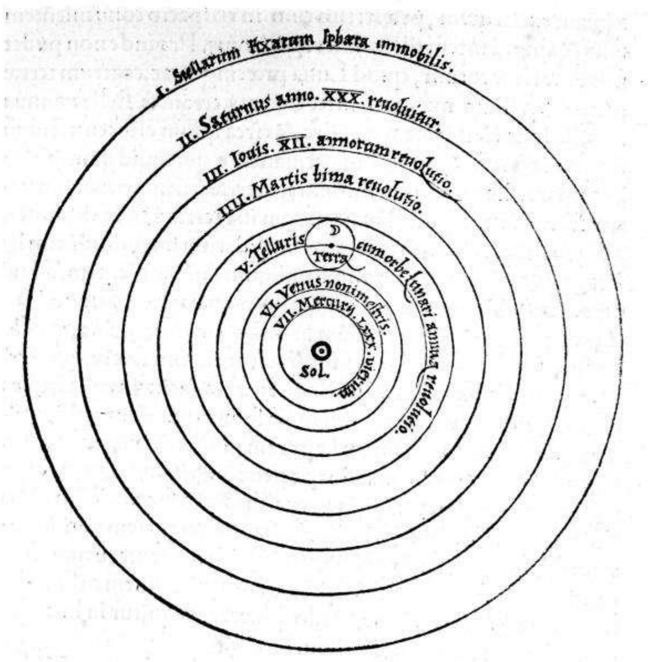
地心说



Copernicus

Heliocentric model

日心说



How to **describe** the orbit?

- verbally
- with mathematical model

How to compare these two models?

- Goodness-of-fit & simplicity
- Intellectual and scholarly judgement

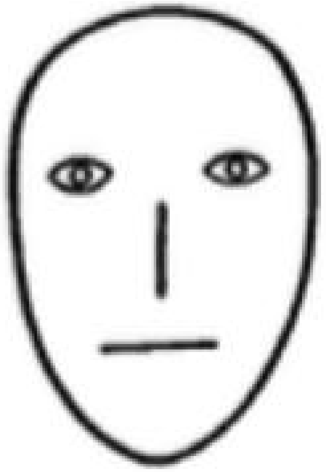
What we've known from this example?

1) Data never speak for themselves. It requires a model to be understood and to be explained.
2) Verbal theorizing alone ultimately cannot substitute for quantitative analysis.
3) There are always several alternative models that vie for explanation of data and we must select among them.
4) Model selection rests on both quantitative evaluation and **intellectual and scholarly judgment**.

Let’s do a cognitive task.

The following training faces are from **two** categories (A and B).

Try to **remember** the faces you have seen, and his category.



A

Let’s do a cognitive task.

The following training faces are from **two** categories (A and B).

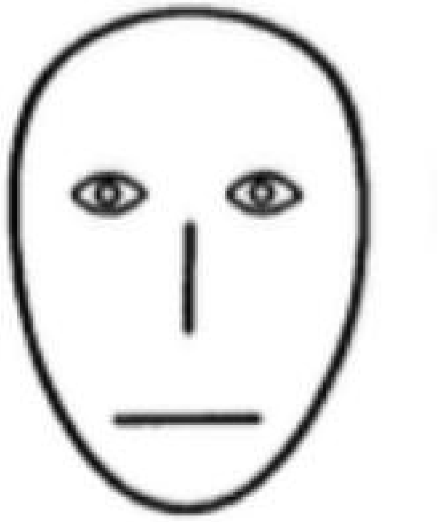
Try to **remember** the faces you have seen, and his category.

B

Let’s do a cognitive task.

The following training faces are from **two** categories (A and B).

Try to **remember** the faces you have seen, and his category.



A

Let’s do a cognitive task.

The following training faces are from **two** categories (A and B).

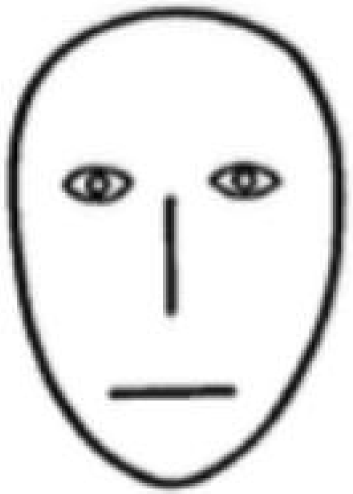
Try to **remember** the faces you have seen, and his category.

B

Now tell me:

1) the category of this face, and the classification confidence (from 0 to 1)

2) whether you have seen it before (recognition), and the probability



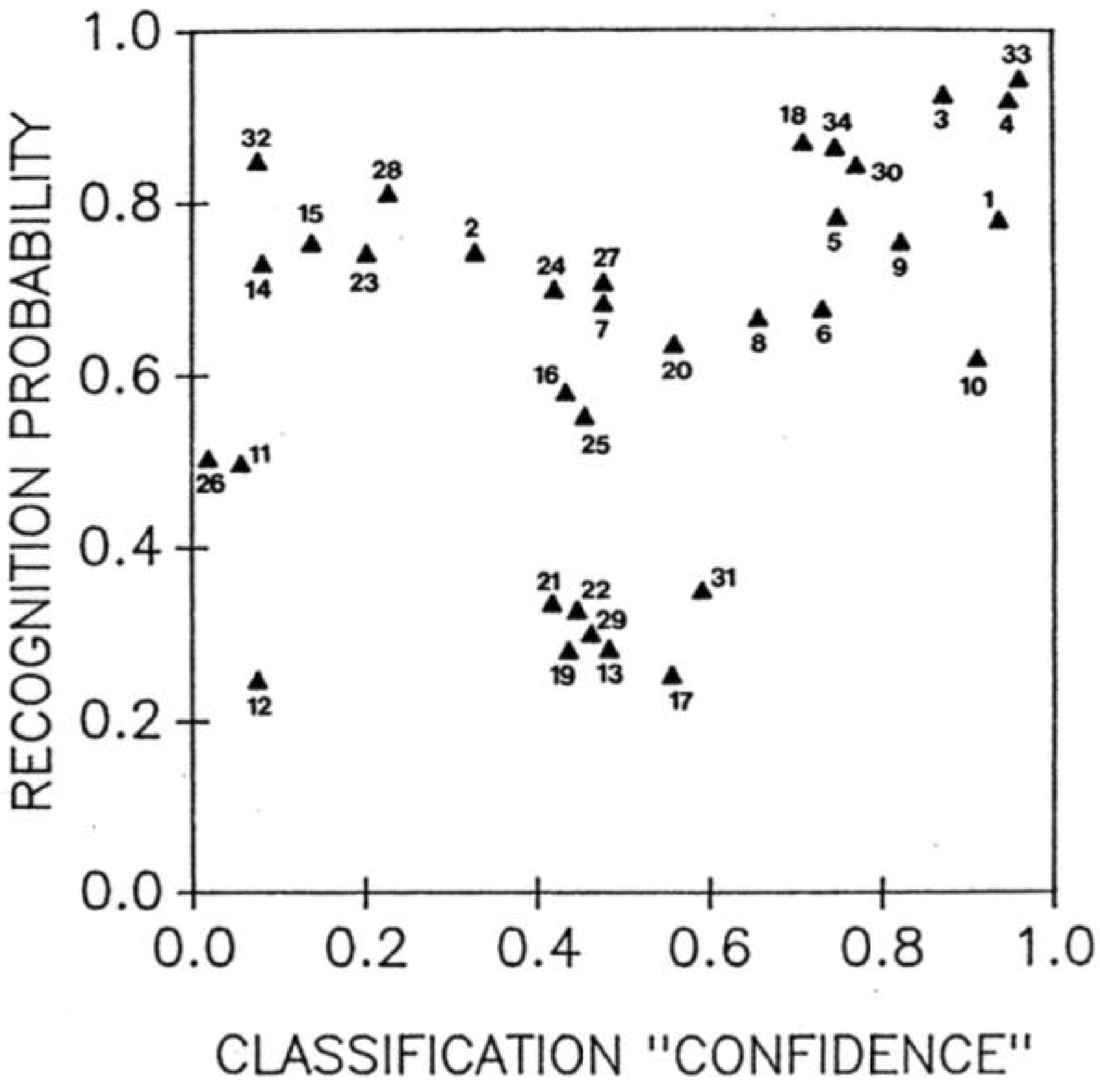
A

Now tell me:

1) the category of this face, and the classification confidence (from 0 to 1)

2) whether you have seen it before (recognition), and the probability

B



Any relationship between *perceptual classification* and *recognition memory*?

Models to study relationship:

- Correlation
- Linear regression
- Fitting a specific function
- Regression with neural network
- …

Nosofsky, R. M. (1991). Tests of an exemplar mode for relating perceptual classification and recognition memory, *Journal of Experimental Psychology*

Two types of models

1. models that simply **describe data**

   Neural Networks, and deep learning…

   Very often to see people just use a model without thinking too much

2. models that **explain the underlying process**

   (e.g. physics, chemistry, cognition, nature…)

   Heliocentric model for astronomy

   Dynamical models for chemical process

   <u>Generalized Context Model</u> for our cognitive task

Generalized Context Model (GCM)

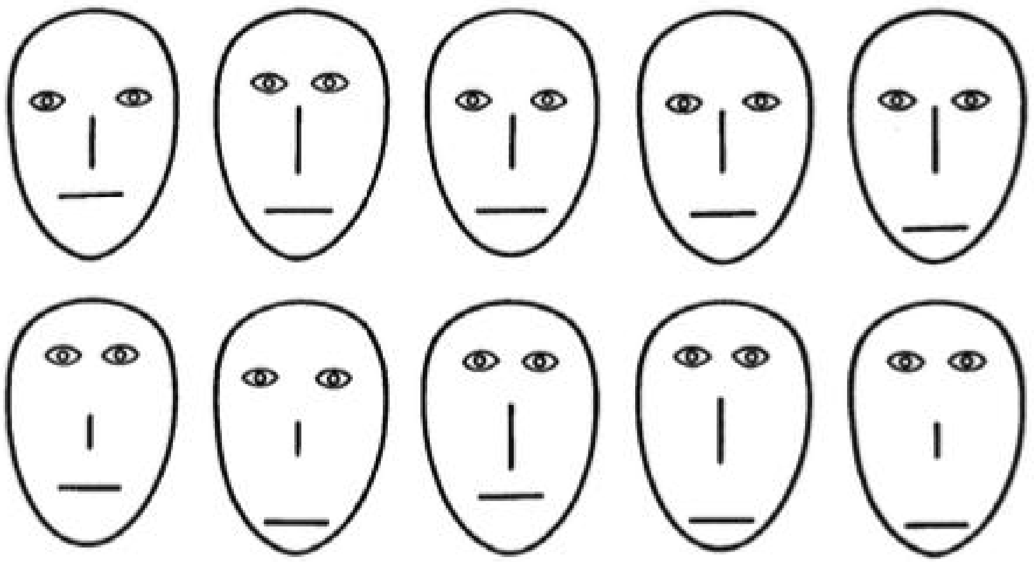
GCM stores every category exemplar encountered during training in memory.

In GCM, the classification procedure would be implemented by adding each stimulus to the pile of faces belonging to the same category.

Remember that each response during training is followed by feedback, so people know whether a face belongs to A or B at the end of each trial.

Following training, GCM has thus built two sets of exemplars, one for each category, and all subsequent test stimuli are classified by referring to those memorized ensembles.

Generalized Context Model



The **distance** between two faces (i&j):

$$d_{ij} = \left(\sum_{k=1}^{K} |x_{ik} - x_{jk}|^2 \right)^{\frac{1}{2}}$$

The **similarity** between two faces:

$$s_{ij} = exp(-c \cdot d_{ij})$$

Decompose a face into 4 dimensions of features:

1. eye height
2. eye separation
3. nose length
4. mouth height

$x_{i\,k}$

i: the index of face

k: the index of the feature

Response probabilities

$$P(R_i = A|i) = \frac{\left(\sum_{j\in A} s_{ij}\right)}{\left(\sum_{j\in A} s_{ij}\right) + \left(\sum_{j\in B} s_{ij}\right)}$$

GCM: Step-by-step

$x_{i\ k}$
i: the index of face
k: the index of the feature

1. The **distance** between two faces (i&j):

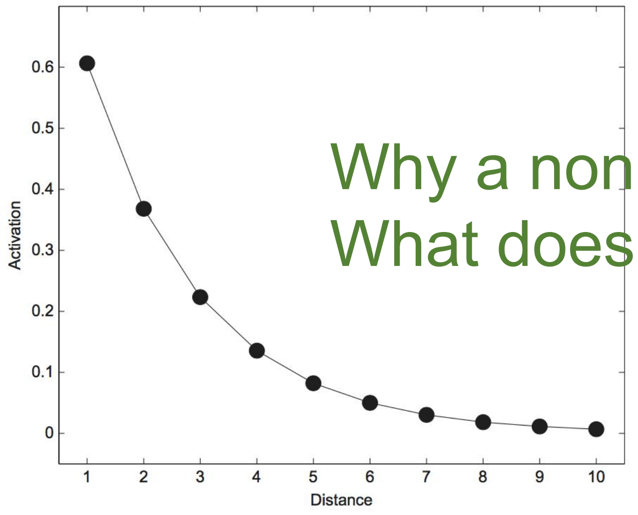
$$d_{ij} = \left(\sum_{k=1}^{K} |x_{ik} - x_{jk}|^2 \right)^{\frac{1}{2}}$$

2. The **similarity** between two faces:

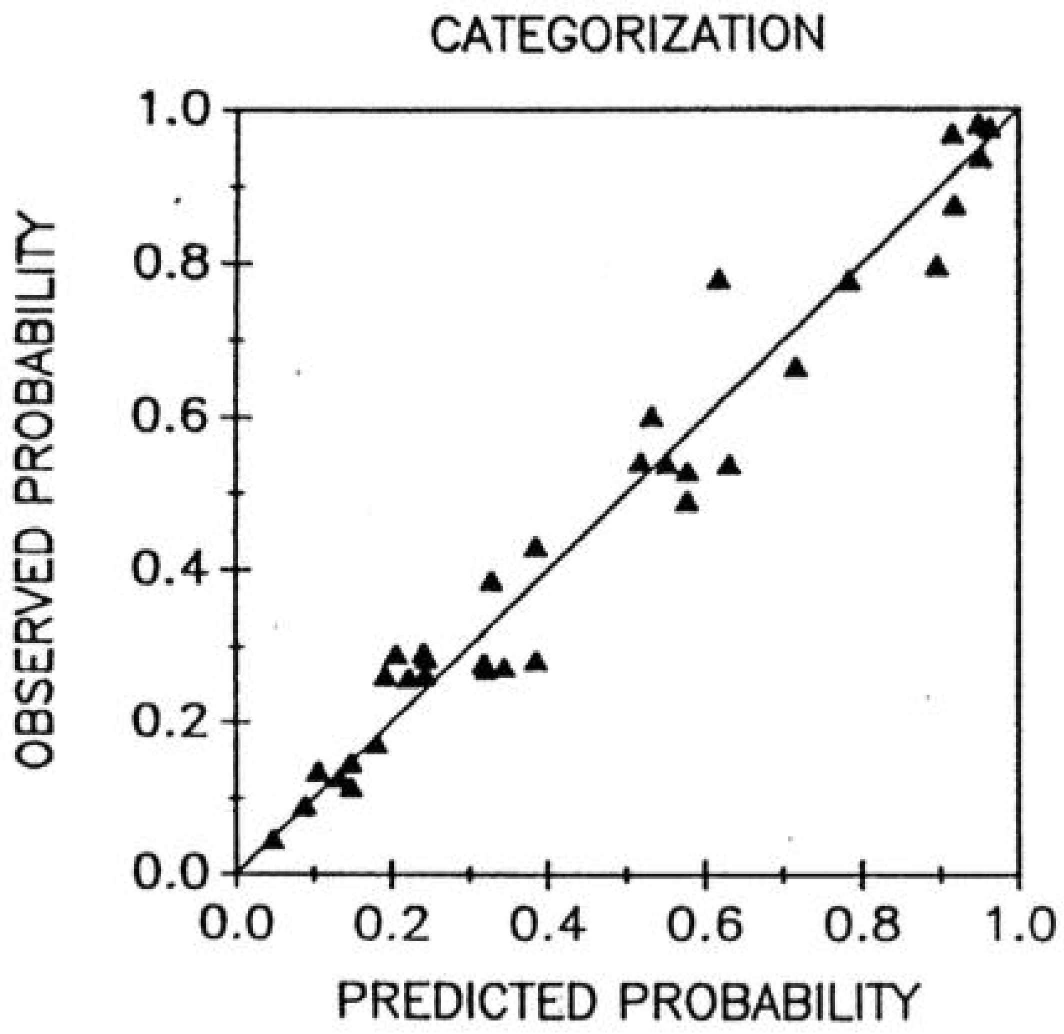
$$s_{ij} = exp(-c \cdot d_{ij})$$

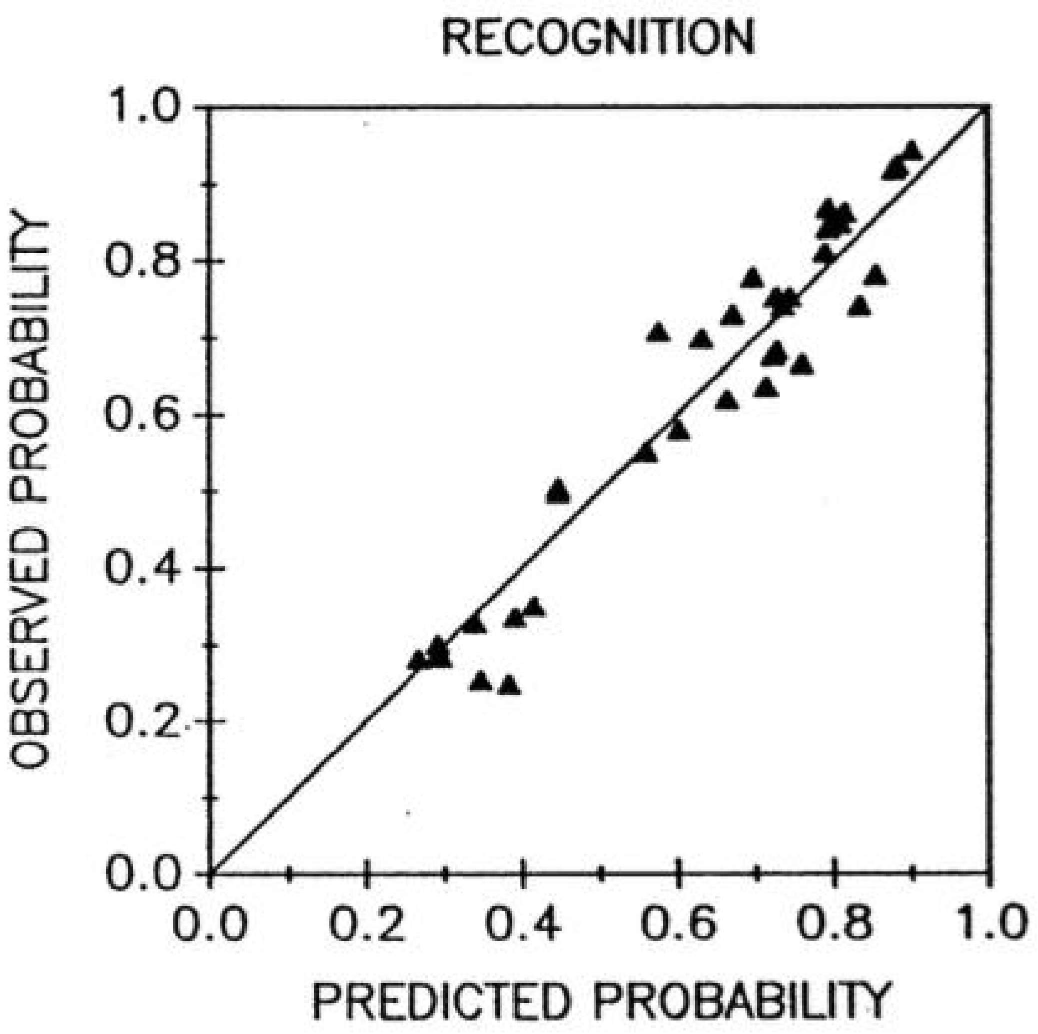
3. Response probabilities (to choose category A)

$$P(R_i = A|i) = \frac{\left(\sum_{j \in A} s_{ij} \right)}{\left(\sum_{j \in A} s_{ij} \right) + \left(\sum_{j \in B} s_{ij} \right)}$$



Why a nonlinear function?
What does it mean?

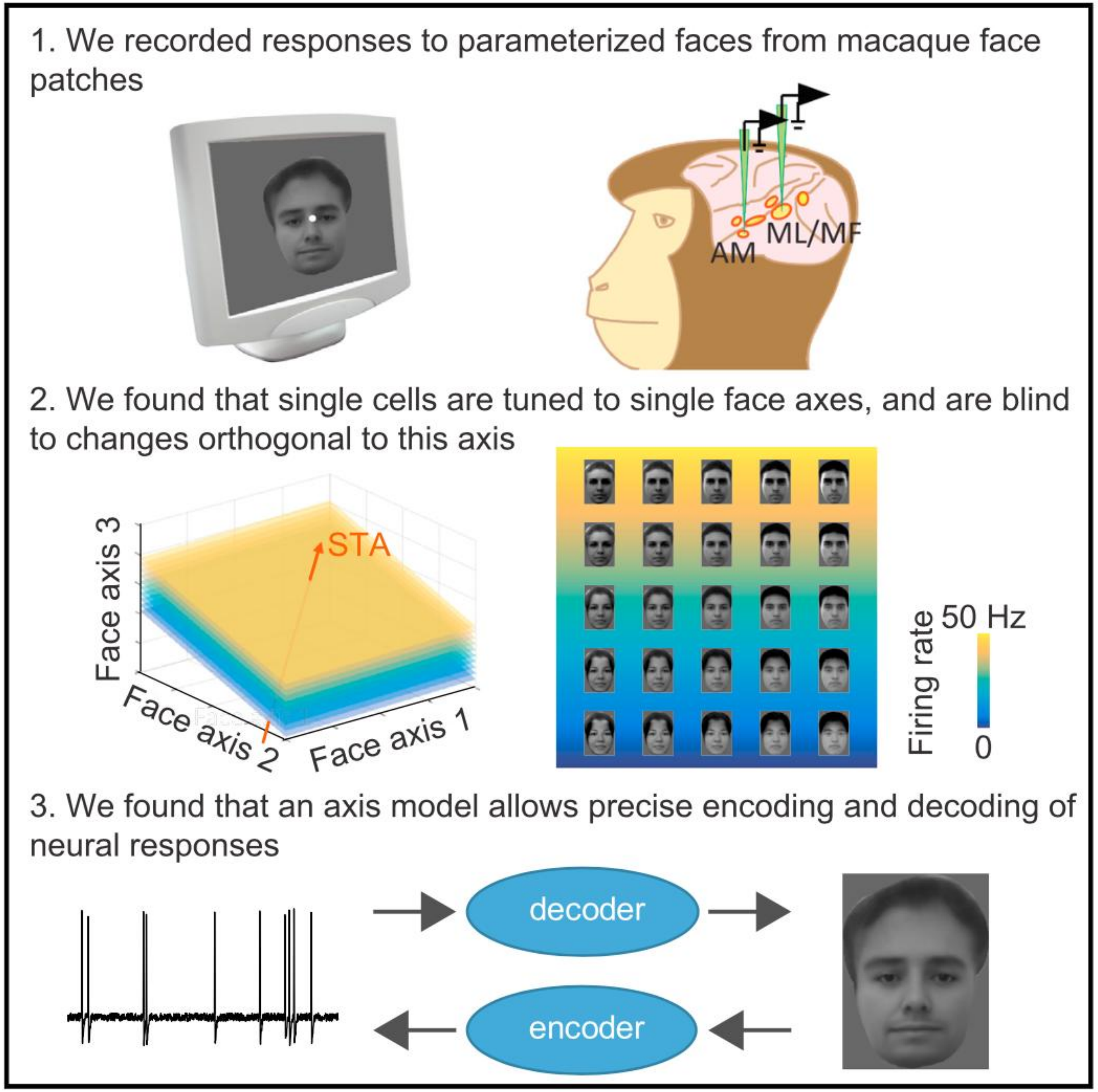




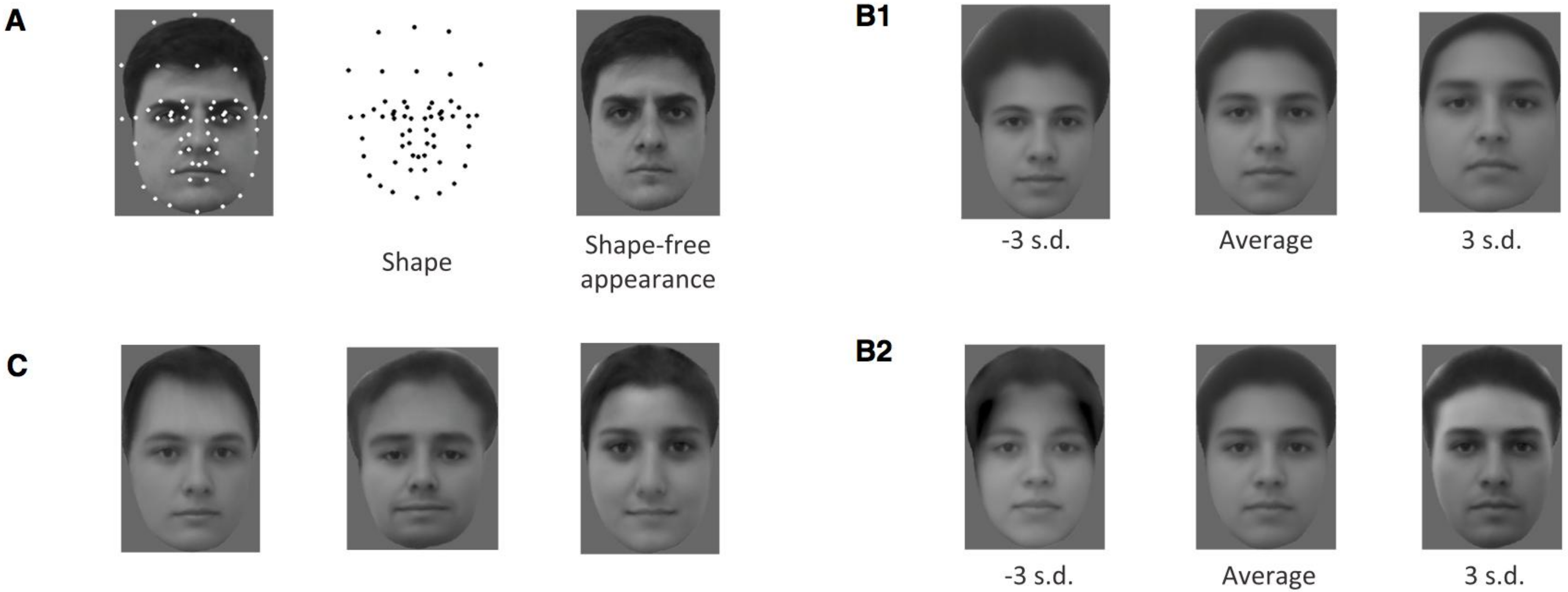
The perceptual classification and recognition memory have a unified account of cognitive processes.

What's next?

- **Other ways to decompose** the face features?
  - PCA? (principle component decomposition)
  - ICA? (Independent component decomposition)
- Can GCM generalize to **real faces? Naturalistic stimuli?**
  - The answer is yes.
  - But you can design experiments to study it.
- Where is **the brain region** to store the information of the category exemplar encountered?
- Where is **the brain region to** calculate x_{ik}, d_{ij}, s_{ij} and $P(i)$.



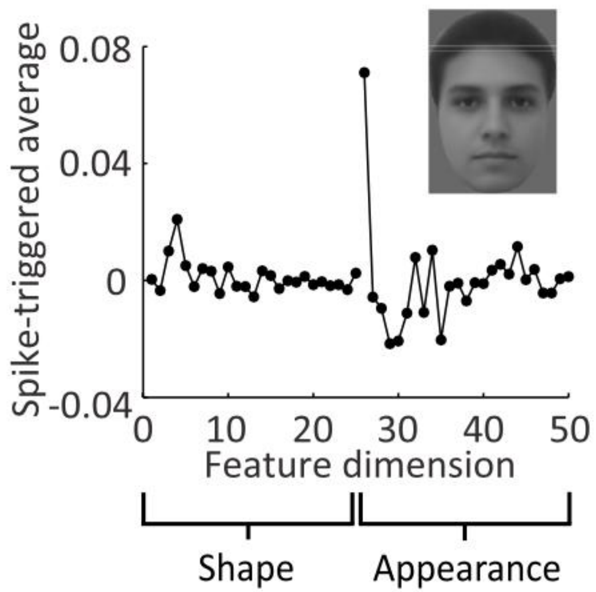
Le Chang and Doris Tsao. (2017), The Code for Facial Identity in the Primate Brain, Cell

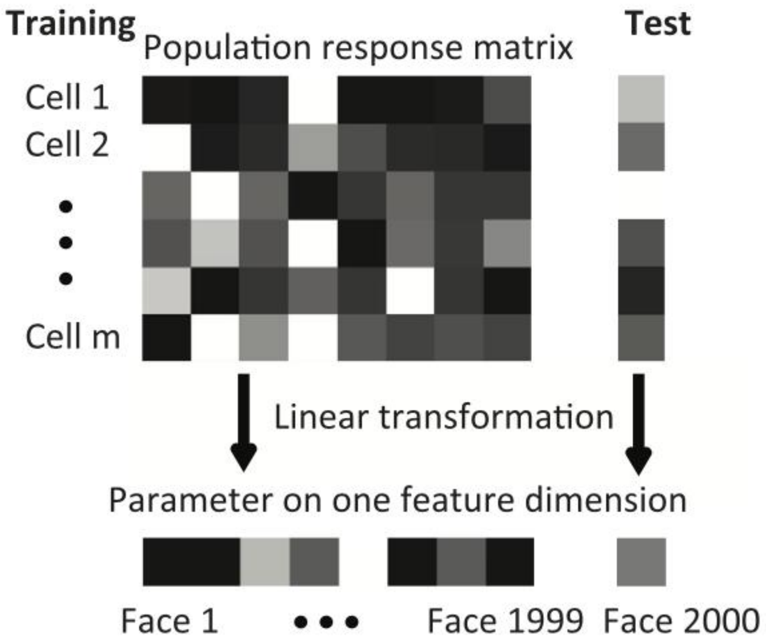


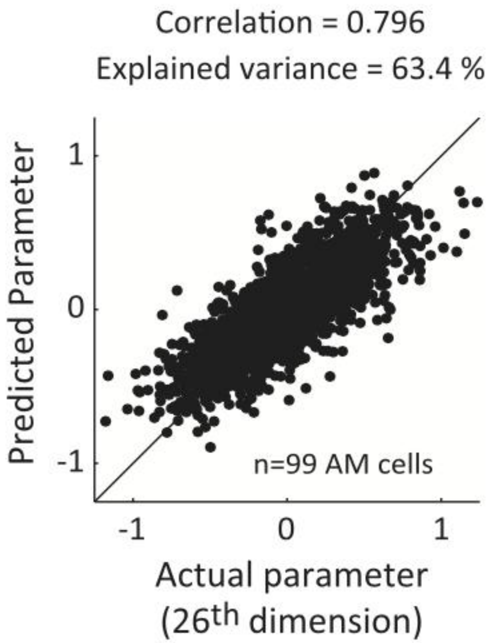
(A) 58 landmark points were labeled on 200 facial images from a face database (FEI face database, left). The positions of these landmarks carry shape information about each facial image (middle). The landmarks were smoothly morphed to match the average landmark positions in the 200 faces, generating an image carrying shape-free appearance information about each face (right).
(B) **PCA** was performed to extract the feature dimensions that account for the largest variability in the database. The first principal component for shape (B1) and appearance (B2) are shown.
(C) Example face stimuli generated by randomly drawing from a face space constructed by the first 25 shape PCs and first 25 appearance PCs.

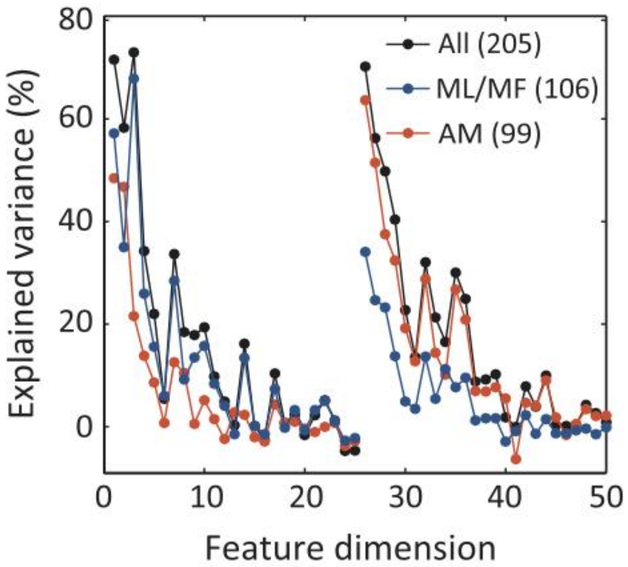
Le Chang and Doris Tsao. (2017), The Code for Facial Identity in the Primate Brain, Cell

D


Spike-triggered average
0.08
0.04
0
-0.04
0
10
20
30
40
50
Feature dimension
Shape
Appearance


Training
Population response matrix
Test
Cell 1
Cell 2
Cell m
Linear transformation
Parameter on one feature dimension
Face 1
Face 1999
Face 2000


Correlation = 0.796
Explained variance = 63.4 %
Predicted Parameter
1
0
-1
n=99 AM cells
-1
0
1
Actual parameter
(26th dimension)


Explained variance (%)
80
60
40
20
0
All (205)
ML/MF (106)
AM (99)
0
10
20
30
40
50
Feature dimension

The process for course project

Ask a question

Make observations and ask questions that can be tested through scientific methods.

Make a hypothesis

Based on observation and previous research, make a prediction or explanation that can be tested with scientific methods.

Conduct an experiment

Select the methodology and tools that can be used to test your hypothesis and design an experiment.

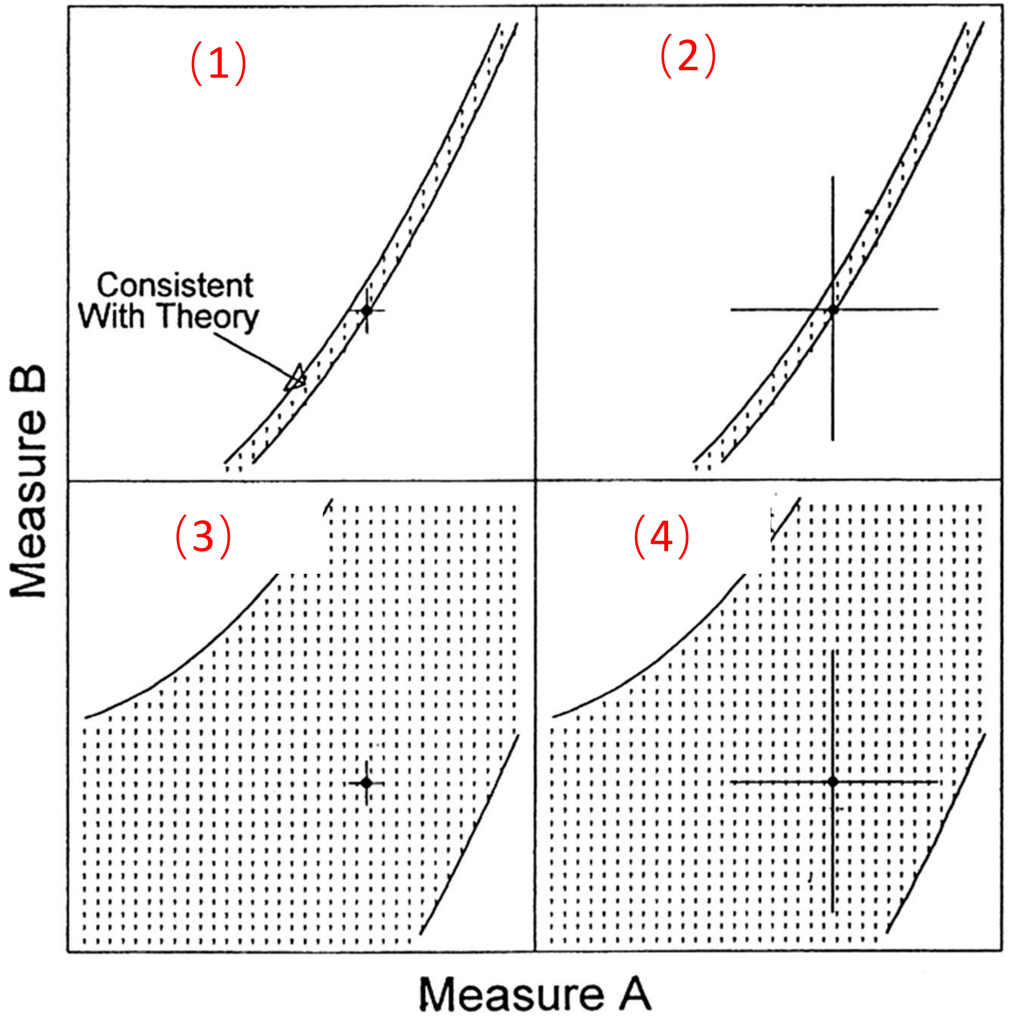
Analyze data

Analyze the data collected through your experiment.

Draw conclusions

Summarize the data collected and form conclusions that relate to your hypothesis.
Write a report.

Scope and Falsifiability



4 models:

The shadowed region is the zone of prediction from model.
The dot with error bar is the real data.

The real data supports which model the most?

(1)

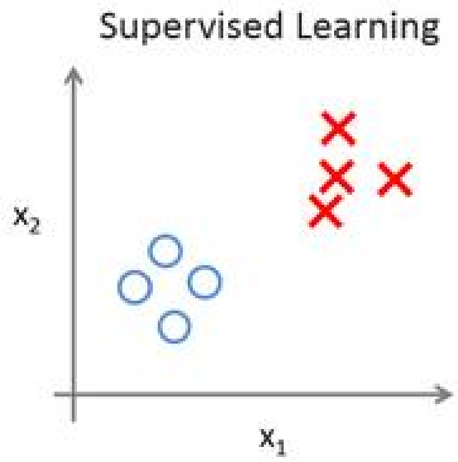
Which two have better quality of data?

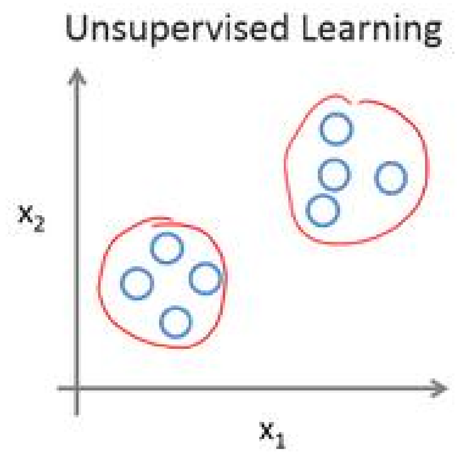
(1 and 3)

Summary of Lecture 2

1. Artificial intelligence > Machine learning > Deep learning
2. Supervised & Unsupervised & Semi-supervised learning

3. Why do we need model? (The 4 points)



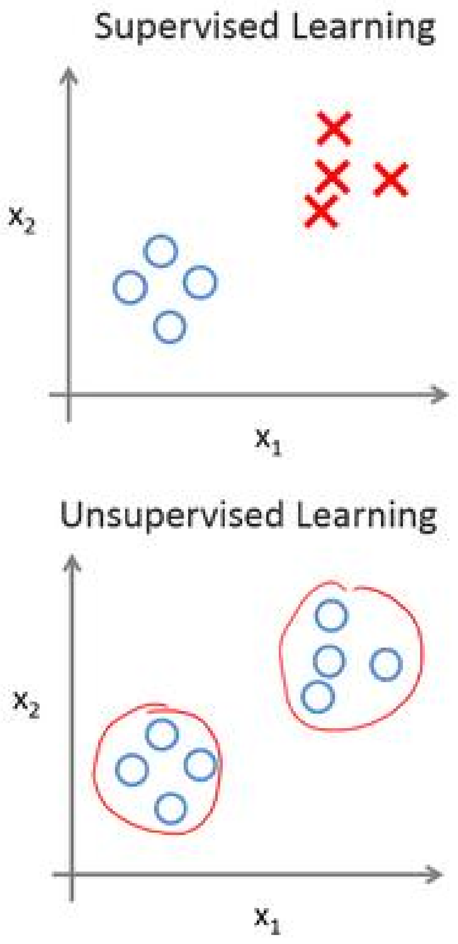


What we've known from this example?

1) Data never speak for themselves. It requires a model to be understood and to be explained.
2) Verbal theorizing alone ultimately cannot substitute for quantitative analysis.
3) There are always several alternative models that vie for explanation of data and we must select among them.
4) Model selection rests on both quantitative evaluation and **intellectual and scholarly judgment**.

Summary of Lecture 2

1. Artificial intelligence > Machine learning > Deep learning
2. Supervised & Unsupervised & Semi-supervised learning
3. Why do we need model? (The 4 points)
4. Generalized Context Model



GCM: Step-by-step

$x_{i\,k}$
i: the index of face
k: the index of the feature

1. The **distance** between two faces (i&j):

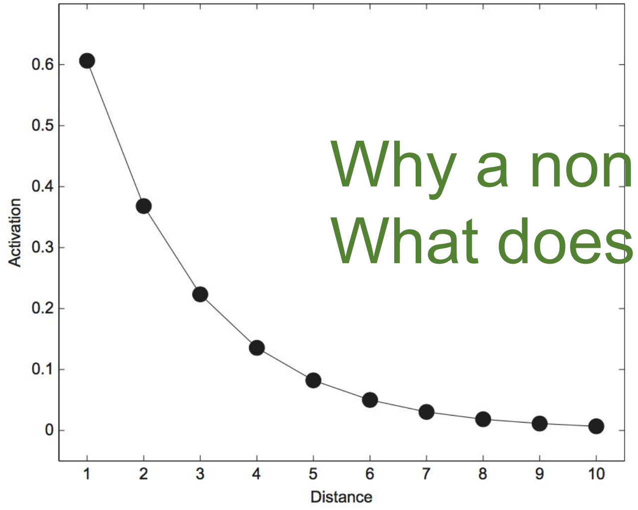
$$d_{ij} = \left(\sum_{k=1}^{K} |x_{ik} - x_{jk}|^2 \right)^{\frac{1}{2}}$$

2. The **similarity** between two faces:

$$s_{ij} = exp(-c \cdot d_{ij})$$

3. Response probabilities (to choose category A)

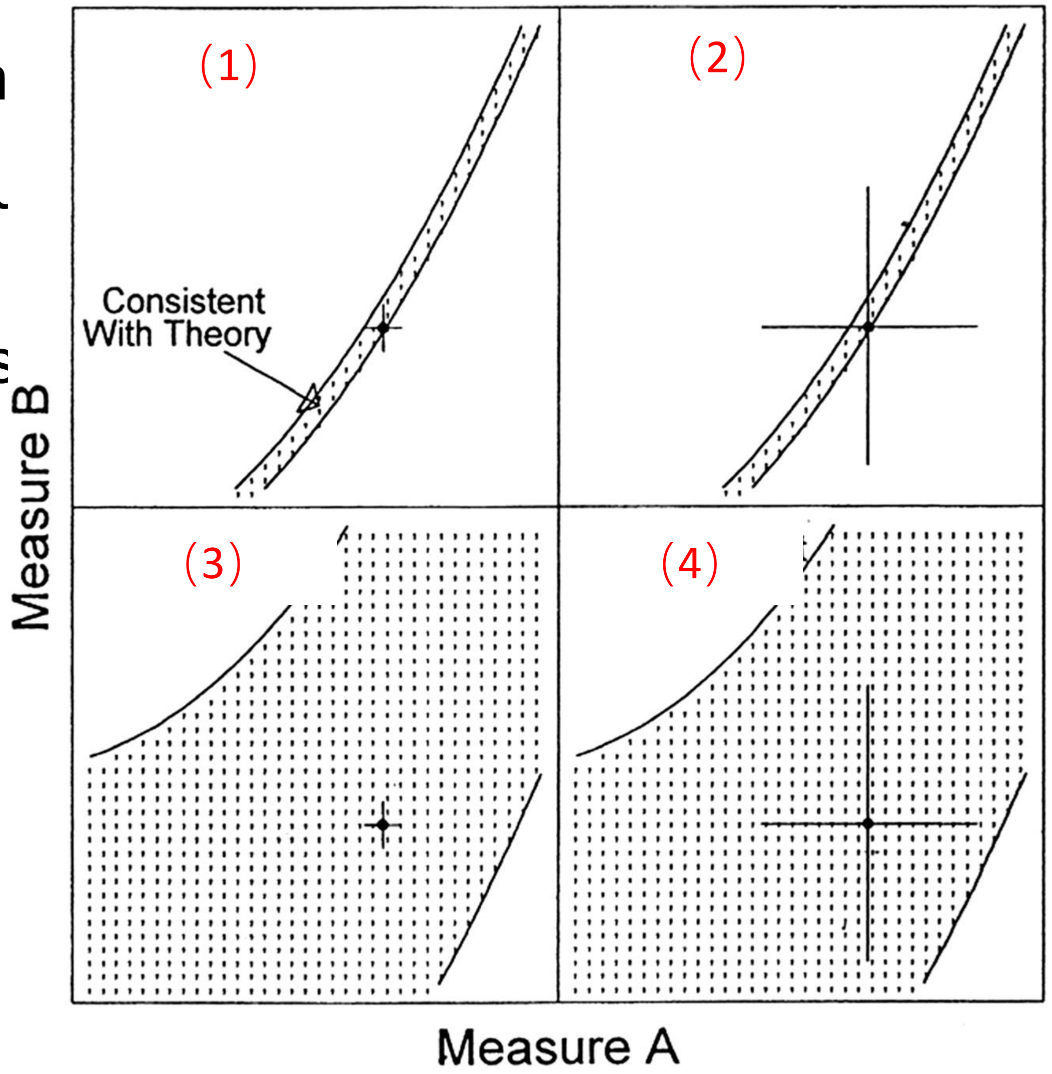
$$P(R_i = A|i) = \frac{\left(\sum_{j \in A} s_{ij}\right)}{\left(\sum_{j \in A} s_{ij}\right) + \left(\sum_{j \in B} s_{ij}\right)}$$



Why a nonlinear function?
What does it mean?

Summary of Lecture 2

1. Artificial intelligence > Machine learnin
2. Supervised & Unsupervised & Semi-su
3. Why do we need model? (The 4 points
4. Generalized Context Model
5. Scope and Falsifiability



Recommended materials

Textbook

- Computational Modeling of Cognition and Behavior, Chapter 1

Must read.

Research Paper

- Le Chang and Doris Tsao. (2017), The Code for Facial Identity in the Primate Brain, Cell

Not obliged. For fun.

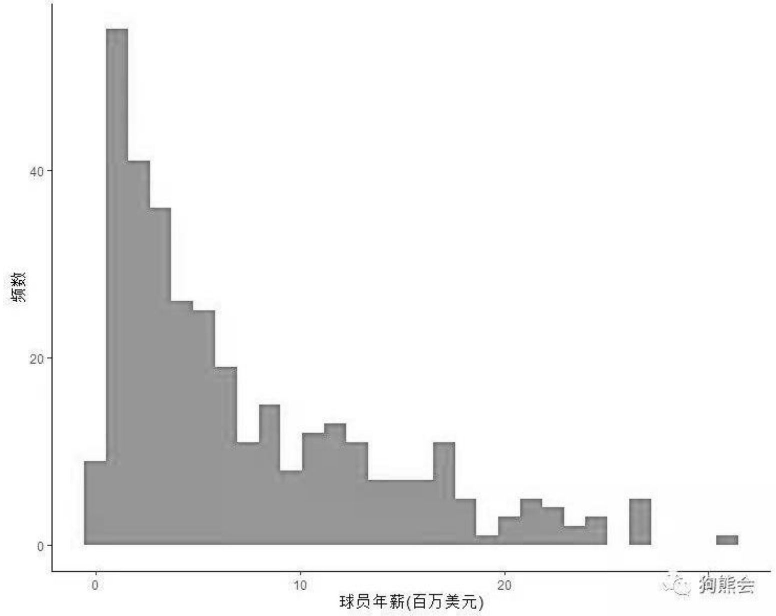
Which model is better?

A or B

The salary of people in China

A: mean

B: median



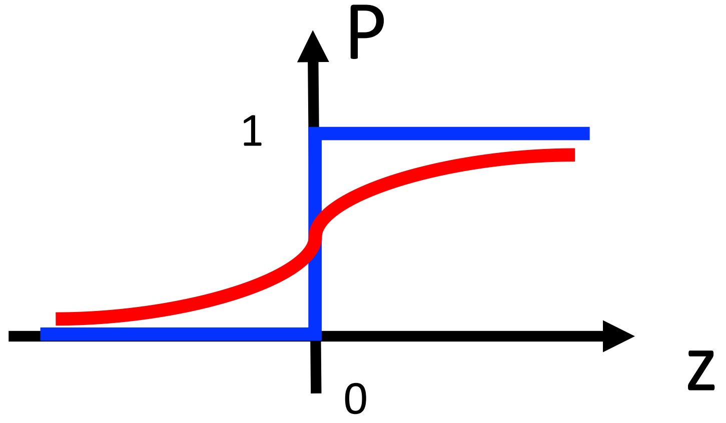
B

The data is not Gaussian. It has a long tail.

You are blind dating with two persons (p1 and p2).
Variable z is your preference of p1 over p2.
Variable P is the probability of choosing p1.

A: $1/(1+e^{-z})$

B: *0 (if z<0), 1(otherwise)*



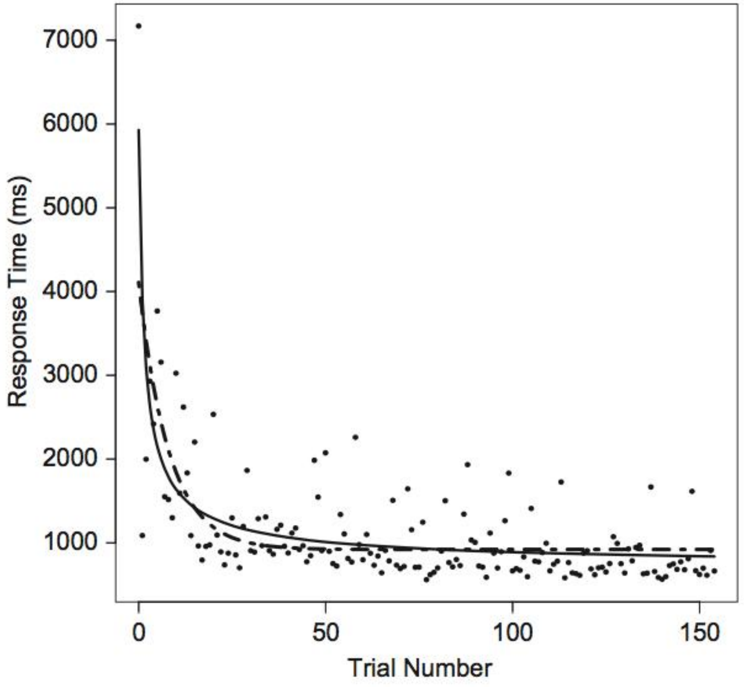
A

Allow some randomness

The relation between response time (RT) and the trials of practice (n):

A: $RT(n) = n^{-\beta}$

B: $RT(n) = e^{-\alpha n}$



B

The learning rate between trials is constant.