Deep Learning CS583

Jia Xu

Stevens Institute of Technology

2020 Fall

about CS583: contacts

about CS583: contacts

lecture time:

6:30-9:00PM Thursdays

Recitation time:

Dishti: 14:30-15:30 Fridays

Sattvik: 10:30-11:30 Fridays

about CS583: contacts

lecture time:

6:30-9:00PM Thursdays

Recitation time:

Dishti: 14:30-15:30 Fridays

Sattvik: 10:30-11:30 Fridays

Office hours:

Instructor: 12:00-14:00 Fridays

Dishti: 13:30-14:30 Fridays

Sattvik: 9:30-10:30 Fridays

about CS583: contacts

lecture time:

6:30-9:00PM Thursdays

Recitation time:

Dishti: 14:30-15:30 Fridays

Sattvik: 10:30-11:30 Fridays

Office hours:

Instructor: 12:00-14:00 Fridays

Dishti: 13:30-14:30 Fridays

Sattvik: 9:30-10:30 Fridays

Email:

Instructor: jxu70@stevens.edu

Dishti: ddave2@stevens.edu

Sattvik: ssahai@stevens.edu

about CS583: grading

about CS583: grading

Grading:

A: excellent empirical ability and deep understanding
B: good empirical ability and good understanding
C: meet basic requirement
D: fail in following up the course content

about CS583: grading

Grading:

A: excellent empirical ability and deep understanding
B: good empirical ability and good understanding
C: meet basic requirement
D: fail in following up the course content

Grading proportional to
the ranking of your total score in the class

about CS583: how to score

about CS583: how to score

Midterm exam: 10%
Final exam: 10%
Presentation: 15%
Q&A participation: 10%
Quizz (3-5): 15%
Homework assignments (4): 40%

about CS583: textbook

about CS583: textbook

Textbook:

Deep Learning. Ian Goodfellow, Yoshua Bengio, and Aaron Courville.
http://www.deeplearningbook.org

Other Readings:

S. Boyd and L. Vandenberghe. Introduction to Applied Linear Algebra. Cambridge University Press, 2018. (Available online.)

Y. Nesterov. Introductory Lectures on Convex Optimization Book. Springer, 2013. (Available online.)

D. S. Watkins. Fundamentals of Matrix Computations. John Wiley & Sons, 2004.

Francois Chollet. Deep learning with Python. Manning Publications Co., 2017. (Available online.)

M. Mohri, A. Rostamizadeh, and A. Talwalkar. Foundations of machine learning. MIT press, 2012.

J. Friedman, T. Hastie, and R. Tibshirani. The elements of statistical learning. Springer series in statistics, 2001. (Available online.)

about CS583: tentative topics

Lecture slides	Main topics	Sections in Textbook “Deep Learning” by Goodfellow, Bengio, Courville
09/03	Introduction, ML basics	Section 5
09/10	Math Background	Section 2-4
09/17	MLP	Section 6
09/24	Stochastic Gradient Descent	Section 6
10/01	Deep Feedforward Networks	Section 6
10/08	Regularization, Optimization	Section 7, 8
10/15	Midterm Exam	All above
10/22	RNN, LSTM	Section 10
10/29	CNN	Section 9
11/5	Attention	beyond textbook
11/12	GAN	Section 20
11/19	Auto-encoders	Section 14
11/26	Thanksgiving	
12/03	Reinforcement learning	beyond textbook
12/10	Final exam	All above

Question?

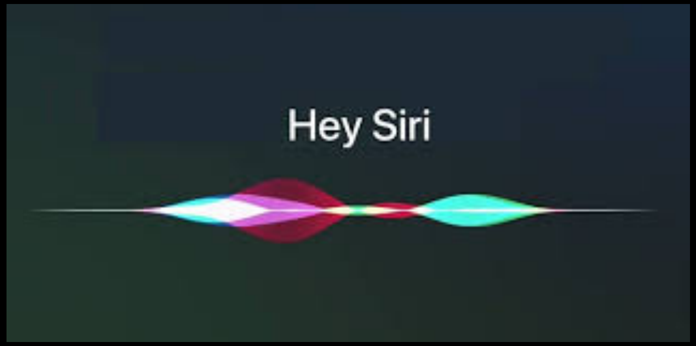
Deep Learning: introduction

Now, deep learning is everywhere

Now, deep learning is everywhere

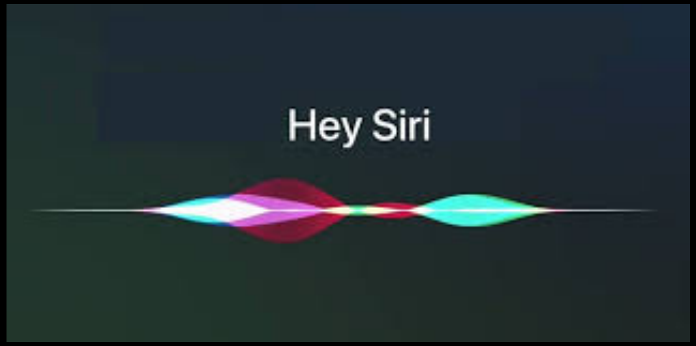
nowadays deep learning
is applied in society, science,
arts, commerce and finance,
literature, military, …

Now, deep learning is everywhere



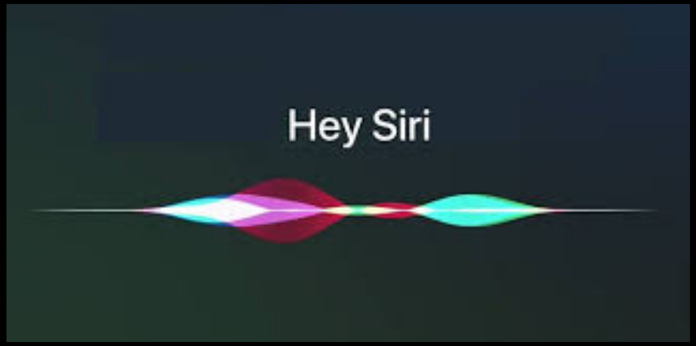
nowadays deep learning
is applied in society, science,
arts, commerce and finance,
literature, military, …

Now, deep learning is everywhere



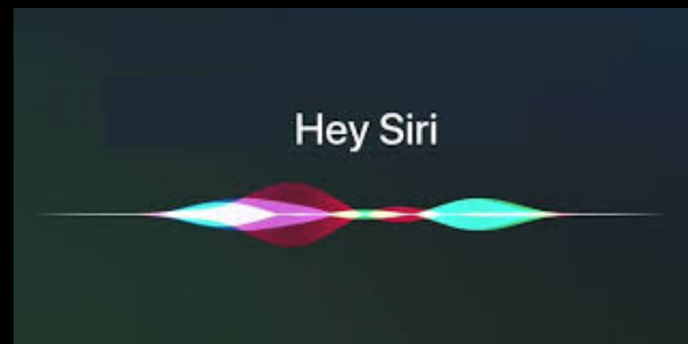
nowadays deep learning
is applied in society, science,
arts, commerce and finance,
literature, military, …

Now, deep learning is everywhere



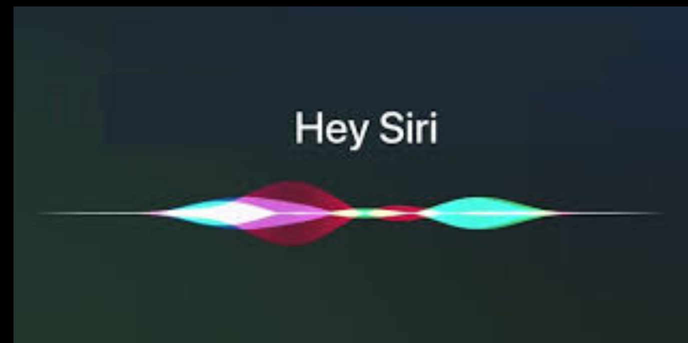
nowadays deep learning
is applied in society, science,
arts, commerce and finance,
literature, military, …

Now, deep learning is everywhere



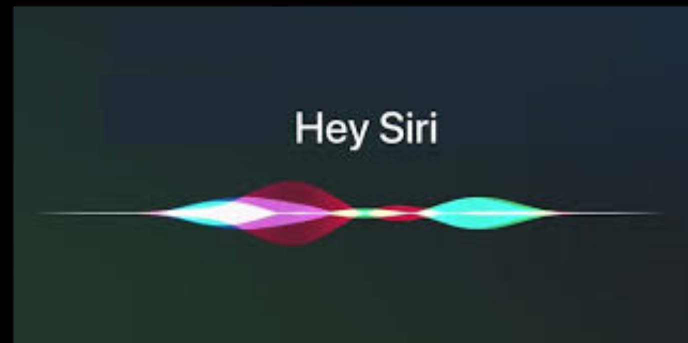
nowadays deep learning
is applied in society, science,
arts, commerce and finance,
literature, military, …

Now, deep learning is everywhere



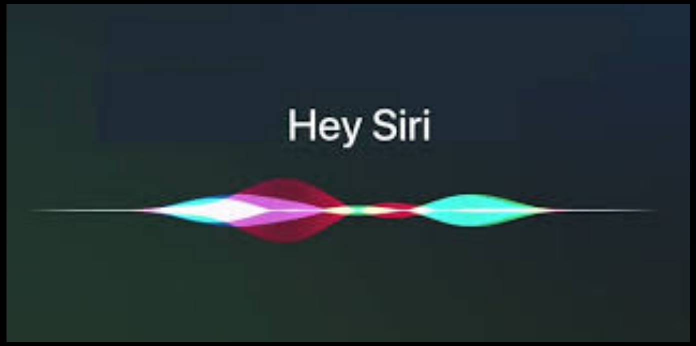
nowadays deep learning
is applied in society, science,
arts, commerce and finance,
literature, military, …

Now, deep learning is everywhere



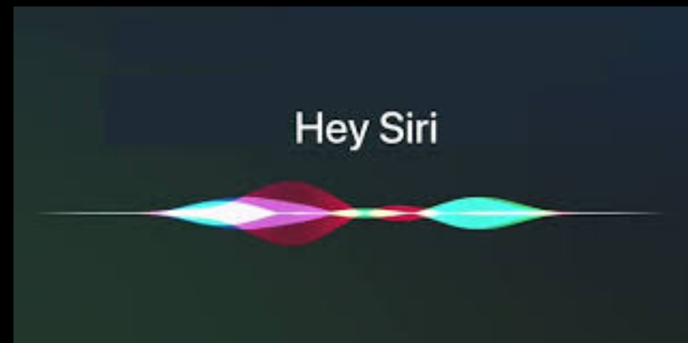
nowadays deep learning
is applied in society, science,
arts, commerce and finance,
literature, military, …

Now, deep learning is everywhere



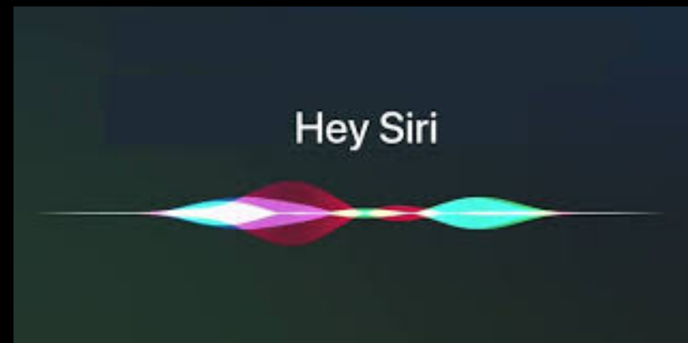
nowadays deep learning
is applied in society, science,
arts, commerce and finance,
literature, military, …

Now, deep learning is everywhere

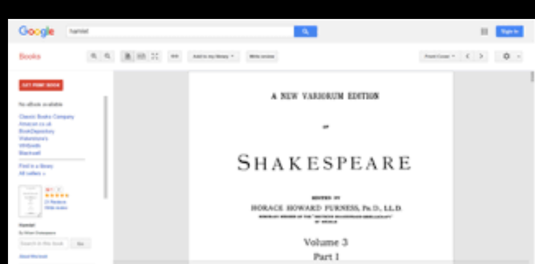


nowadays deep learning
is applied in society, science,
arts, commerce and finance,
literature, military, …

Now, deep learning is everywhere



nowadays deep learning
is applied in society, science,
arts, commerce and finance,
literature, military, …



Back in time,

...

How was deep learning introduced and developed?

...

What is the connection of deep learning with other fields?

from CS, statistics to machine learning

by Kirk Borne

examples of
AI, machine learning and deep learning

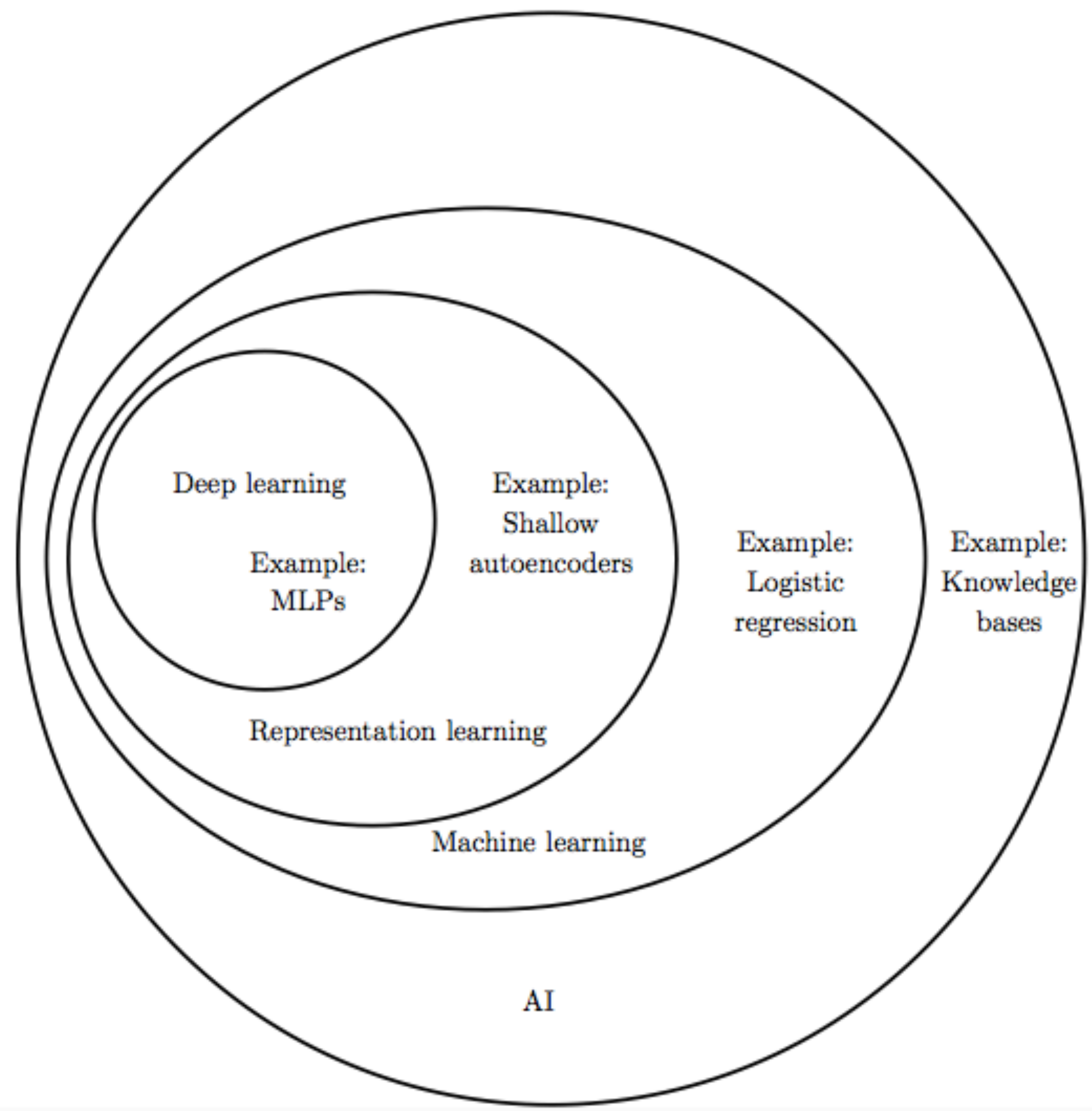
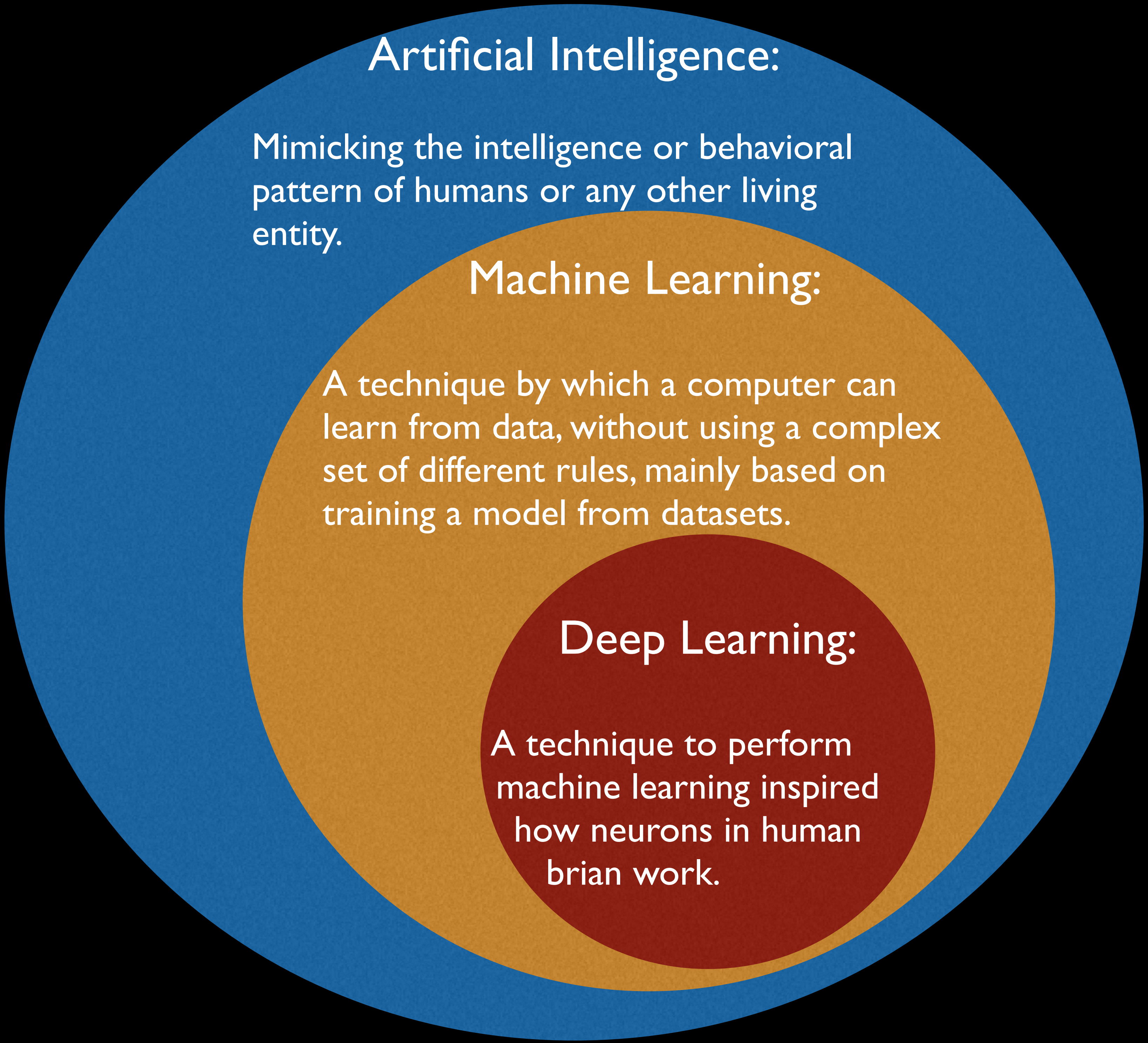
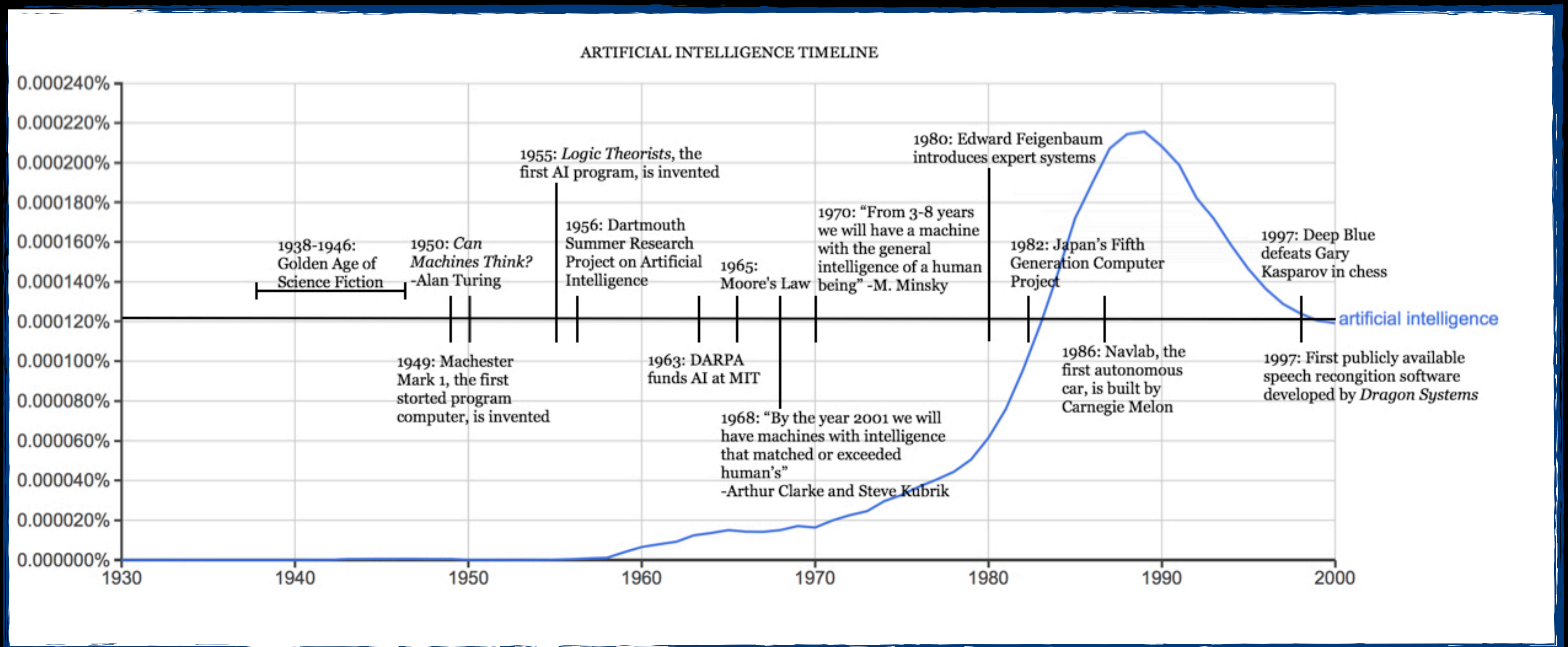


Figure 1.4

AI, machine learning, deep learning

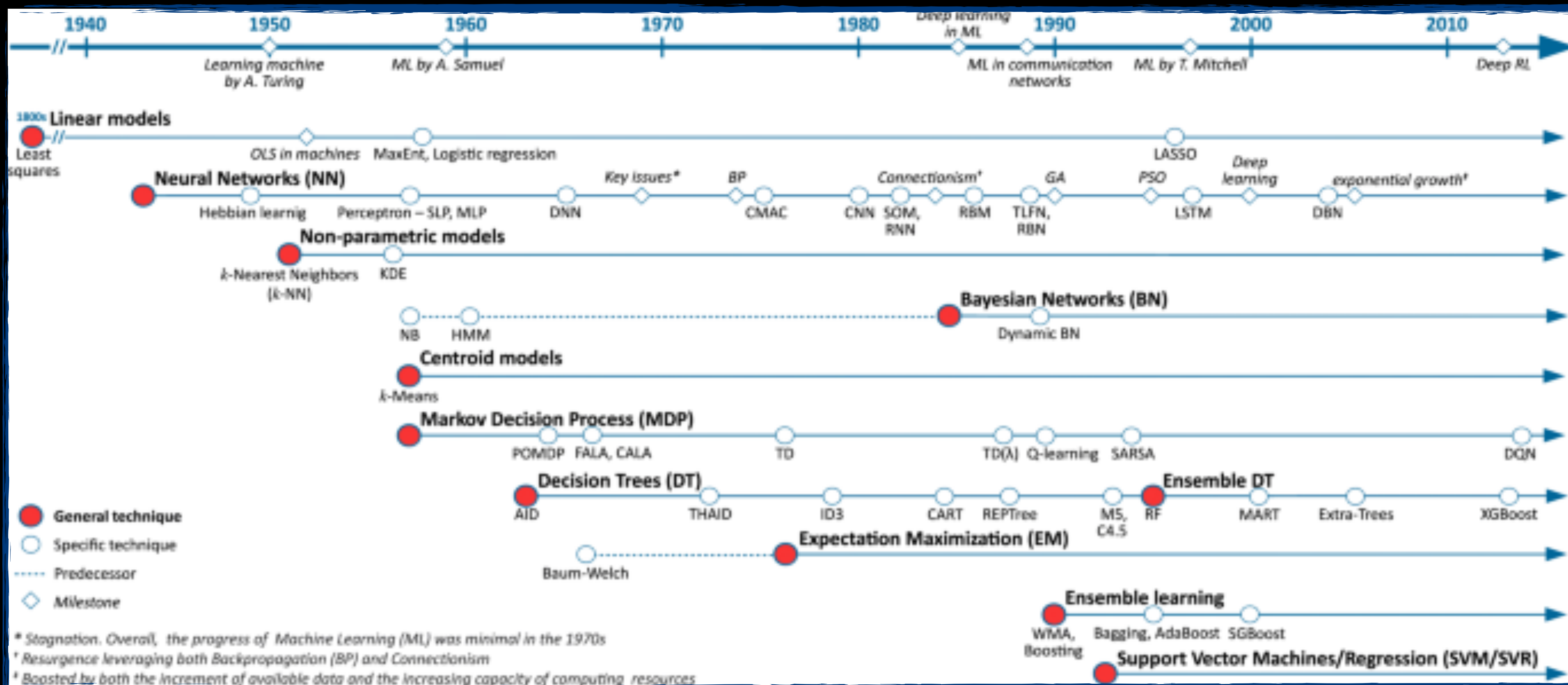


AI history



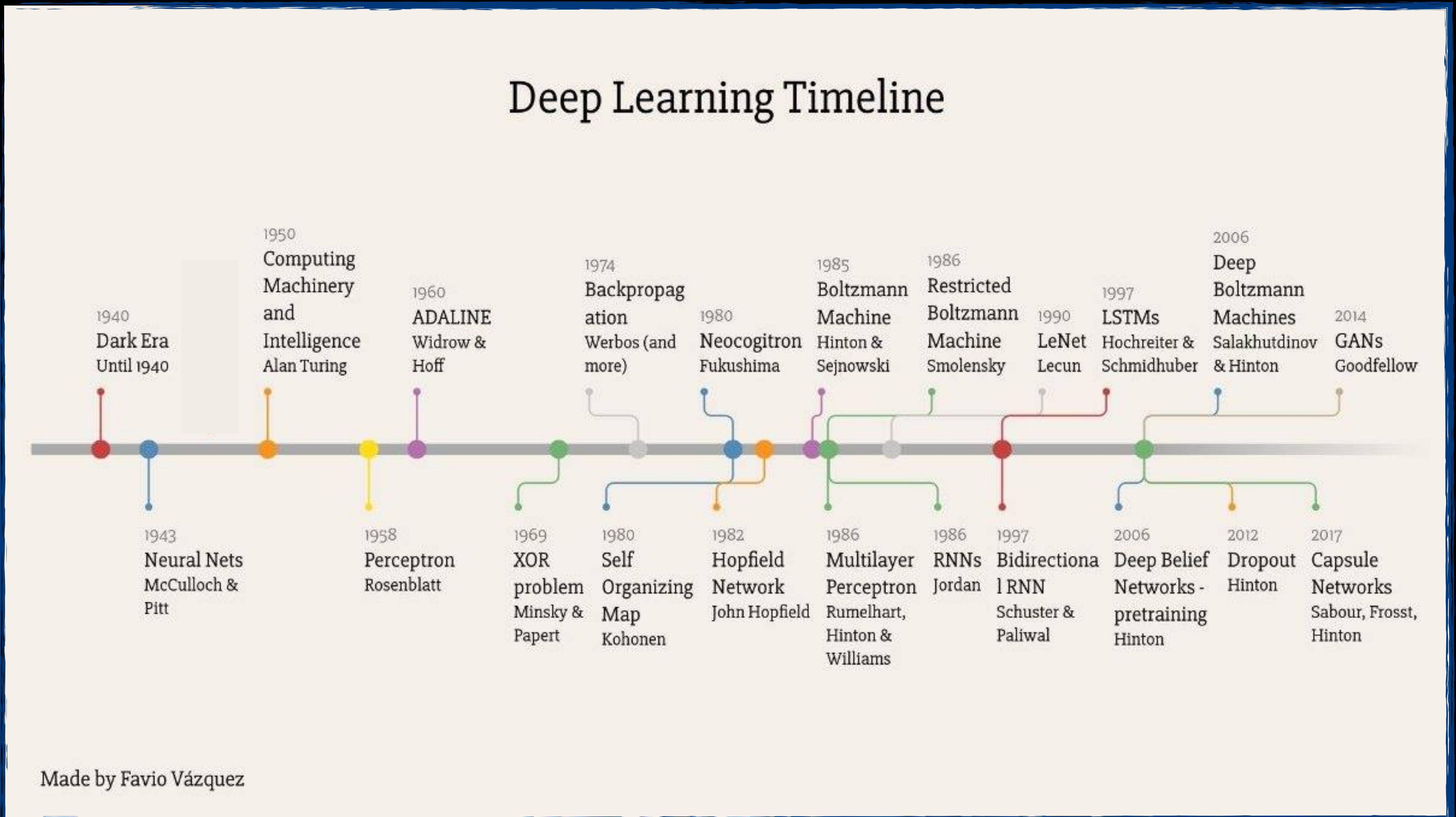
http://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/

machine learning history



https://jisajournal.springeropen.com/articles/10.1186/s13174-018-0087-2

deep learning history

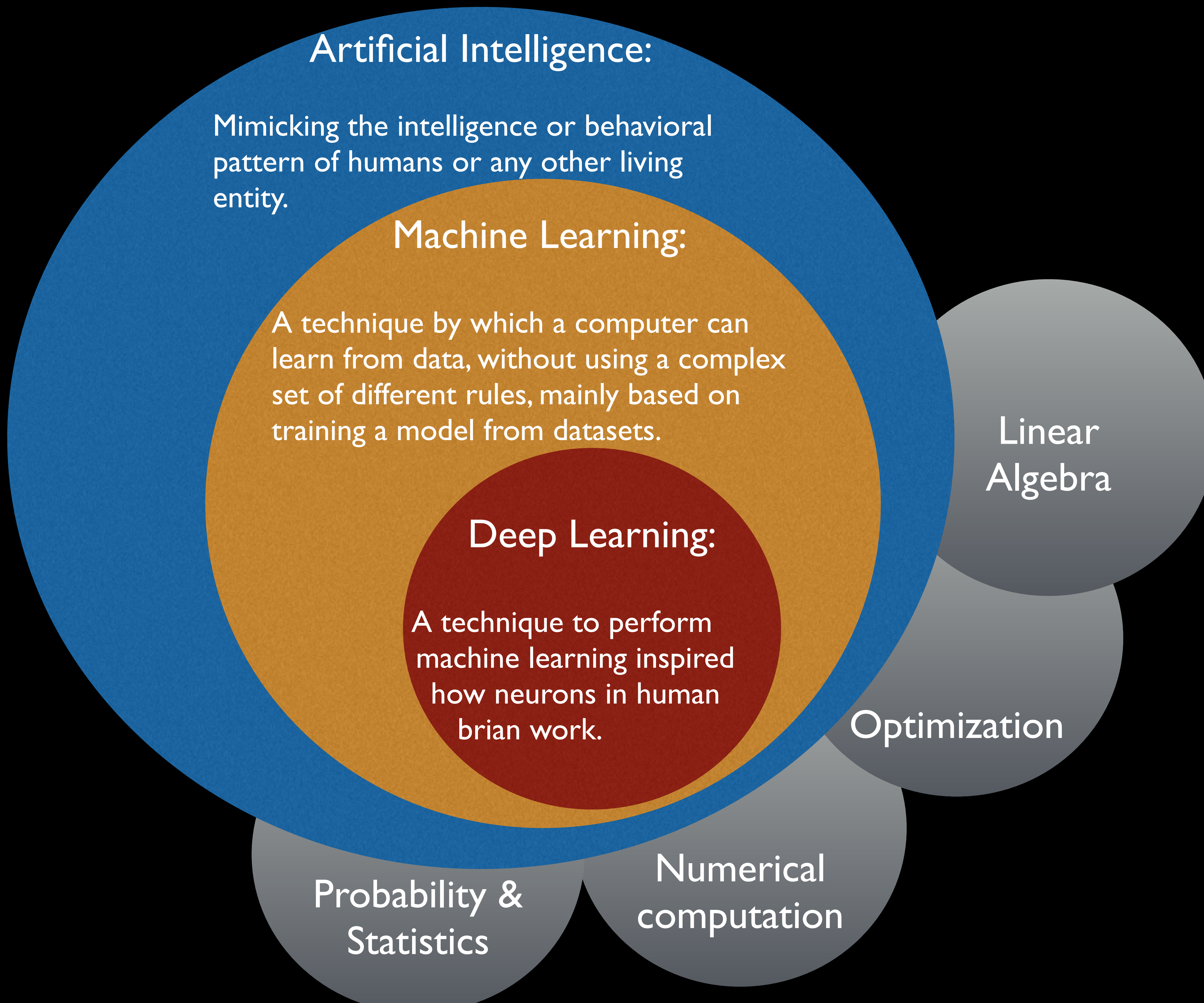


deep learning history

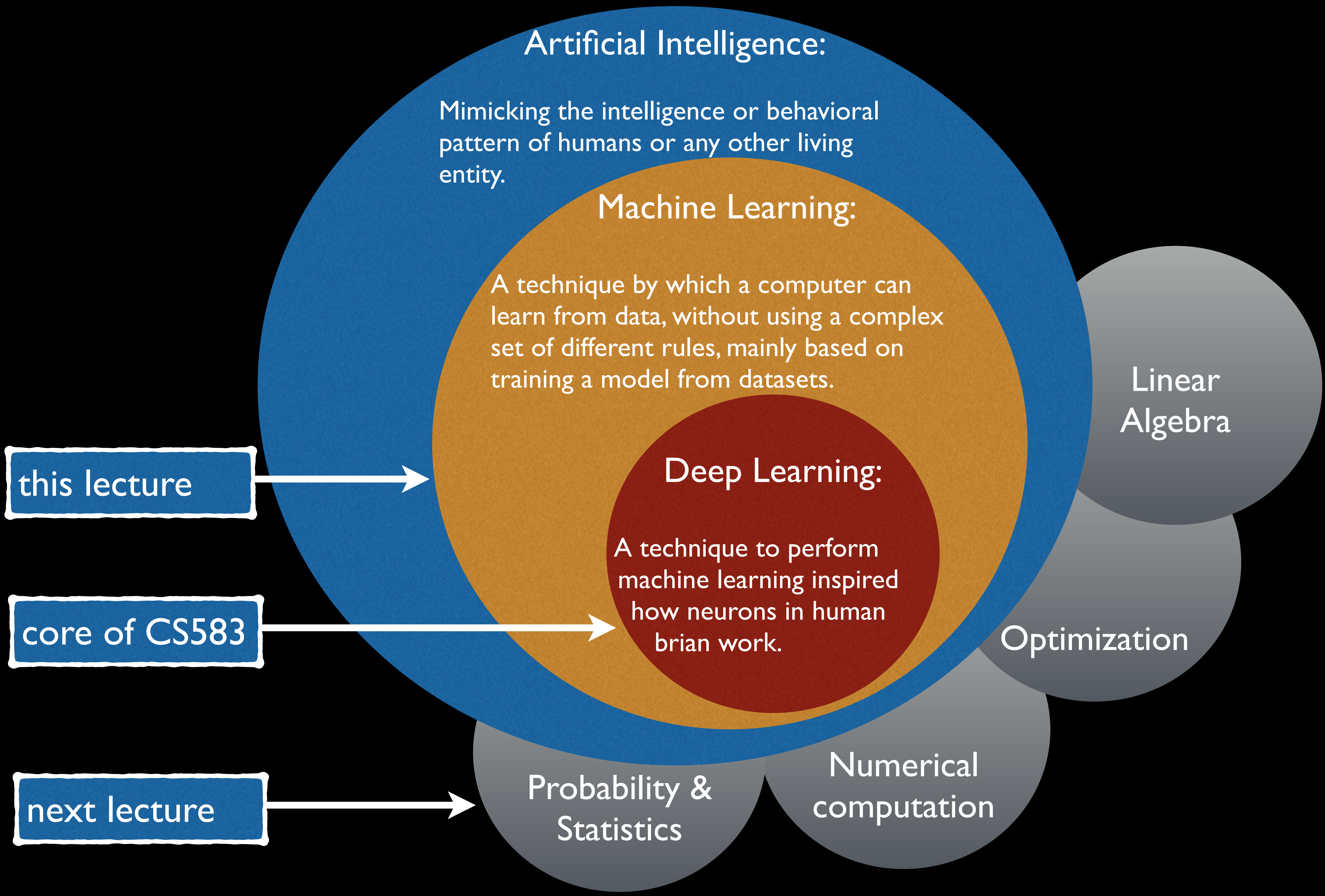
deep learning history

1943: McCulloch and Pitts, a computational model for neural networks based on threshold logic.
1958: Frank Rosenblatt, the perceptron, a two-layer computer neural network using simple addition and subtraction.
1980: Kunihiko Fukushima proposes the Neoconitron, a hierarchical, multilayered artificial neural network for handwriting recognition.
Mid-2000s: Geoffrey Hinton and Ruslan Salakhutdinov, "deep learning" — a many-layered neural network could be pre-trained one layer at a time.
2009: NIPS Workshop on Deep Learning for Speech Recognition
2015: Facebook DeepFace
2016: Google DeepMind's algorithm AlphaGo
…

AI, machine learning, deep Learning



AI, machine learning, deep Learning



Machine Learning

learning from examples

learning from examples

- detect credit card fraud

learning from examples

- detect credit card fraud

given past transactions
and
current purchase

current purchase
valid or fraud ?

learning from examples

- detect credit card fraud

given past transactions
and
current purchase

current purchase
valid or fraud ?

input

label

learning from examples

- detect credit card fraud

given past transactions
and
current purchase

current purchase
valid or fraud ?

input

label

example = input + label

learning from examples

- detect credit card fraud

given past transactions
and
current purchase

current purchase
valid or fraud ?

predict the answer
from past examples

learning from examples

- detect credit card fraud

given past transactions
and
current purchase

current purchase
valid or fraud ?

predict the answer
from past examples

- regression tasks in healthcare

learning from examples

- detect credit card fraud

given past transactions
and
current purchase

current purchase
valid or fraud ?

predict the answer
from past examples

- regression tasks in healthcare
- face recognition (e.g. AiCure in healthcare)
- natural language processing

…

learning from examples

- detect credit card fraud

given past transactions
and
current purchase

current purchase
valid or fraud ?

predict the answer
from past examples

- regression tasks in healthcare
- face recognition (e.g. AiCure in healthcare)
- natural language processing

…

can big data help to **generically**
improve the prediction accuracy ?

learning from examples

- detect credit card fraud

given past transactions
and
current purchase

current purchase
valid or fraud ?

predict the answer
from past examples

- regression tasks in healthcare
- face recognition (e.g. AiCure in healthcare)
- natural language processing

…

can big data help to **generically**
improve the prediction accuracy ?

machine learning example

binary classification

label (classify)
emails as
spam or ham

From: howtobeabillioner.com
Subject: lottery

You've been selected
among millions to receive....

...viagra...

...Rolex...

From: mike@tsinghua.edu.cn
Subject: manuscript

hi Nick,

Please find attached the
corrections to our
manuscript.
...

machine learning example

binary classification

label (classify)
emails as
spam or ham

From: howtobeabillioner.com
Subject: lottery

You've been selected
among millions to receive....

...viagra...

...Rolex...

From: mike@tsinghua.edu.cn
Subject: manuscript

hi Nick,

Please find attached the
corrections to our
manuscript.
...

given an email
extract two features:
x_1 and x_2 (real numbers)

machine learning example

binary classification

label (classify)
emails as
spam or ham

From: howtobeabillioner.com
Subject: lottery

You've been selected
among millions to receive....

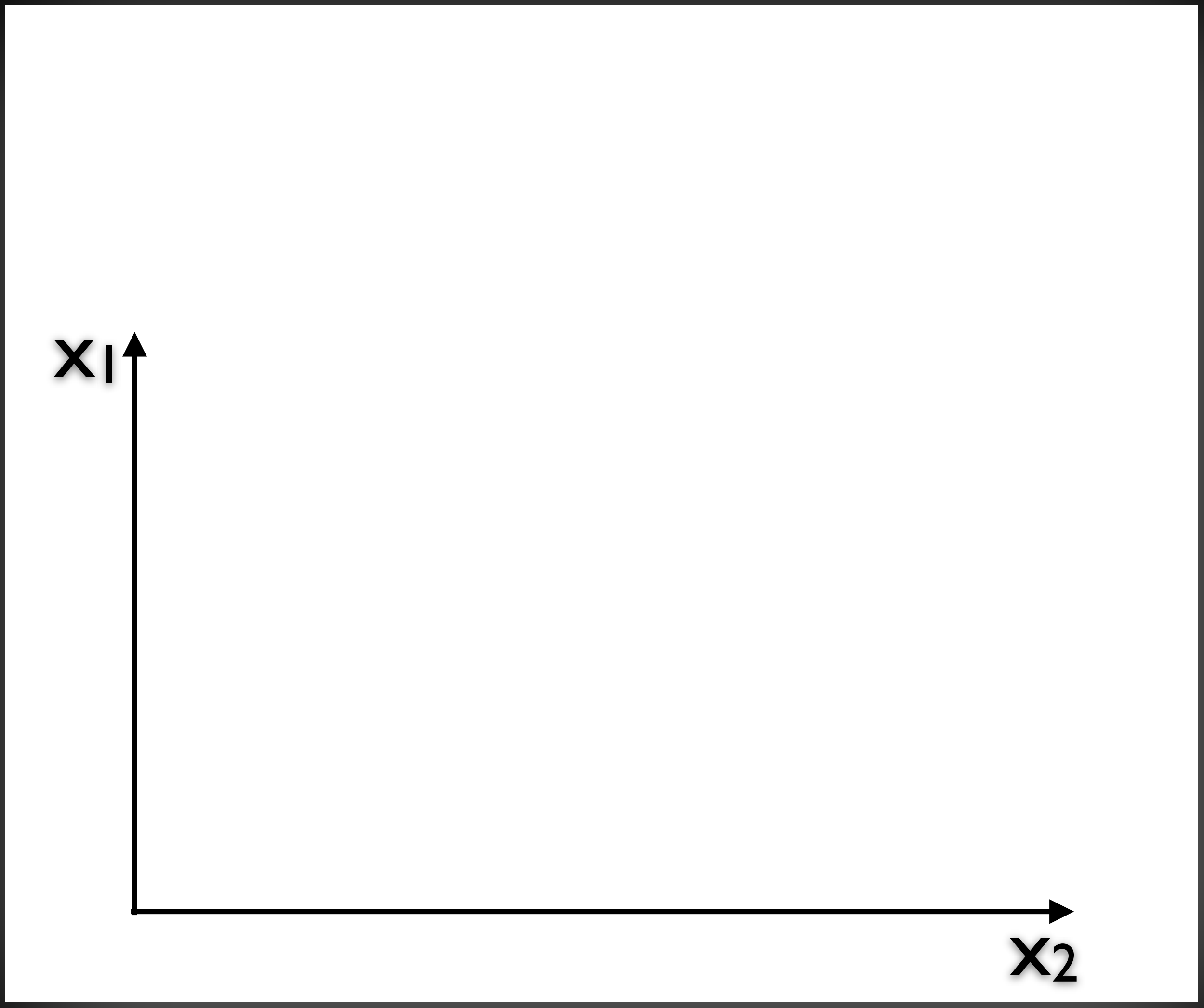
...viagra...

...Rolex...

From: mike@tsinghua.edu.cn
Subject: manuscript

hi Nick,

Please find attached the
corrections to our
manuscript.
...



given an email
extract two features:
x_1 and x_2 (real numbers)

machine learning example

binary classification

label (classify)
emails as
spam or ham

From: howtobeabillioner.com
Subject: lottery

You've been selected
among millions to receive....

...viagra...

...Rolex...

From: mike@tsinghua.edu.cn
Subject: manuscript

hi Nick,

Please find attached the
corrections to our
manuscript.
...

given an email
extract two features:
x_1 and x_2 (real numbers)

observations (training set)

machine learning example

binary classification

label (classify)
emails as
spam or ham

From: howtobeabillioner.com
Subject: lottery

You've been selected
among millions to receive....

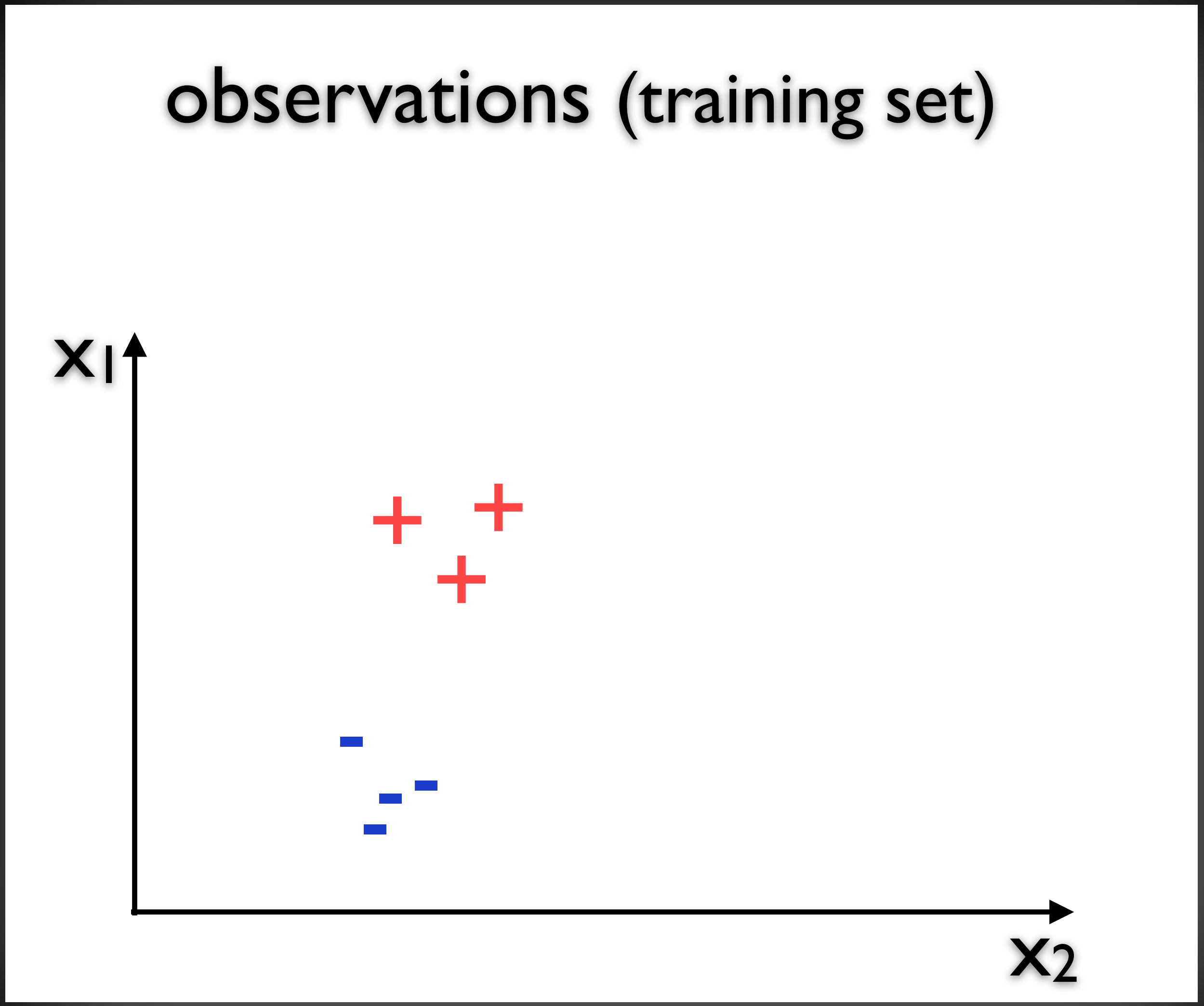
...viagra...

...Rolex...

From: mike@tsinghua.edu.cn
Subject: manuscript

hi Nick,

Please find attached the
corrections to our
manuscript.
...



given an email
extract two **features**:
x_1 and x_2 (real numbers)

machine learning example

binary classification

label (classify)
emails as
spam or ham

From: howtobeabillioner.com
Subject: lottery

You've been selected
among millions to receive....

...viagra...

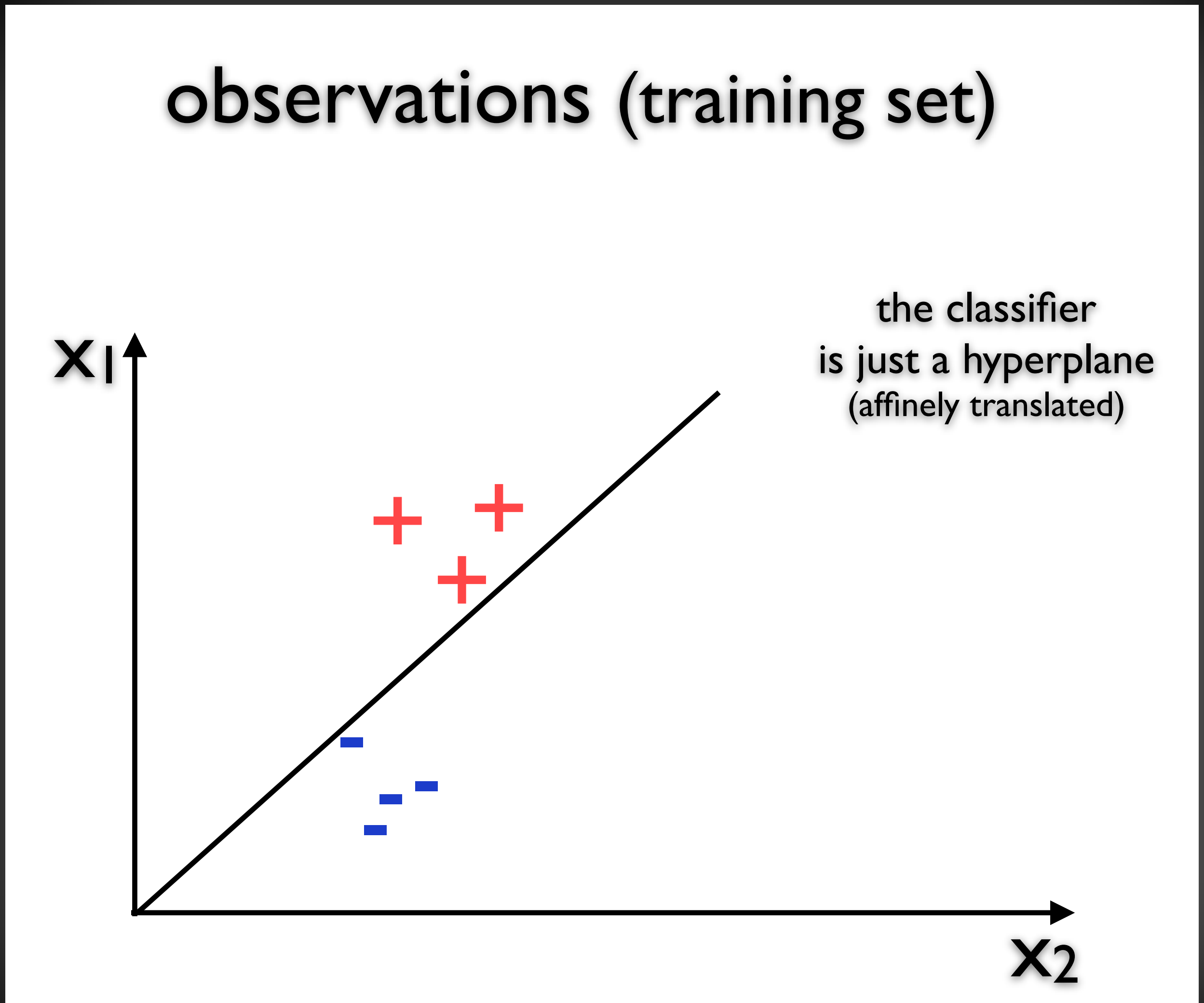
...Rolex...

From: mike@tsinghua.edu.cn
Subject: manuscript

hi Nick,

Please find attached the
corrections to our
manuscript.
...

given an email
extract two features:
x_1 and x_2 (real numbers)



machine learning example

binary classification

label (classify)
emails as
spam or ham

From: howtobeabillioner.com
Subject: lottery

You've been selected
among millions to receive....

...viagra...

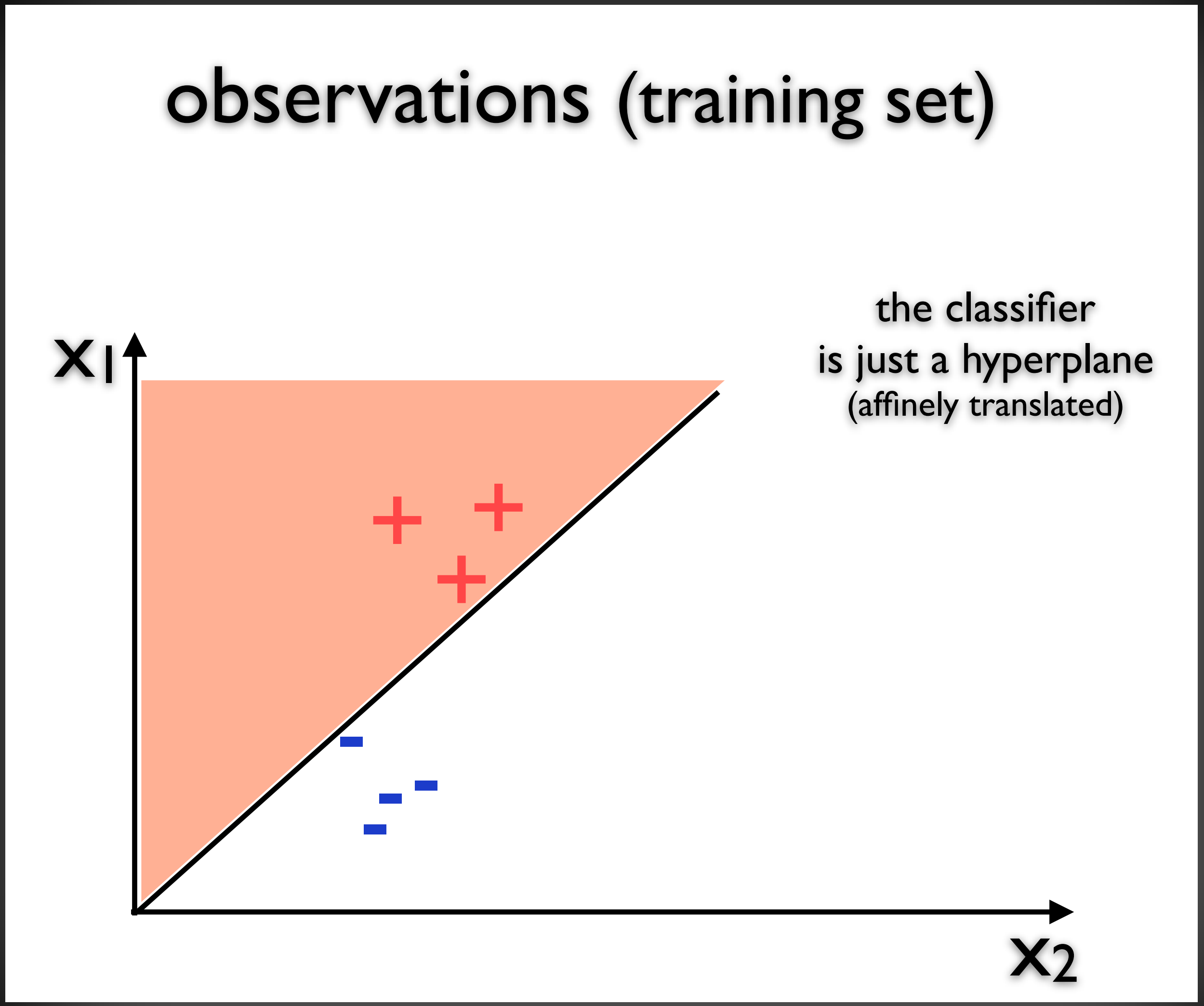
...Rolex...

From: mike@tsinghua.edu.cn
Subject: manuscript

hi Nick,

Please find attached the
corrections to our
manuscript.
...

given an email
extract two features:
x_1 and x_2 (real numbers)



machine learning example

binary classification

label (classify)
emails as
spam or ham

From: howtobeabillioner.com
Subject: lottery

You've been selected
among millions to receive....

...viagra...

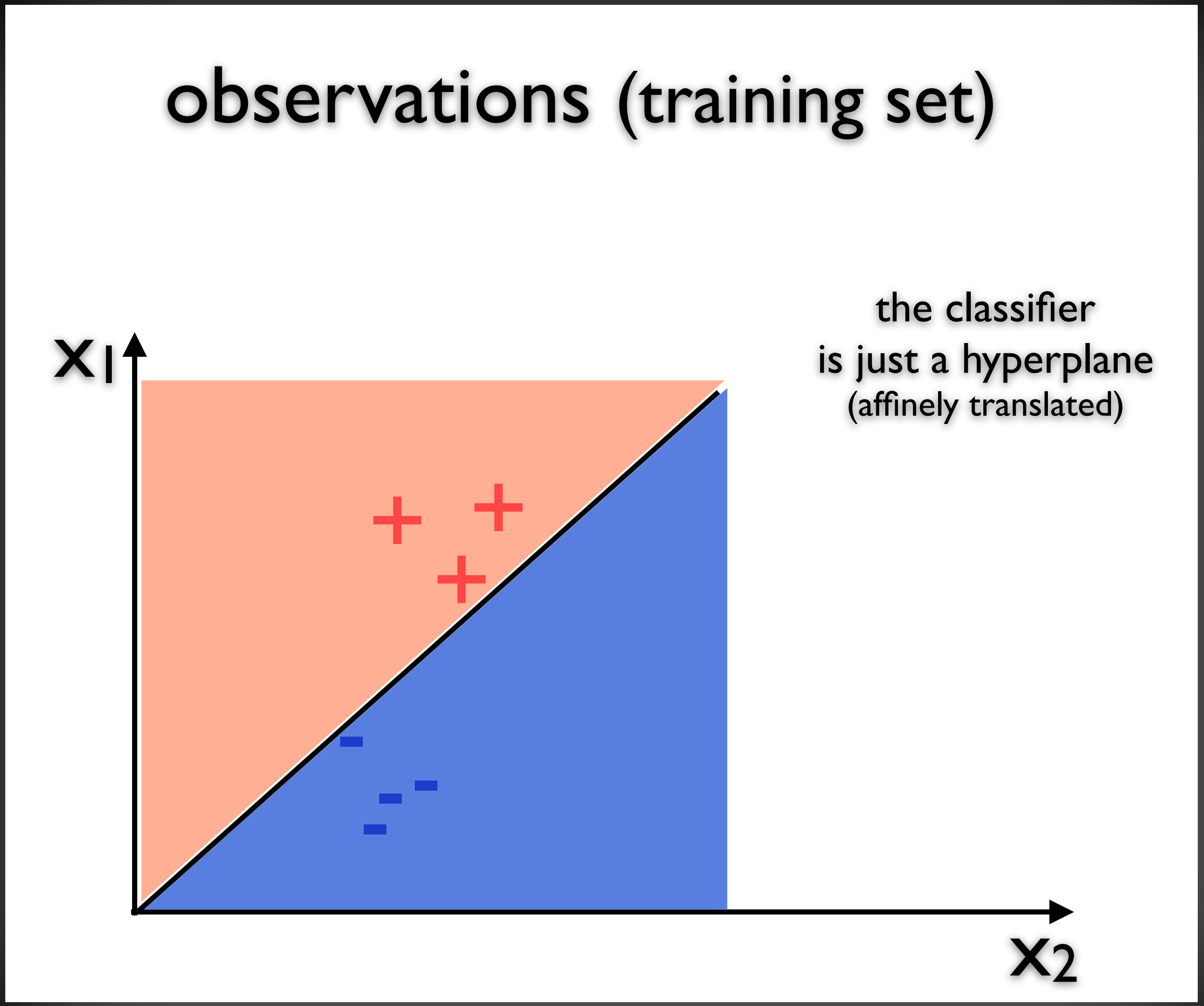
...Rolex...

From: mike@tsinghua.edu.cn
Subject: manuscript

hi Nick,

Please find attached the
corrections to our
manuscript.
...

given an email
extract two features:
x_1 and x_2 (real numbers)



machine learning example

binary classification

label (classify)
emails as
spam or ham

From: howtobeabillioner.com
Subject: lottery

You've been selected
among millions to receive....

...viagra...

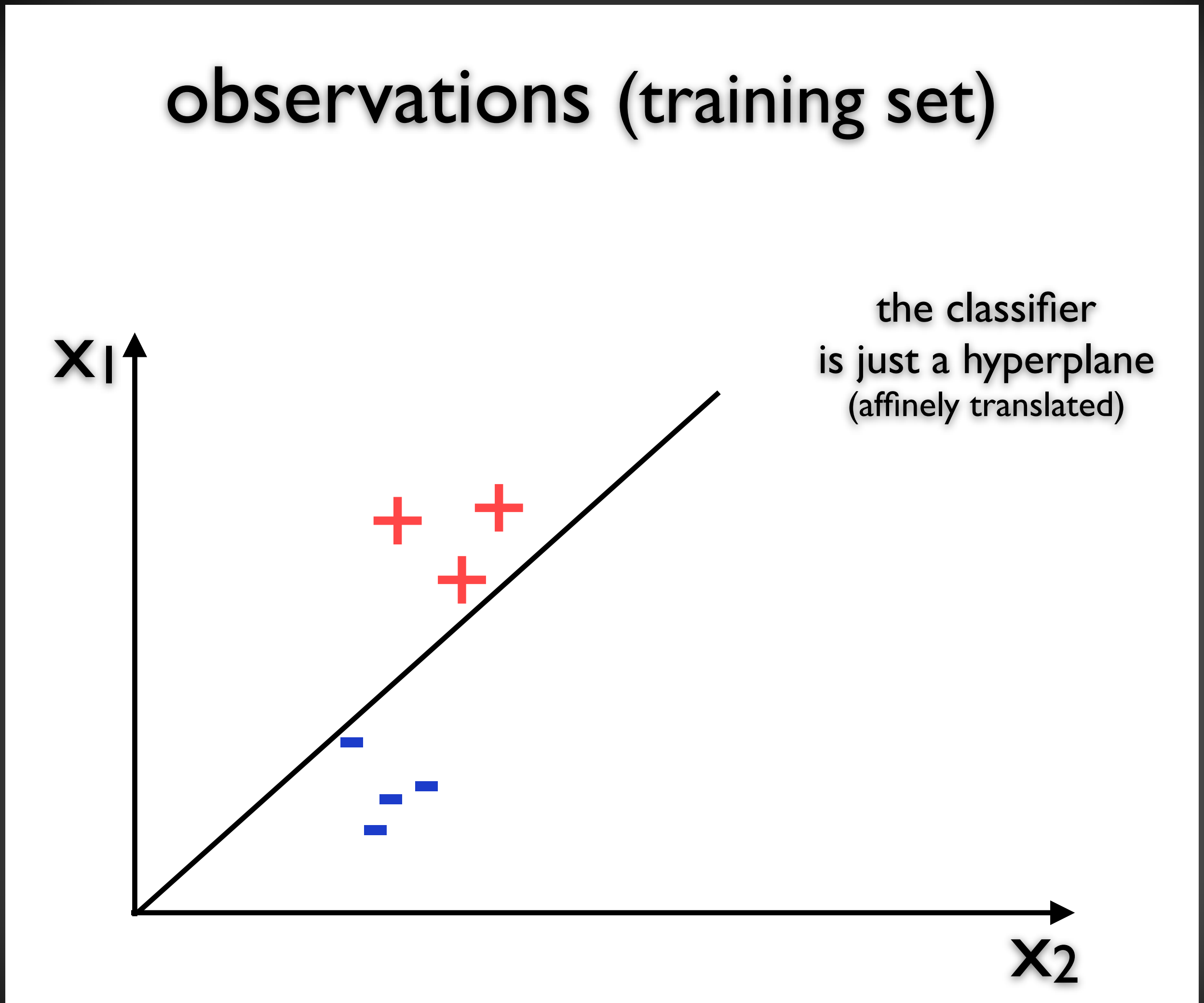
...Rolex...

From: mike@tsinghua.edu.cn
Subject: manuscript

hi Nick,

Please find attached the
corrections to our
manuscript.
...

given an email
extract two features:
x_1 and x_2 (real numbers)



machine learning example

binary classification

label (classify)
emails as
spam or ham

From: howtobeabillioner.com
Subject: lottery

You've been selected
among millions to receive....

...viagra...

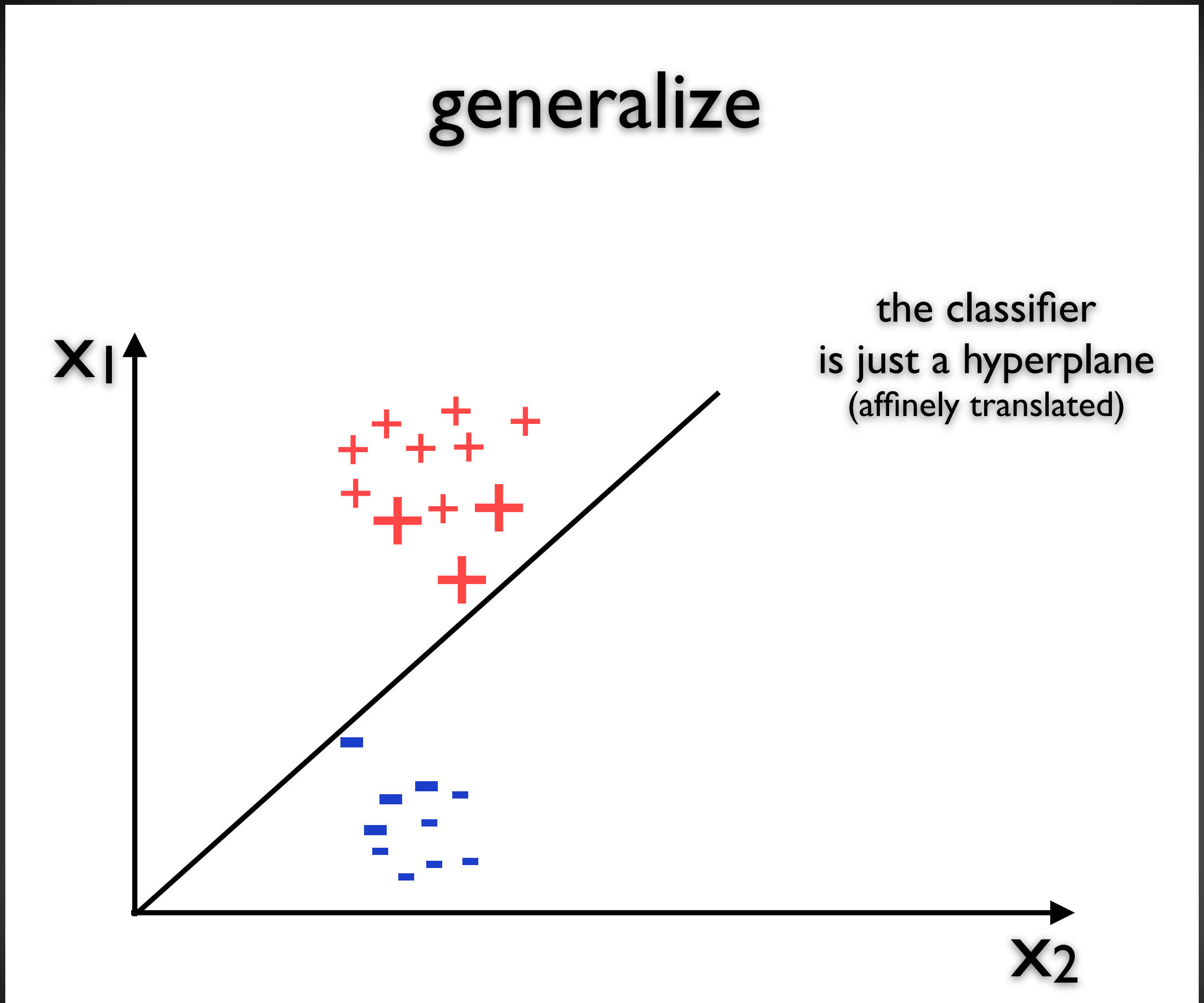
...Rolex...

From: mike@tsinghua.edu.cn
Subject: manuscript

hi Nick,

Please find attached the
corrections to our
manuscript.
...

given an email
extract two features:
x_1 and x_2 (real numbers)



machine learning example

binary classification

label (classify)
emails as
spam or ham

From: howtobeabillioner.com
Subject: lottery

You've been selected
among millions to receive....

...viagra...

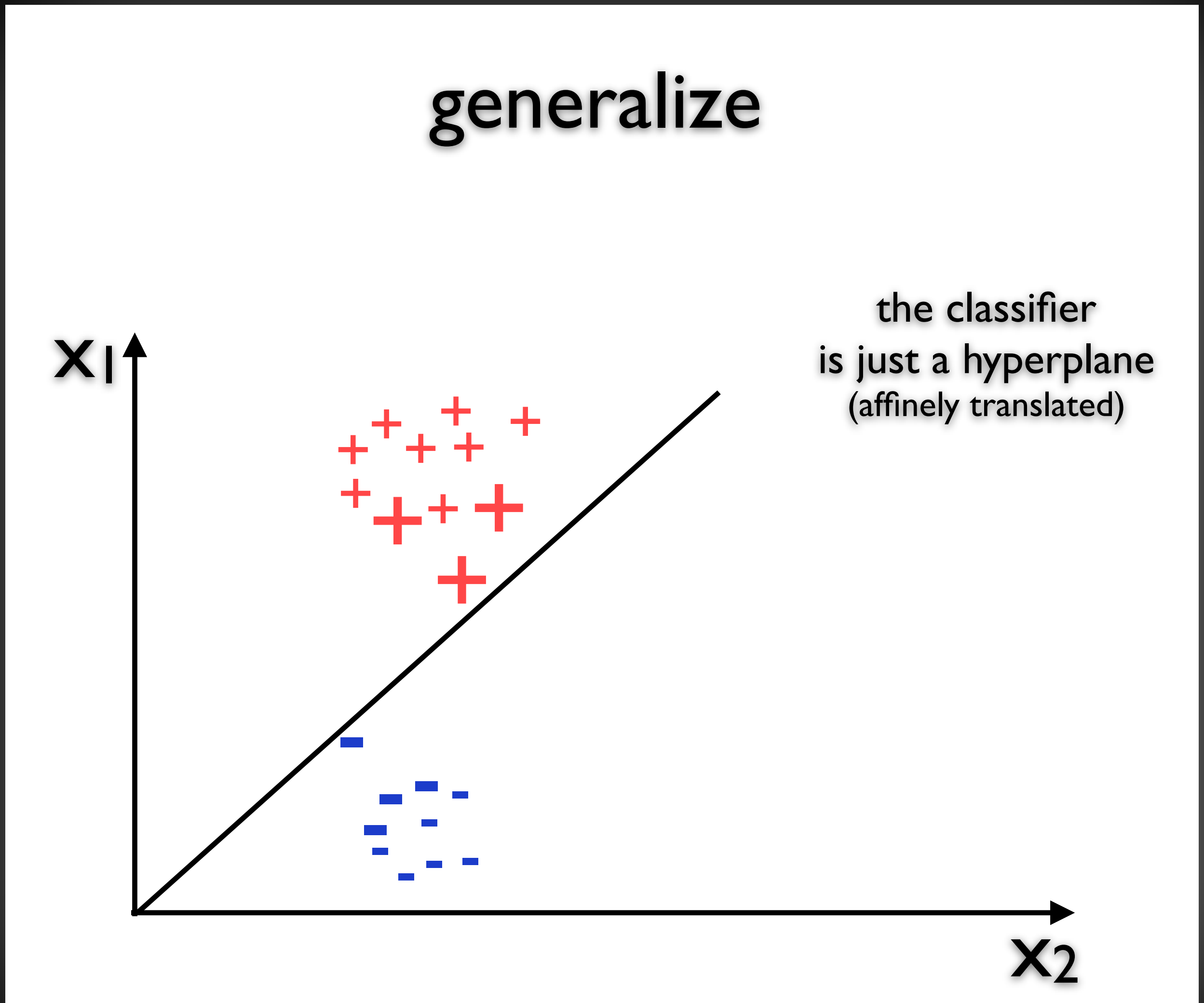
...Rolex...

From: mike@tsinghua.edu.cn
Subject: manuscript

hi Nick,

Please find attached the
corrections to our
manuscript.
...

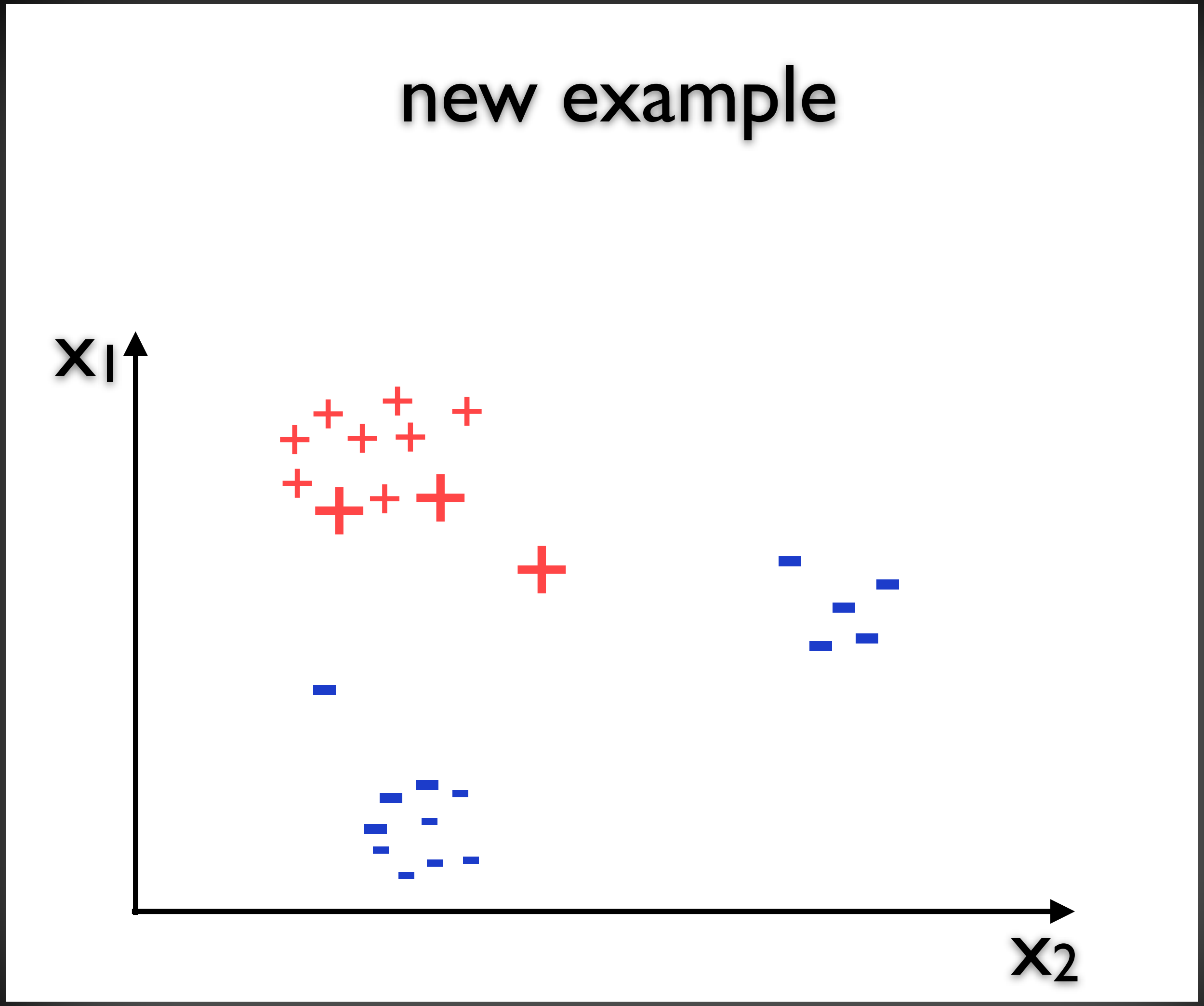
given an email
extract two features:
x_1 and x_2 (real numbers)



this example: no generalization error

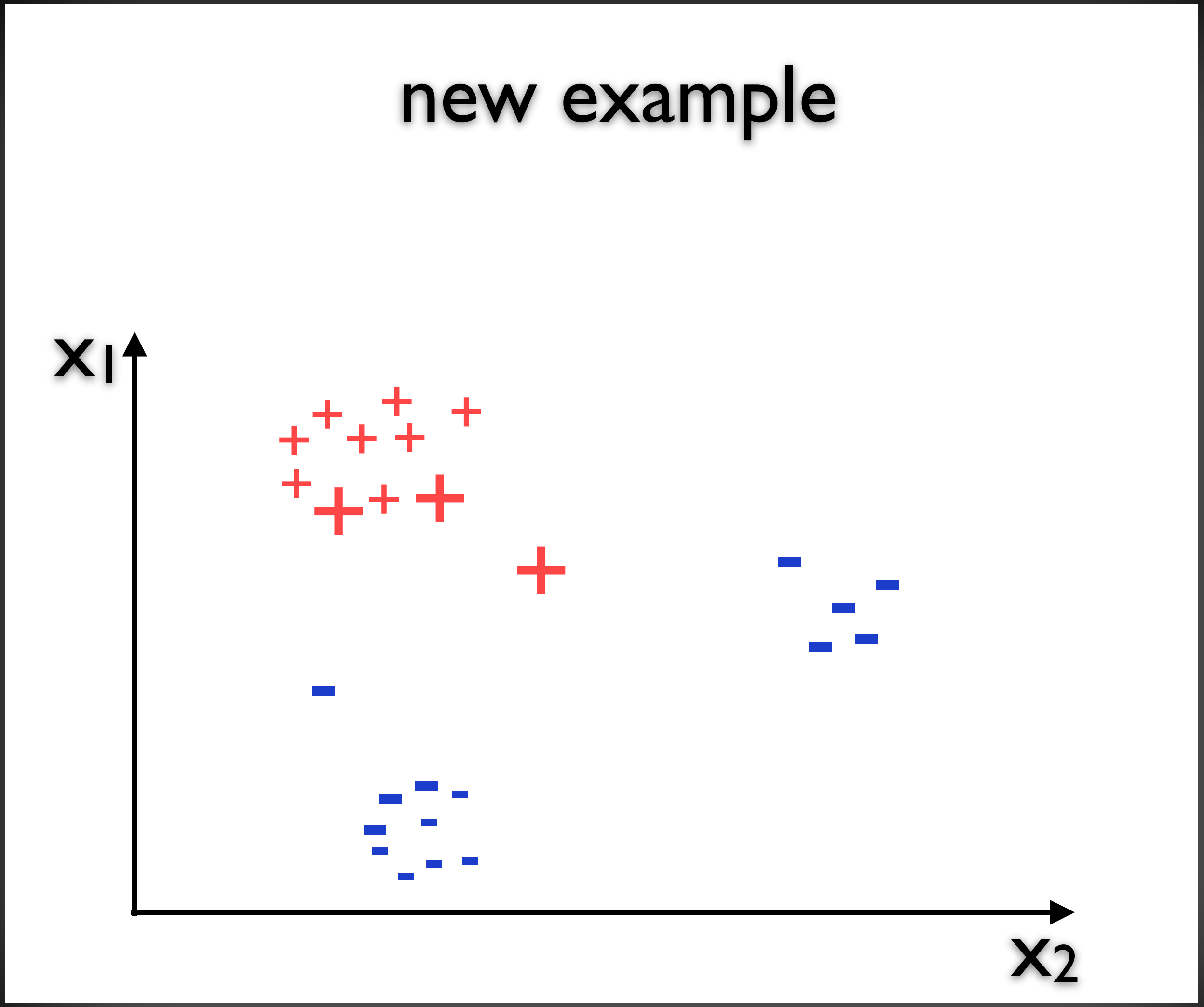
machine learning example

who constructs the classifier?



machine learning example

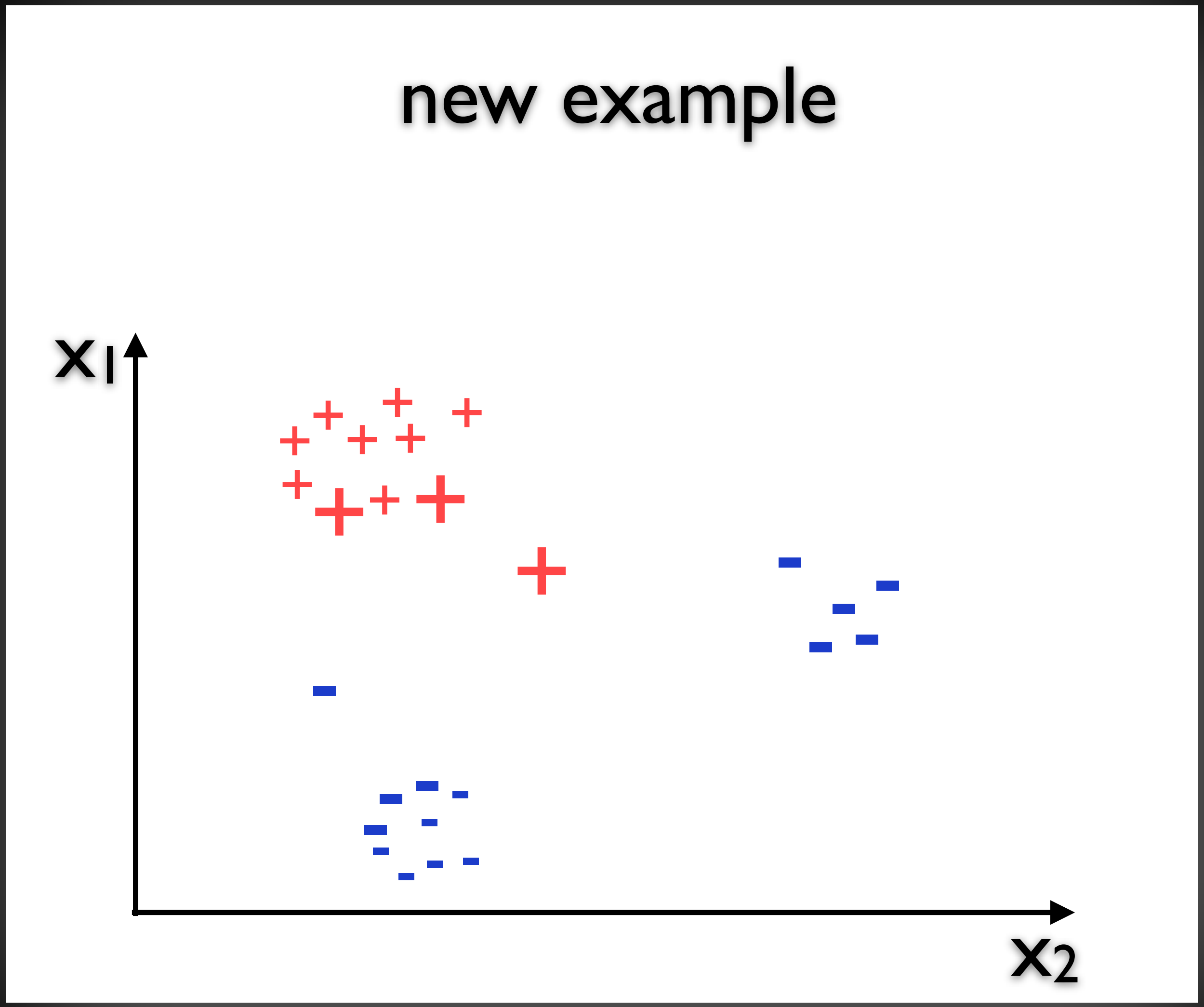
who constructs the classifier?



learner

machine learning example

who constructs the classifier?



(Y, (x_1,x_2)) , ...

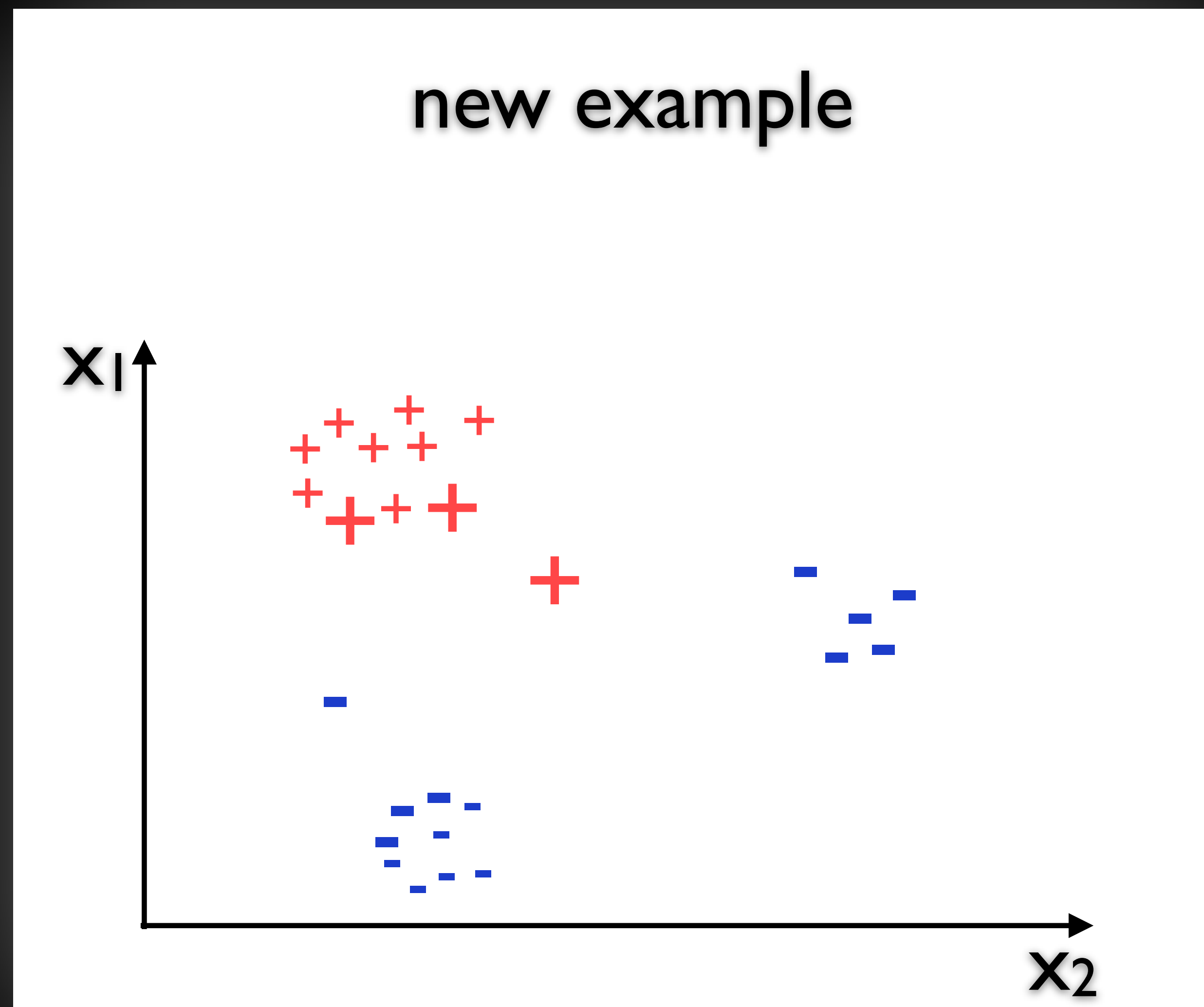
learner

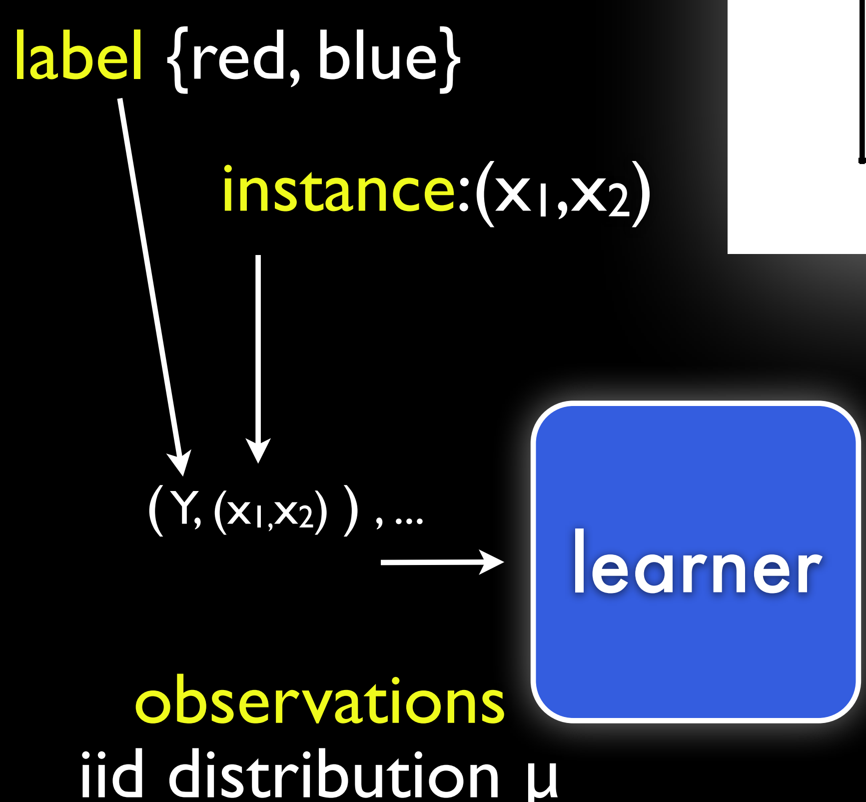
observations

iid distribution μ

machine learning example

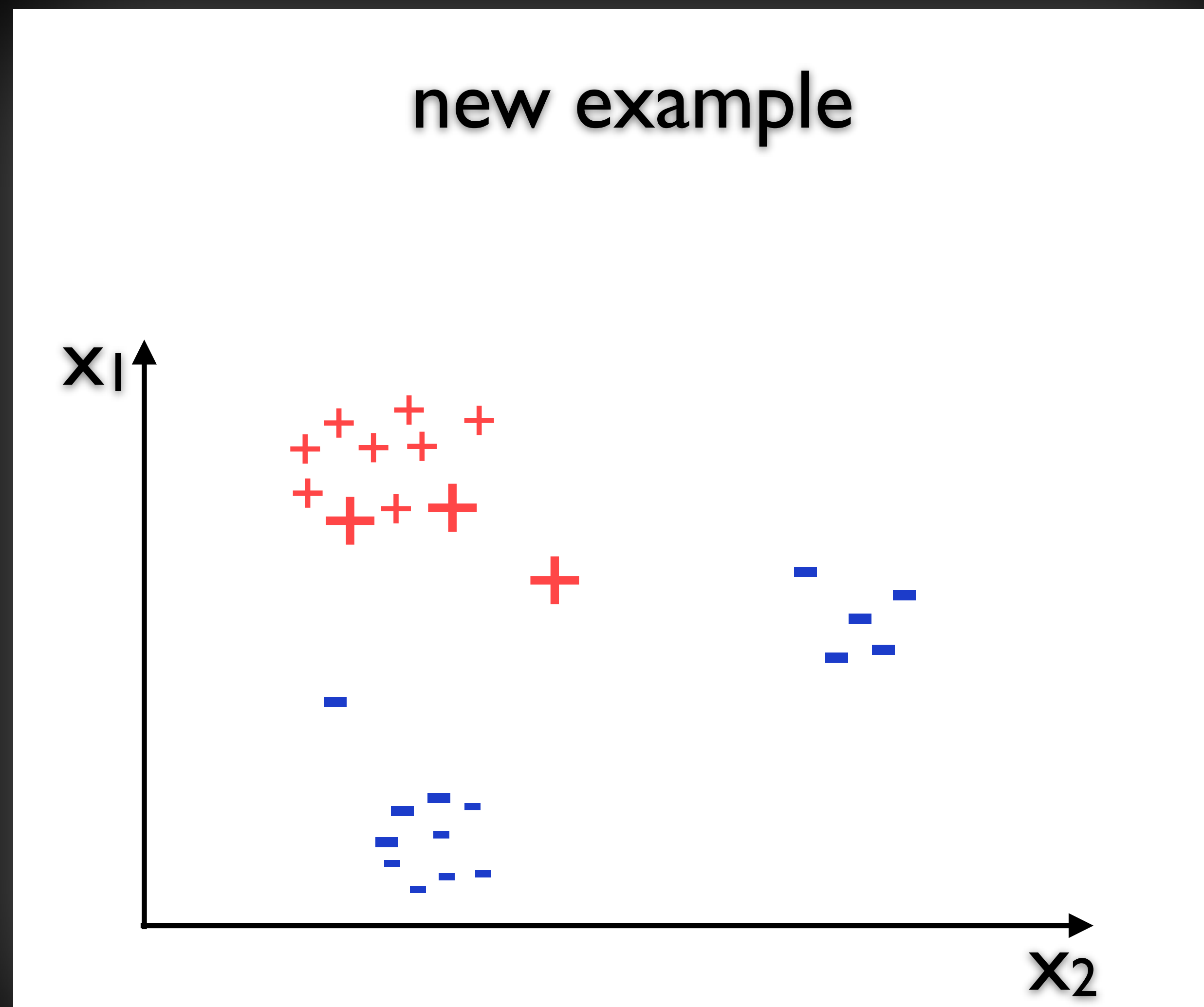
who constructs the classifier?





machine learning example

who constructs the classifier?



label {red, blue}

instance:(x_1,x_2)

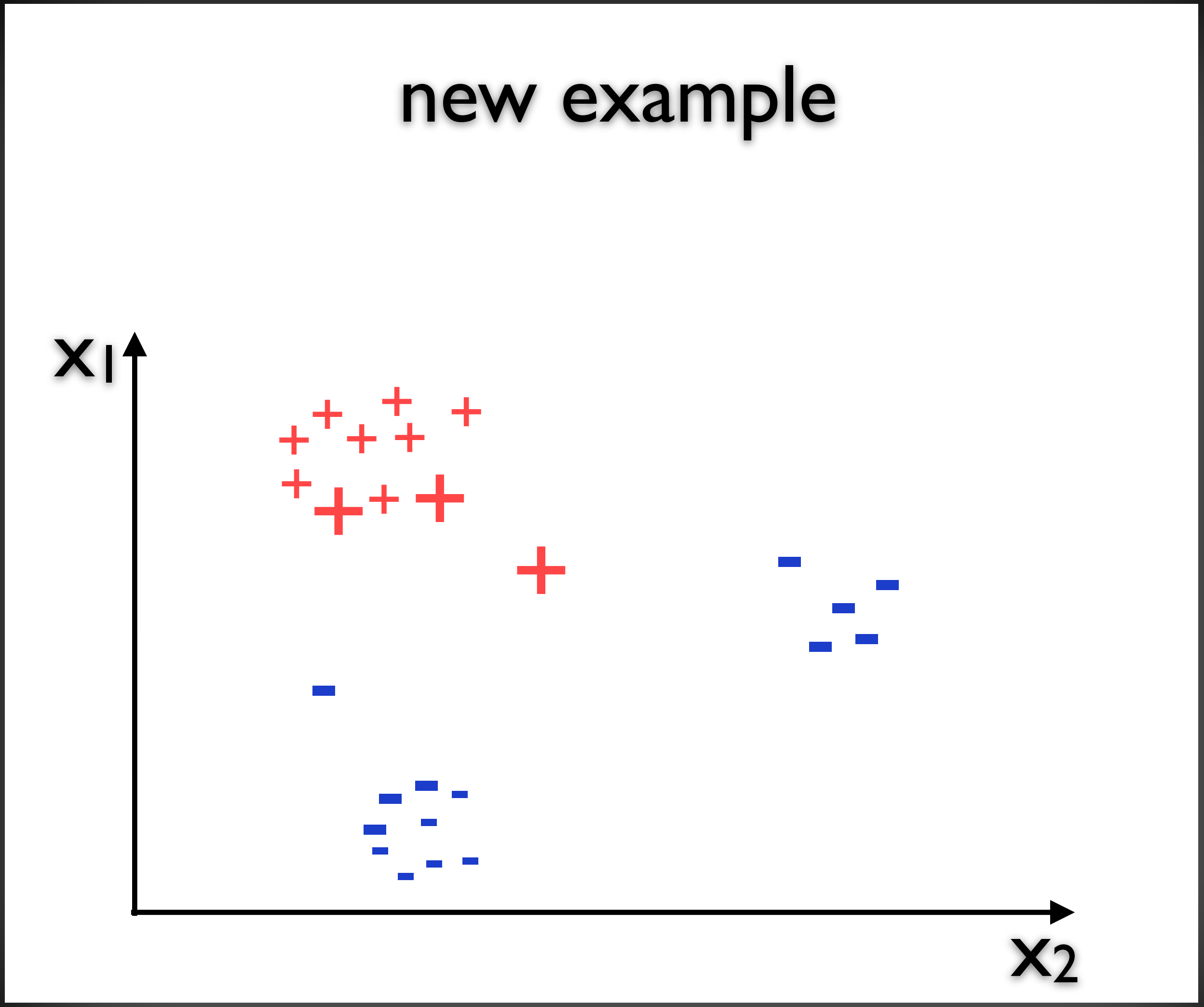
$(Y, (x_1,x_2))$, ... → learner → classifier

observations

iid distribution μ

machine learning example

who constructs the classifier?



label {red, blue}

instance:(x_1,x_2)

$(Y, (x_1,x_2))$, ...

learner

classifier

observations

iid distribution μ

target concept: unknown

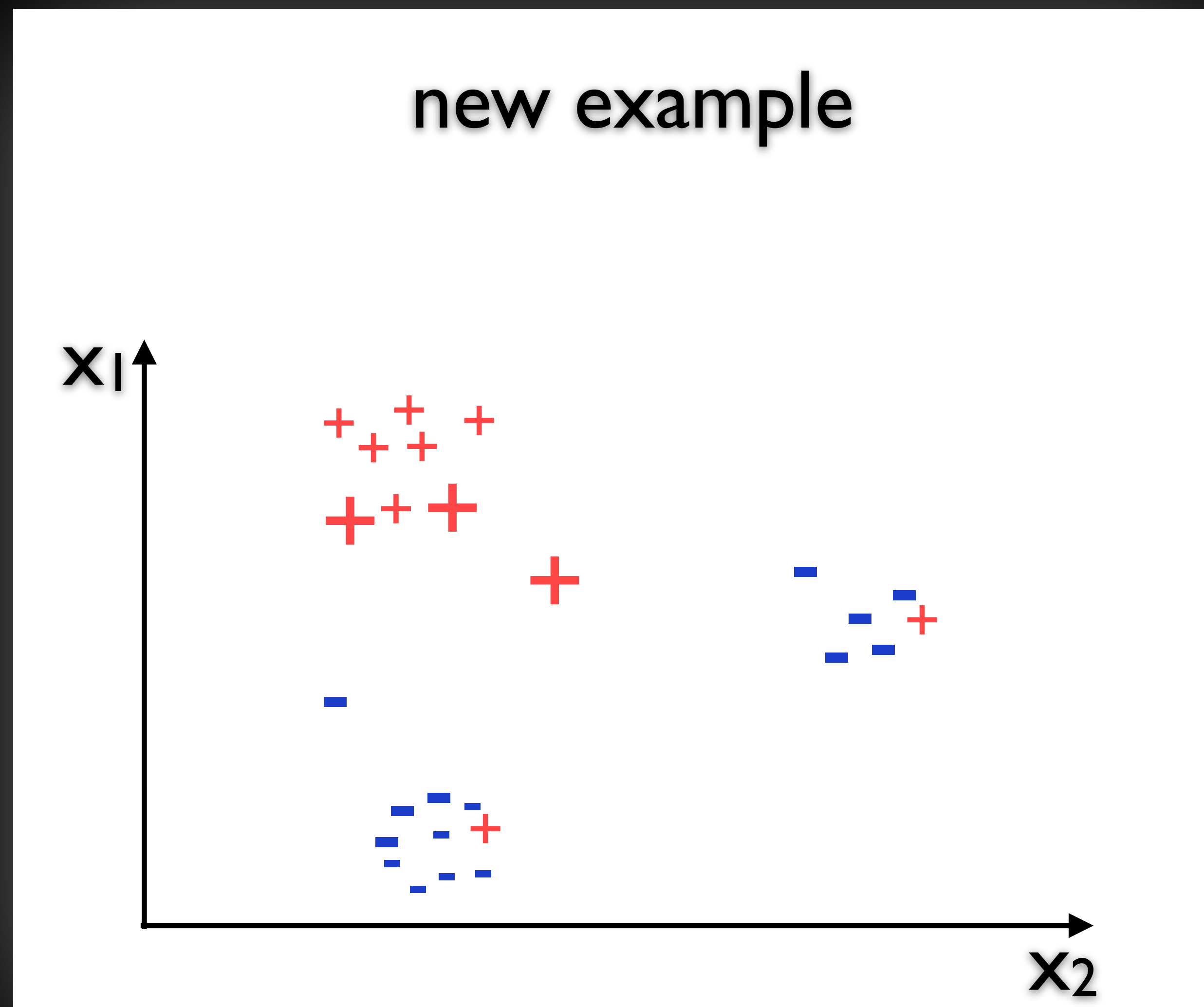
concept class: known

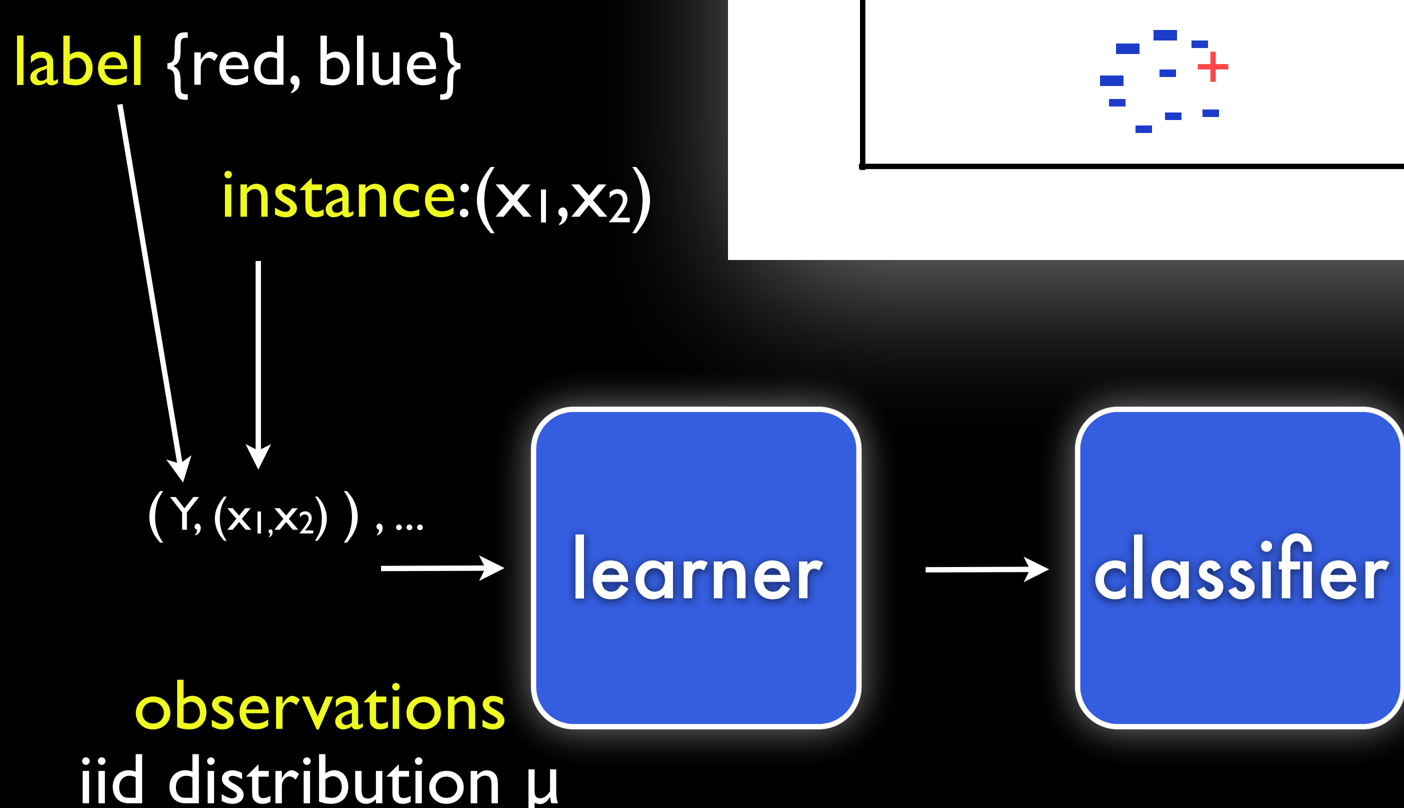
goal: find out the target

among the possible concepts

machine learning example

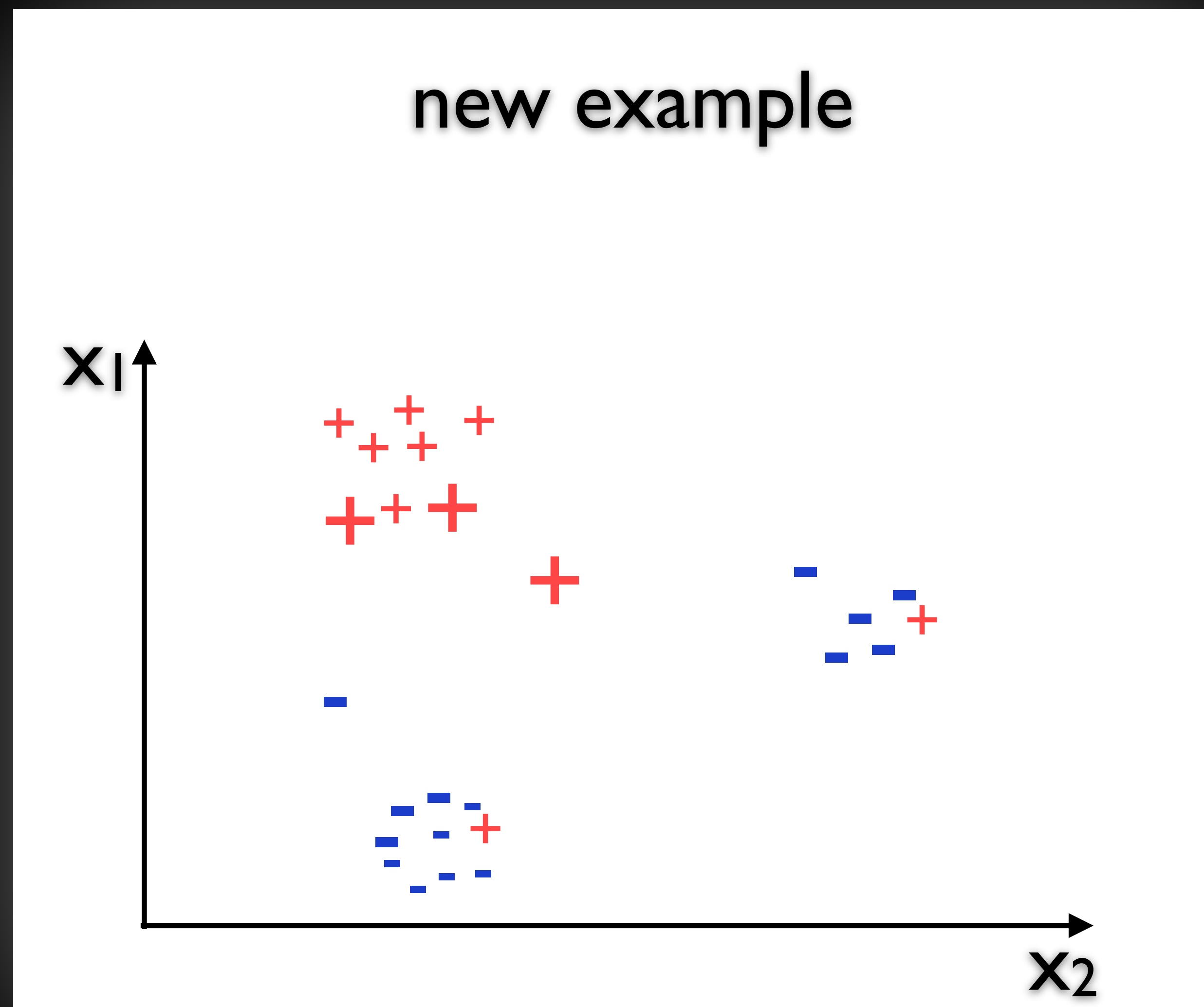
who constructs the classifier?





machine learning example

who constructs the classifier?



label {red, blue}

instance:(x_1,x_2)

$(Y, (x_1,x_2))$, ...

learner

classifier

observations

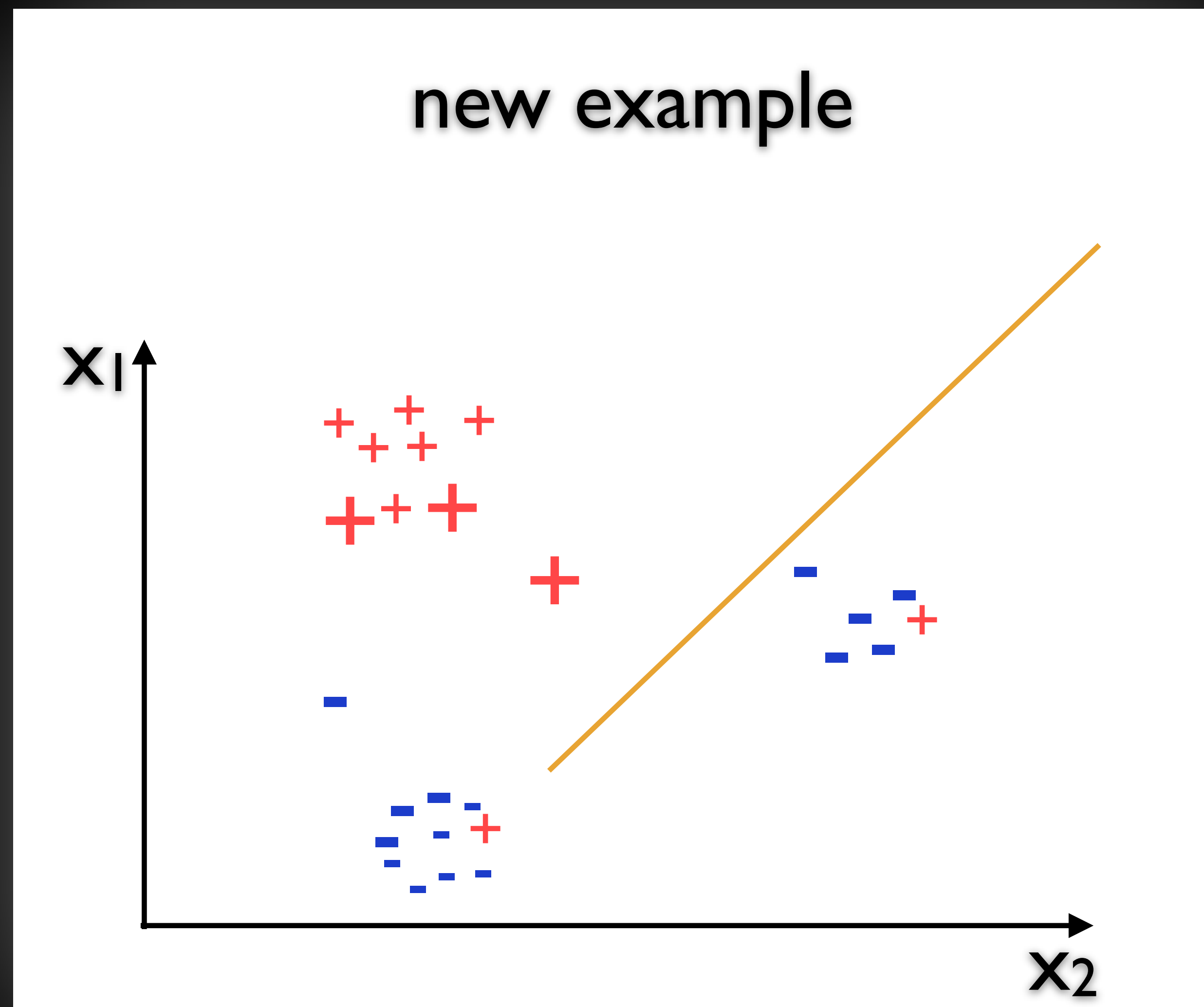
iid distribution μ

but...our learner can only learn affine lines

$x+\beta$, where $|\beta| \leq 2$

machine learning example

who constructs the classifier?



label {red, blue}

instance:(x_1,x_2)

$(Y, (x_1,x_2))$, ...

learner → classifier

observations

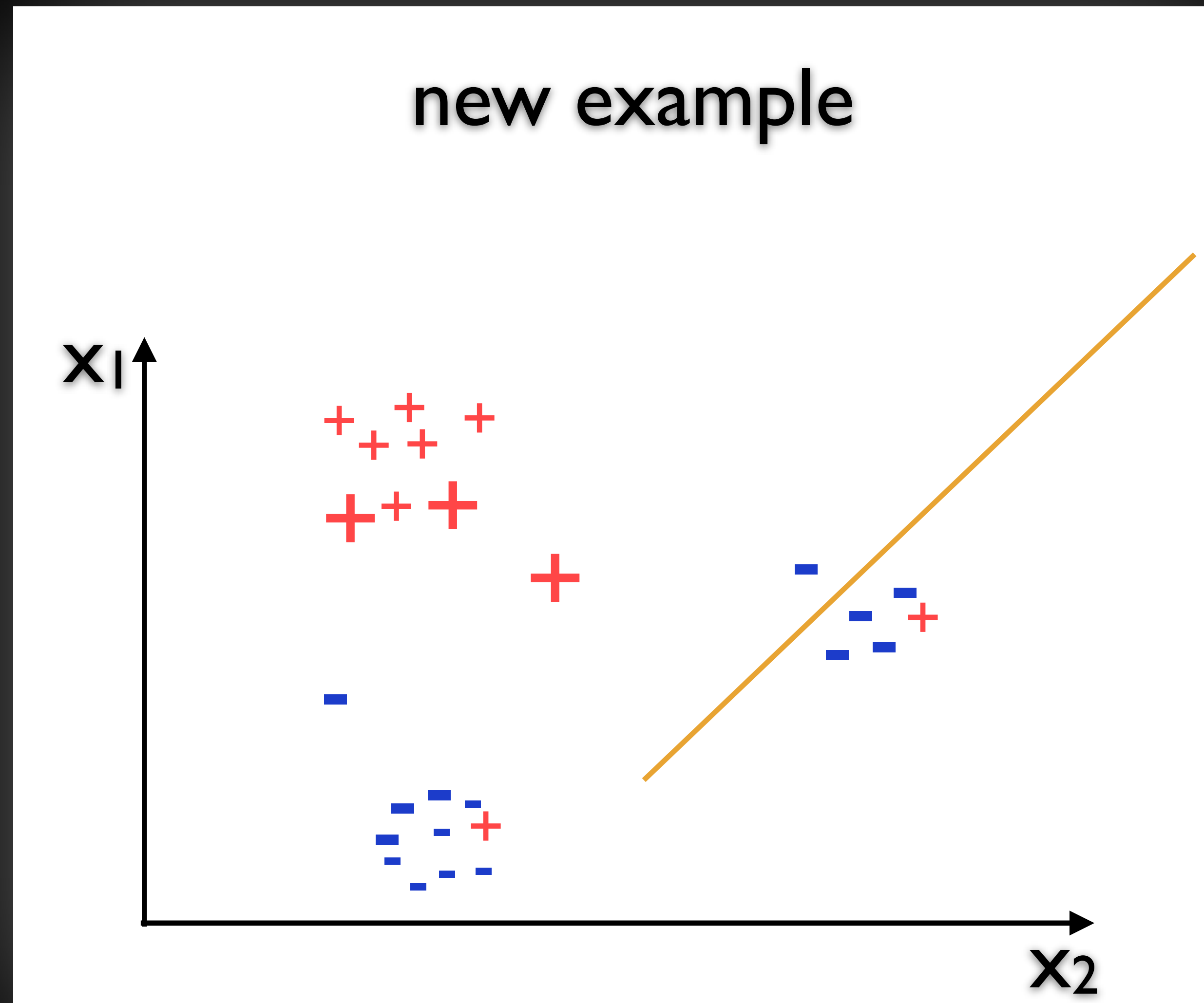
iid distribution μ

but...our learner can only learn affine lines

$x+\beta$, where $|\beta|\leq 2$

machine learning example

who constructs the classifier?



label {red, blue}

instance:(x_1,x_2)

$(Y, (x_1,x_2))$, ...

learner

classifier

observations

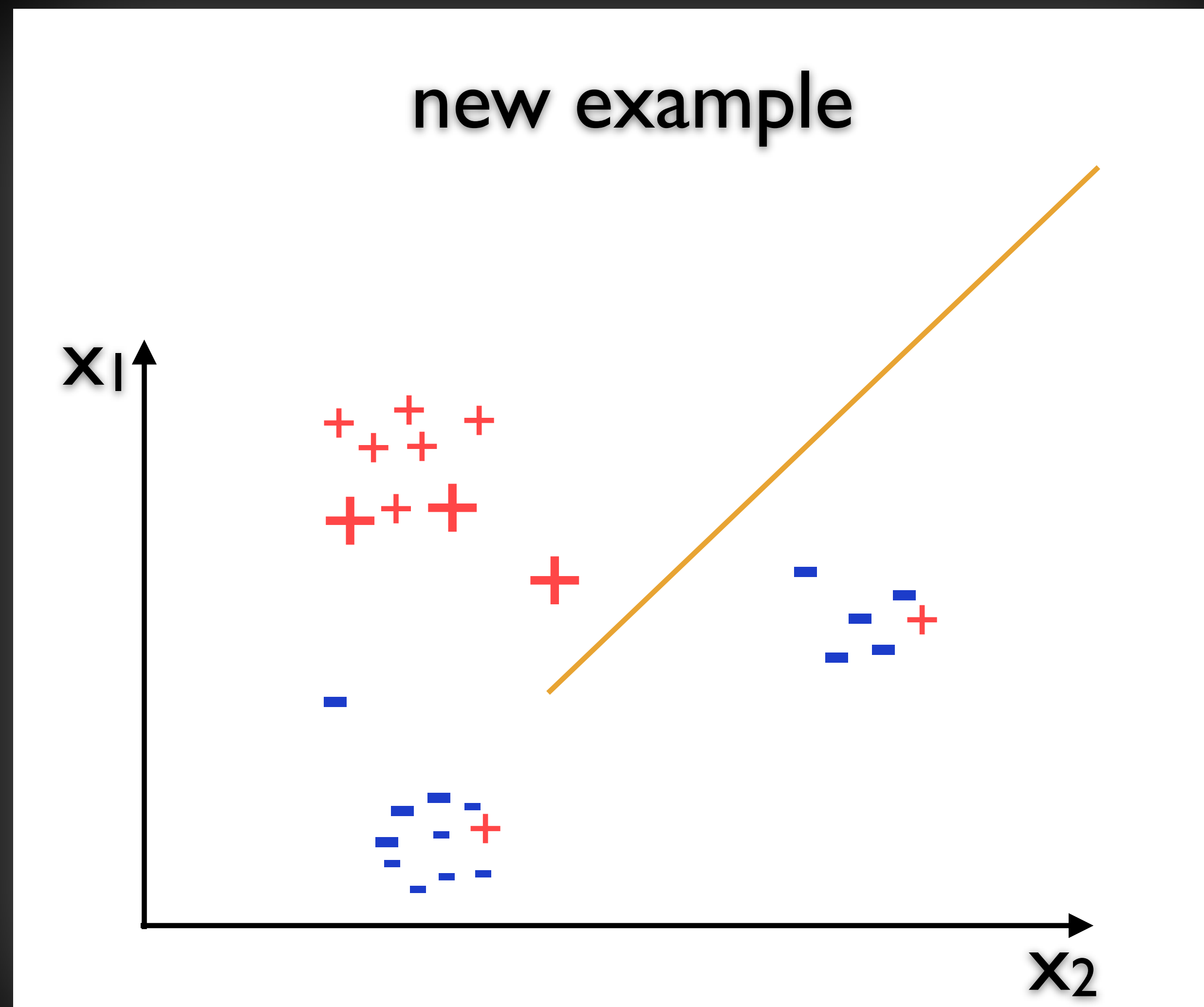
iid distribution μ

but...our learner can only learn affine lines

$x+\beta$, where $|\beta|\leq 2$

machine learning example

who constructs the classifier?



label {red, blue}

instance:(x_1,x_2)

$(Y, (x_1,x_2))$, ...

learner → classifier

observations

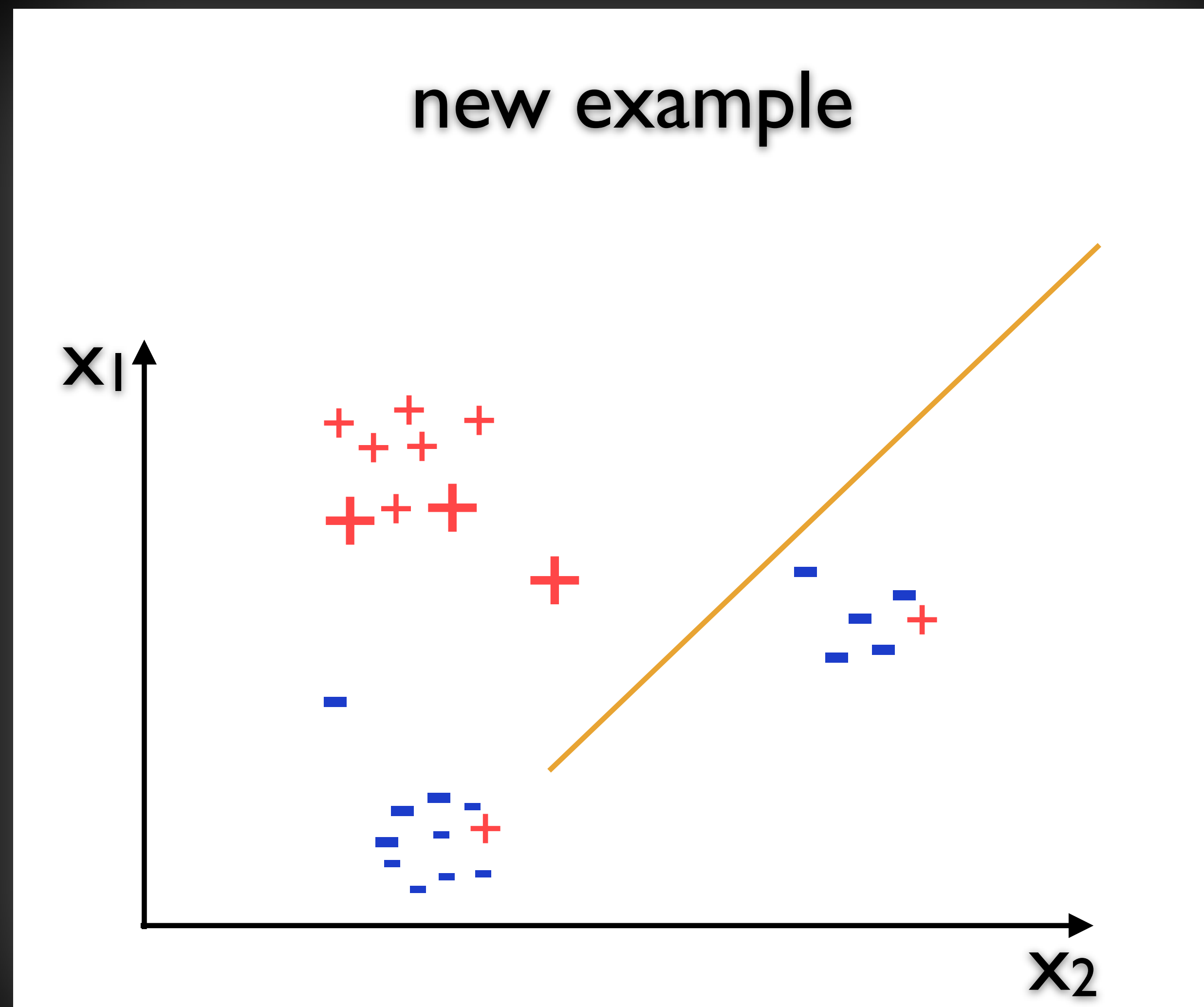
iid distribution μ

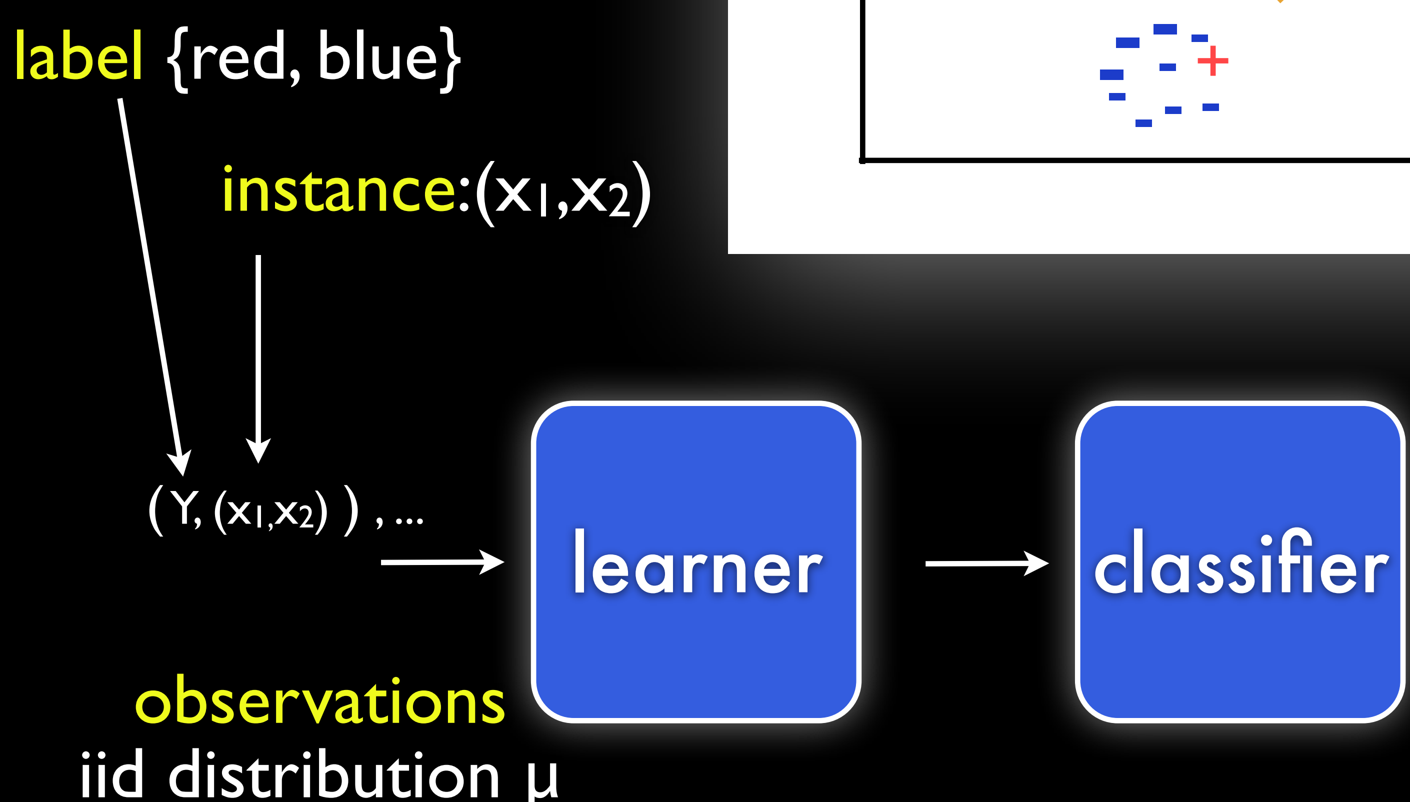
but...our learner can only learn affine lines

$x+\beta$, where $|\beta| \leq 2$

machine learning example

who constructs the classifier?



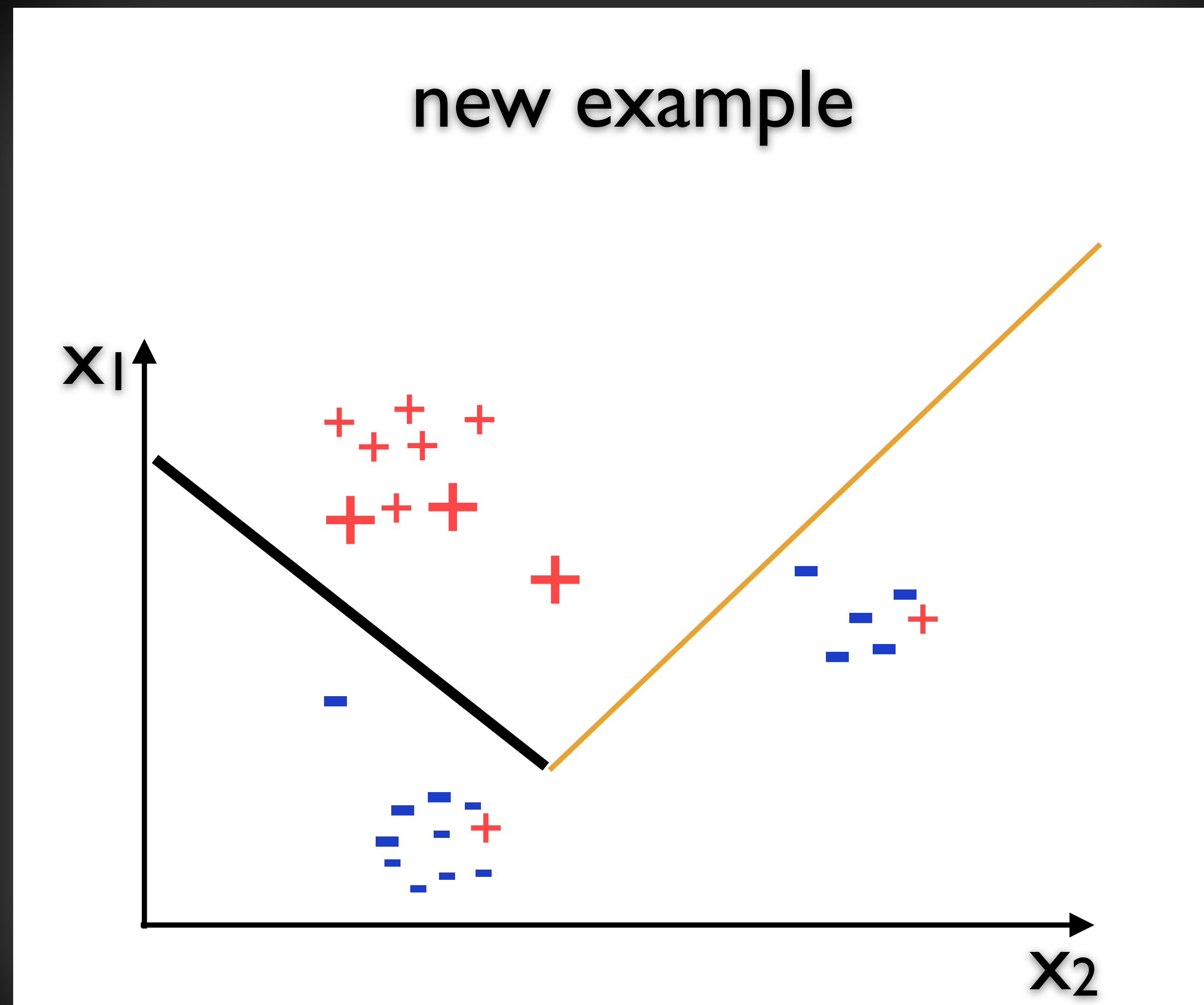


but...our learner can only learn affine lines

$x+\beta$, where $|\beta|\leq 2$

machine learning example

who constructs the classifier?



label {red, blue}

instance:(x_1,x_2)

$(Y, (x_1,x_2))$, ...

learner

classifier

observations

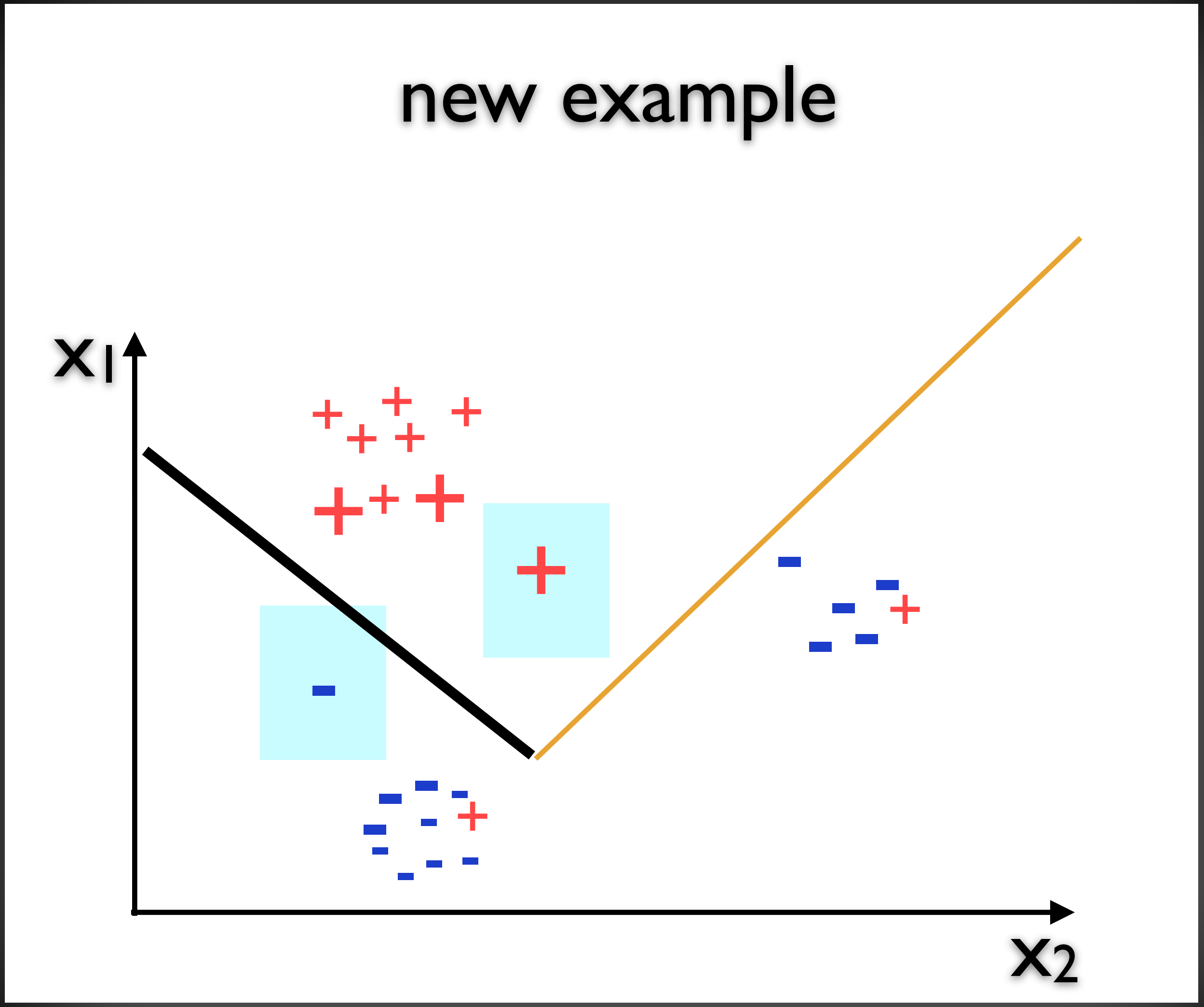
iid distribution μ

but...our learner can only learn affine lines

$x+\beta$, where $|\beta|\leq 2$

machine learning example

who constructs the classifier?



label {red, blue}

instance:(x_1,x_2)

(Y, (x_1,x_2)) , ... → learner → classifier

observations

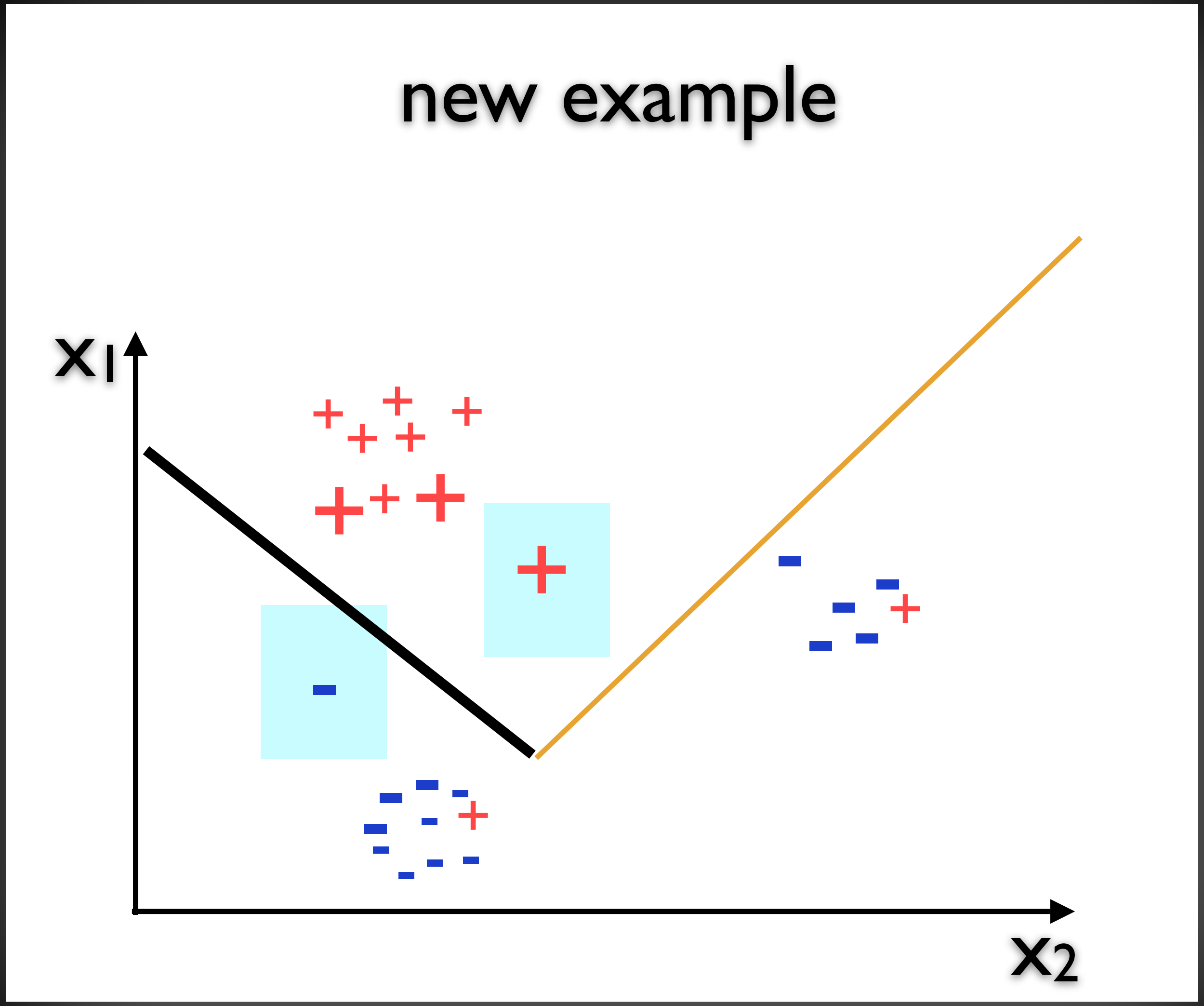
iid distribution μ

but...our learner can only learn affine lines

$x+\beta$, where $|\beta| \leq 2$

machine learning example

who constructs the classifier?



the black line has the form
$0.9x + 1.03$

but…

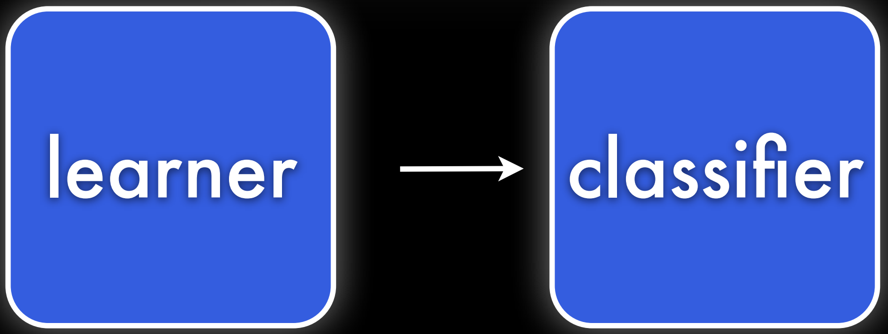
$$\begin{array}{rl} & 0.2(x+2) \\ + & 0.2(x+1.2) \\ + & 0.3(x+1.3) \\ \hline & 0.9x + 1.03 \end{array}$$

label {red, blue}

instance:(x_1,x_2)

$(Y, (x_1,x_2))$, ...

observations
iid distribution μ



but...our learner can only learn affine lines
$x+\beta$, where $|\beta|\leq 2$

No imbalanced data: yes, we care about majority

Imbalanced data: we care minority

Sometimes we care about minor

For example, 99% negative and 1% positive

We do NOT want to classify everything as negative, although it will give us 99% accuracy.

Error rate — precision and recall — true positive, false positive, true negative, false negative. F-measure —> result. 99% classify everything as negative — bad F-measure score.

workflow of machine learning

Please 101
find 33
attached 25
the 948
corrections 11
to 253
our 92
manuscript 2
...

1 (spam)

learner

classifier

workflow of machine learning

(features, spam or not) , (feature, spam or not) , ...

pool of
observations

Please 101
find 33
attached 25
the 948
corrections 11
to 253
our 92
manuscript 2
...

learner

classifier

1 (spam)

workflow of machine learning

instance

(features, spam or not) , (feature, spam or not) , ...

pool of
observations

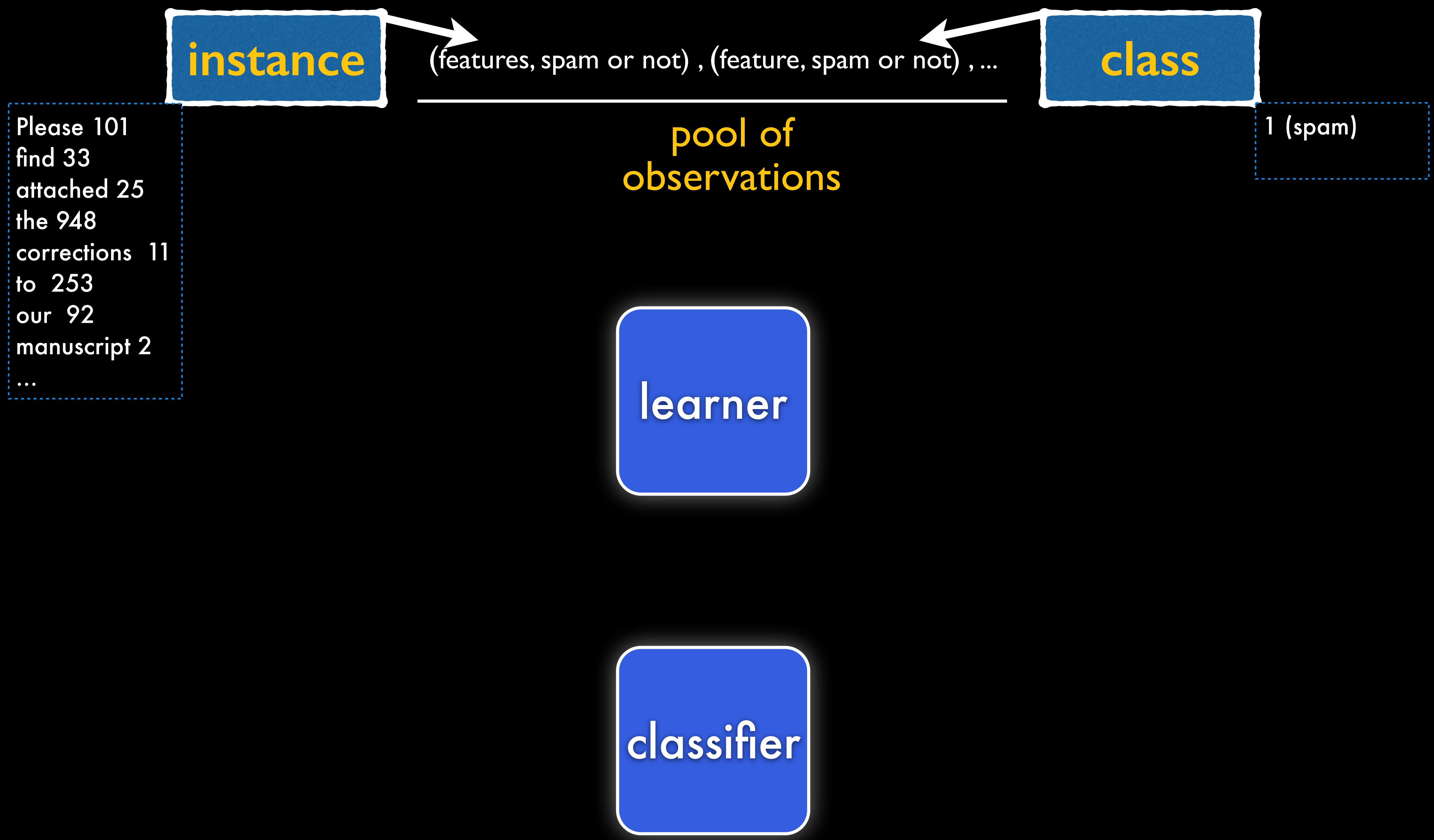
Please 101
find 33
attached 25
the 948
corrections 11
to 253
our 92
manuscript 2
...

1 (spam)

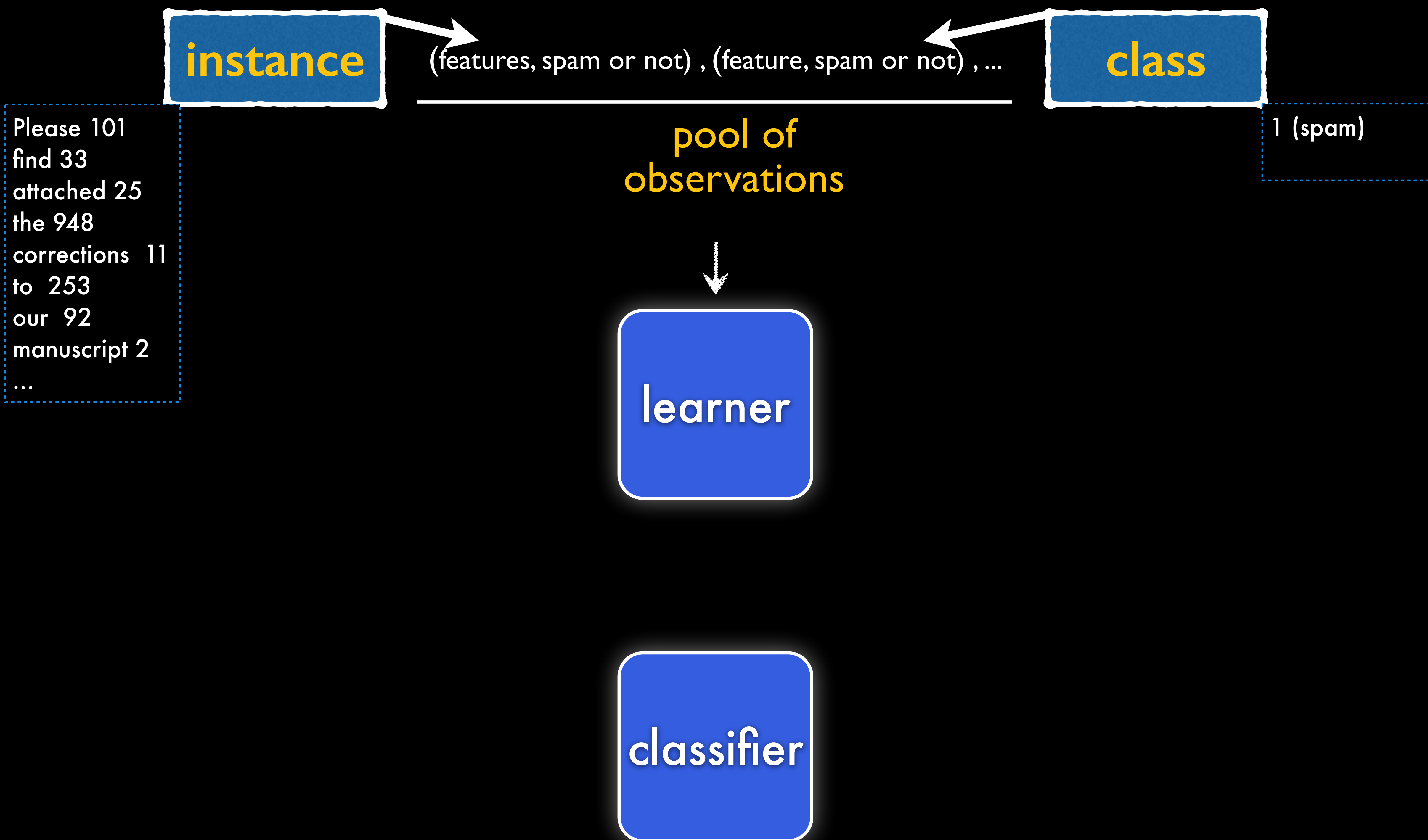
learner

classifier

workflow of machine learning



workflow of machine learning



workflow of machine learning

instance (features, spam or not) , (feature, spam or not) , … **class**

Please 101
find 33
attached 25
the 948
corrections 11
to 253
our 92
manuscript 2
…

1 (spam)

pool of observations

learner

learning method, model fitting, …

classifier

workflow of machine learning

instance

(features, spam or not) , (feature, spam or not) , ...

class

Please 101
find 33
attached 25
the 948
corrections 11
to 253
our 92
manuscript 2
...

1 (spam)

pool of
observations

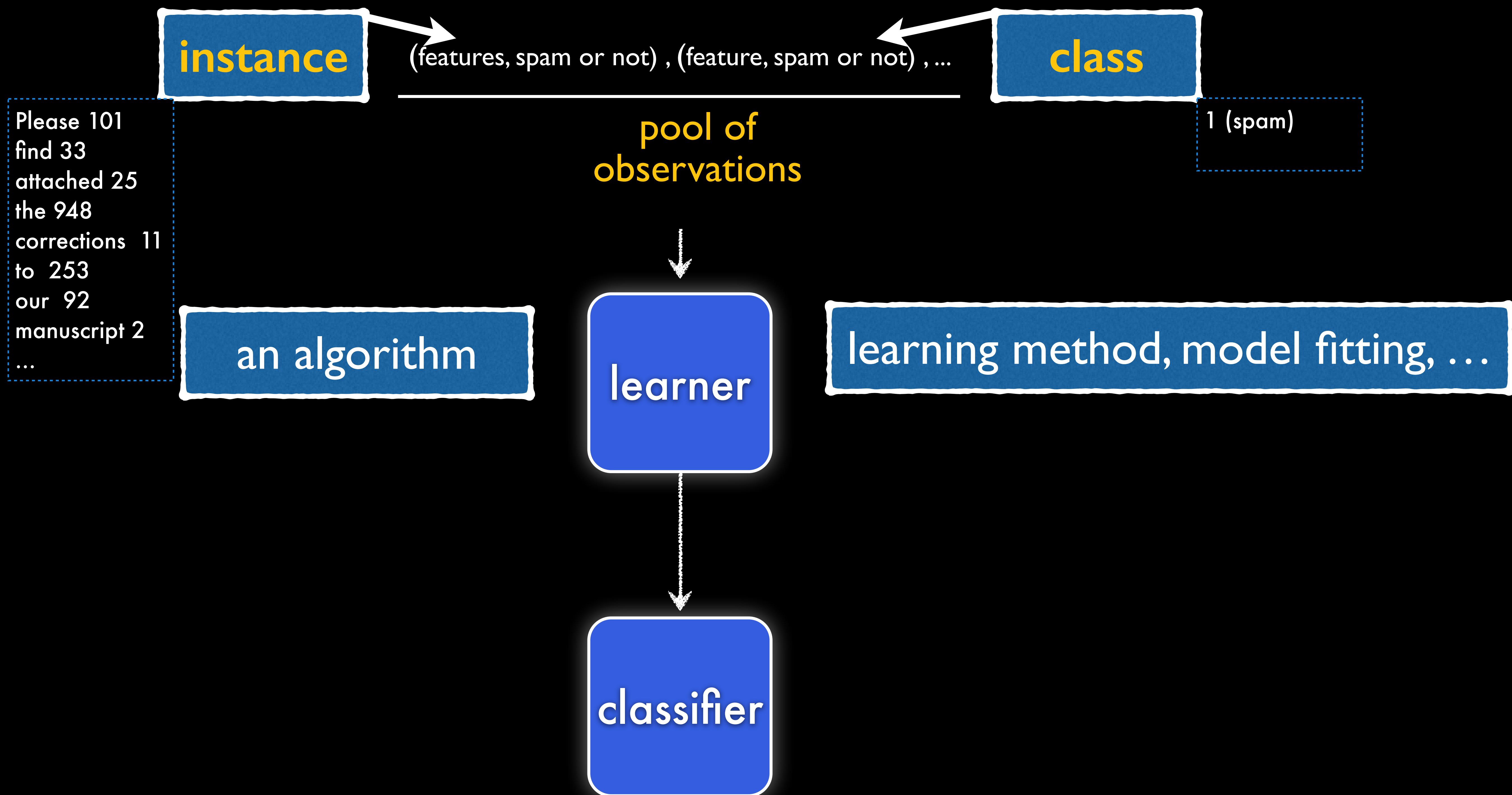
an algorithm

learner

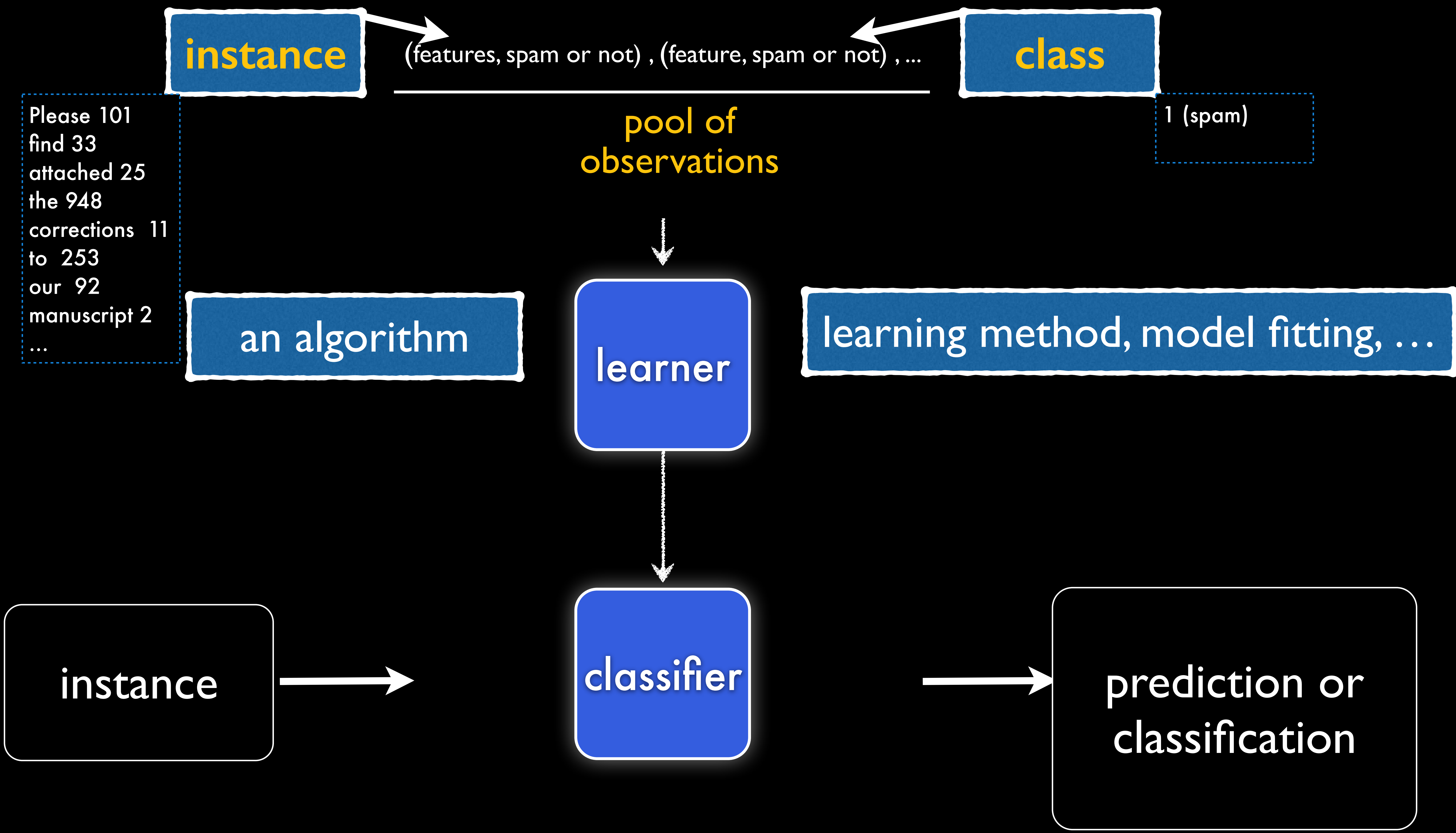
learning method, model fitting, ...

classifier

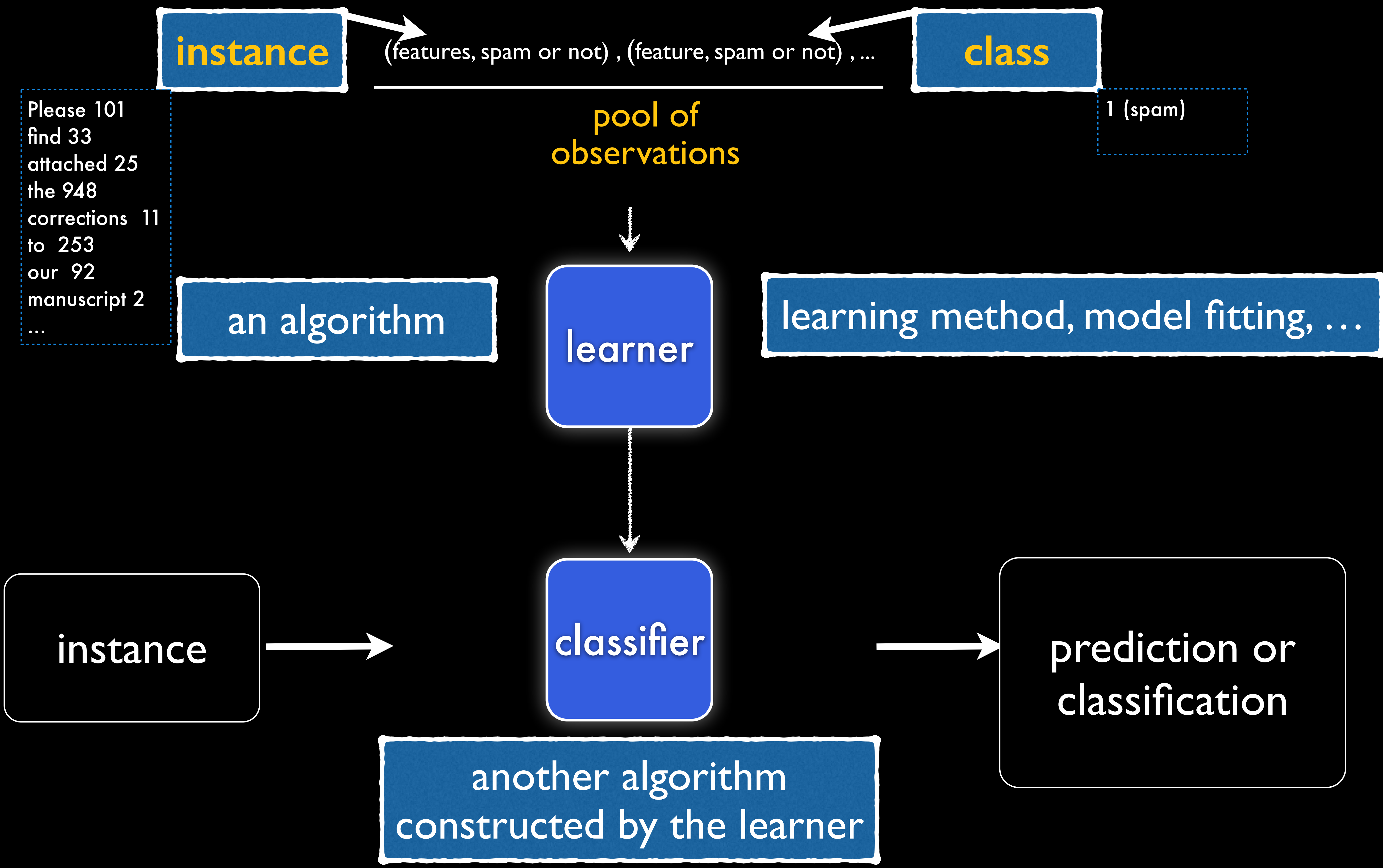
workflow of machine learning



workflow of machine learning



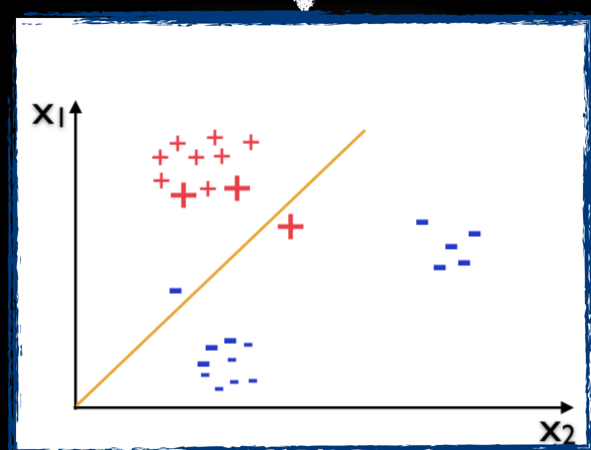
workflow of machine learning



workflow of machine learning

instance → (features, charge or not) , (feature, charge or not) , ... ← class

pool of observations



instance → classifier → prediction or classification

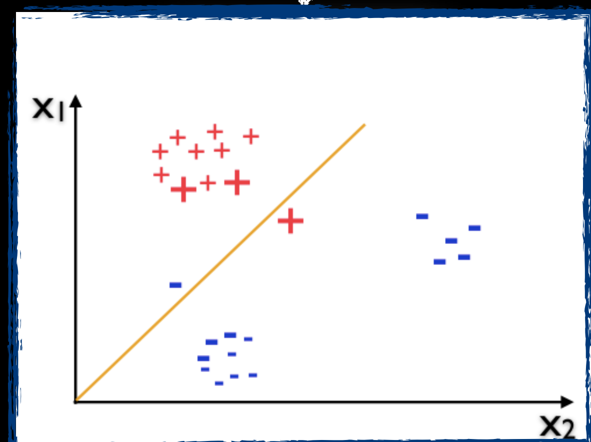
another algorithm constructed by the learner

workflow of machine learning

instance → (features, charge or not) , (feature, charge or not) , ... ← class

pool of observations

hyperplane



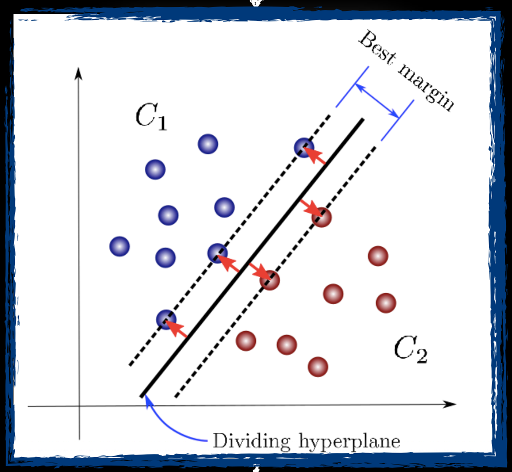
instance → classifier → prediction or classification

another algorithm constructed by the learner

workflow of machine learning

instance → (features, charge or not) , (feature, charge or not) , ... ← **class**

pool of observations



instance → **classifier** → prediction or classification

another algorithm constructed by the learner

workflow of machine learning

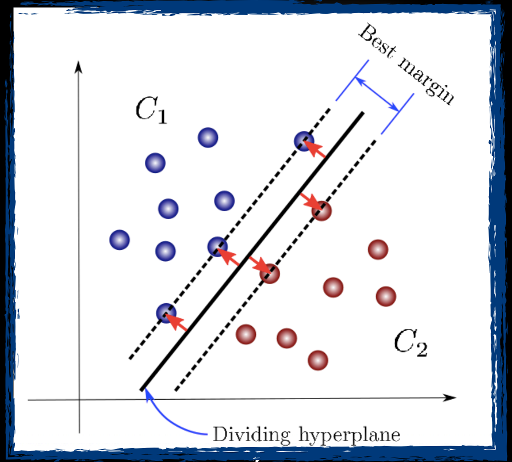
instance

(features, charge or not) , (feature, charge or not) , ...

class

pool of
observations

SVM



instance

classifier

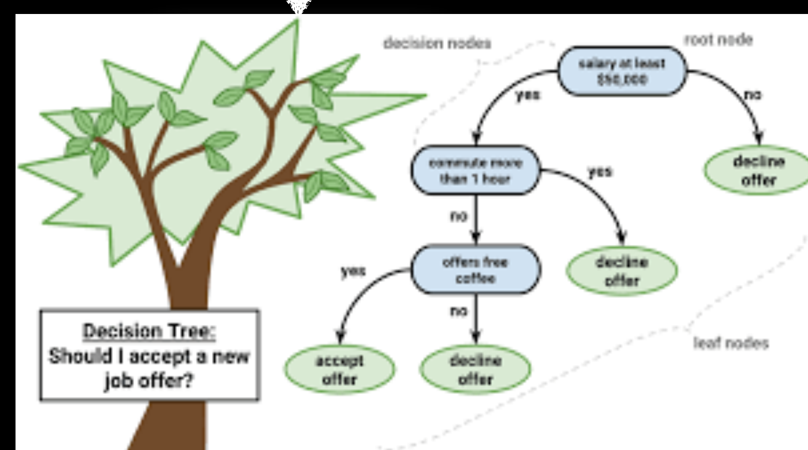
prediction or
classification

another algorithm
constructed by the learner

workflow of machine learning

instance → (features, charge or not) , (feature, charge or not) , ... ← **class**

pool of observations



instance → **classifier** → prediction or classification

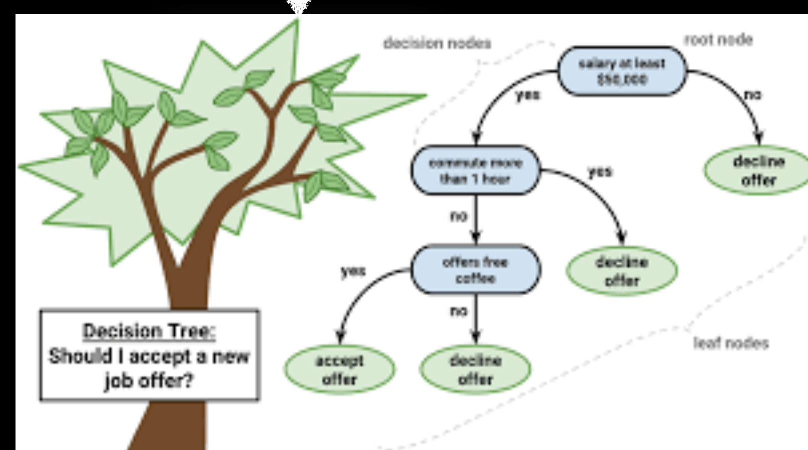
another algorithm constructed by the learner

workflow of machine learning

instance (features, charge or not) , (feature, charge or not) , ... **class**

pool of observations

Decision Tree



instance → classifier → prediction or classification

another algorithm constructed by the learner

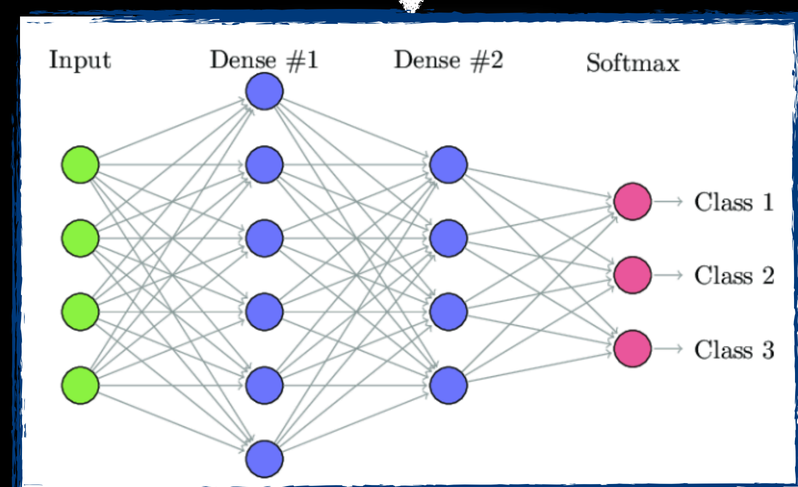
workflow of machine learning

instance

(features, charge or not) , (feature, charge or not) , ...

class

pool of
observations



instance

classifier

prediction or
classification

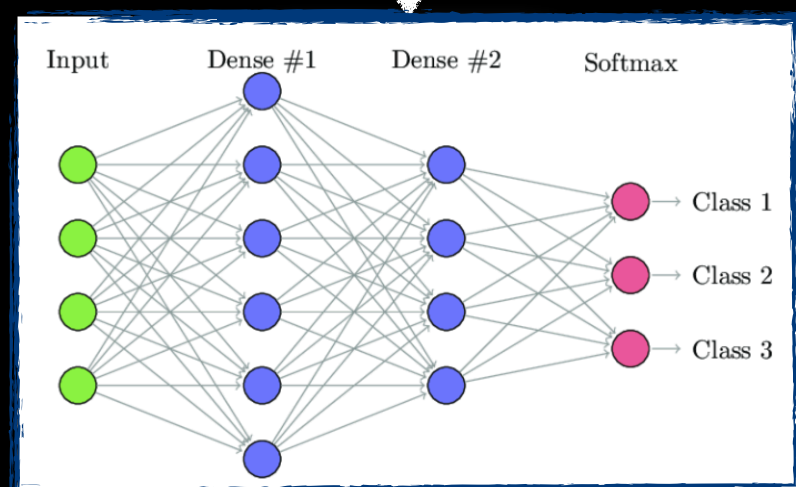
another algorithm
constructed by the learner

workflow of machine learning

instance (features, charge or not) , (feature, charge or not) , ... **class**

pool of observations

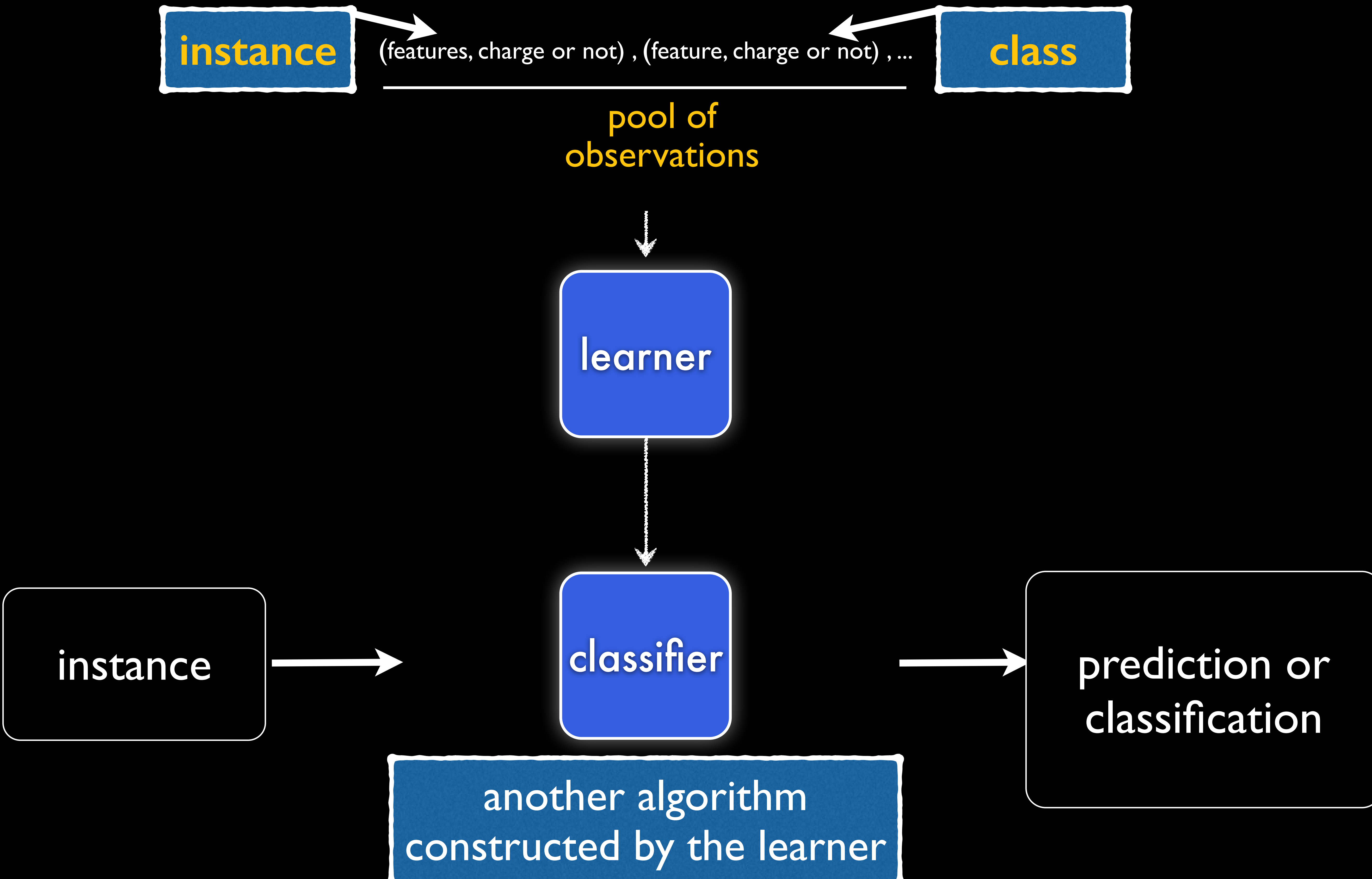
Neural Networks (NN)



instance → **classifier** → prediction or classification

another algorithm constructed by the learner

workflow of machine learning



workflow of machine learning

instance (features, charge or not) , (feature, charge or not) , ... **class**

- input: observation ***x***
- output: decision ***c***
- Bayes decision rule assuming probabilistic dependencies between ***x*** and ***c***:

$$x \to \hat{c} = \arg\max_c \left\{ pr(c|x) \right\}$$ **discriminative**

$$= \arg\max_c \left\{ pr(c) \cdot pr(x|c) \right\}$$ **generative**

instance → classifier → prediction or classification

another algorithm constructed by the learner

workflow of machine learning

instance → (features, charge or not) , (feature, charge or not) , ... ← class

- input: observation ***x***
- output: decision ***c***
- Bayes decision rule assuming probabilistic dependencies between ***x*** and ***c***:

$$
\begin{aligned}
x \rightarrow \hat{c} &= \arg\max_{c} \left\{ pr(c|x) \right\} \\
&= \arg\max_{c} \left\{ pr(c) \cdot pr(x|c) \right\}
\end{aligned}
$$

disc

ge

modeling

- define model of Pr(c|x) or Pr(c) and Pr(x|c)

learning/training

- learn model parameters

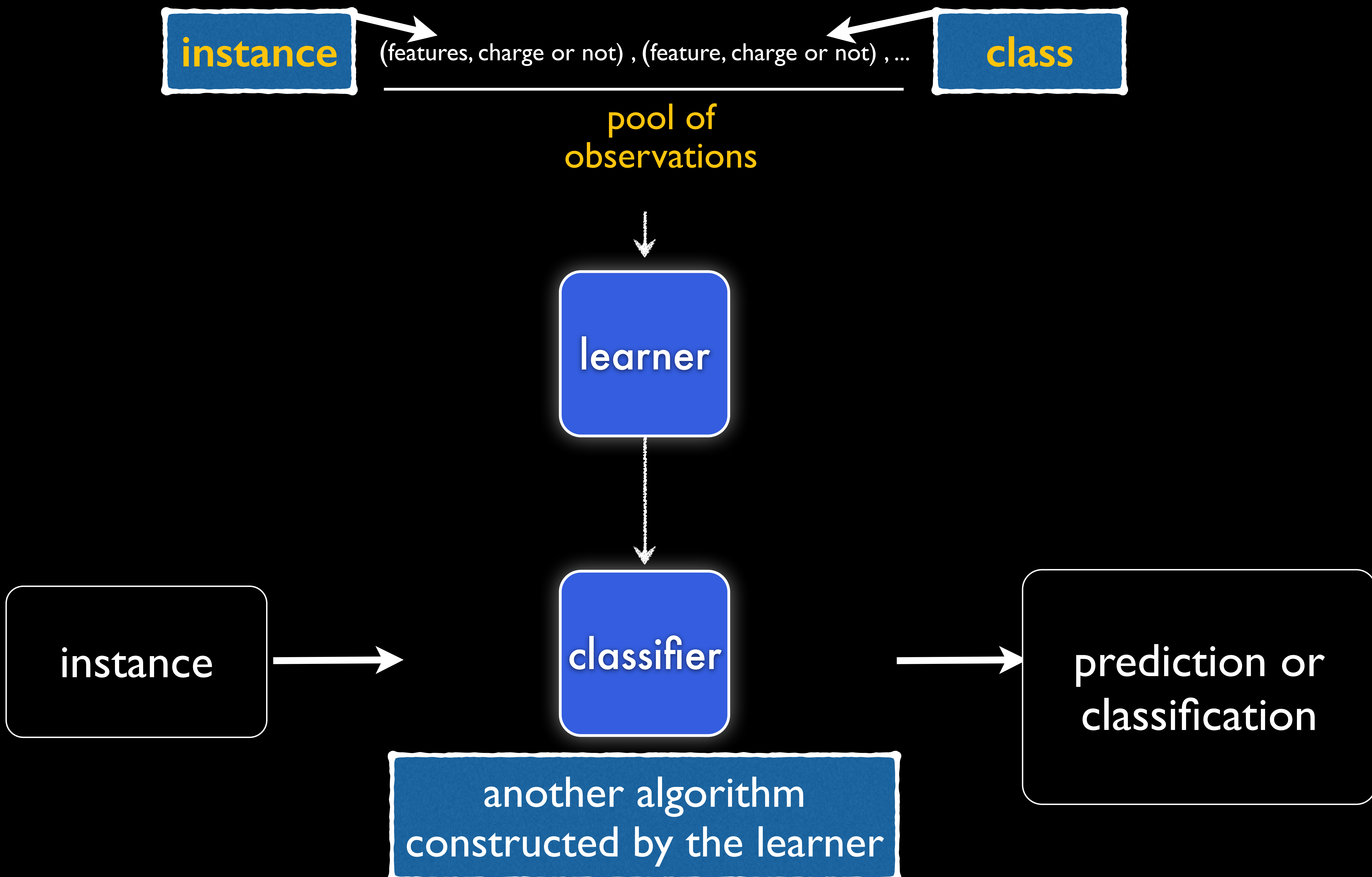
search/decoding

- find the class maximizing the probability distribution

instance → classifier → prediction or classification

another algorithm constructed by the learner

workflow of machine learning



Machine Learning

What is Machine Learning ?

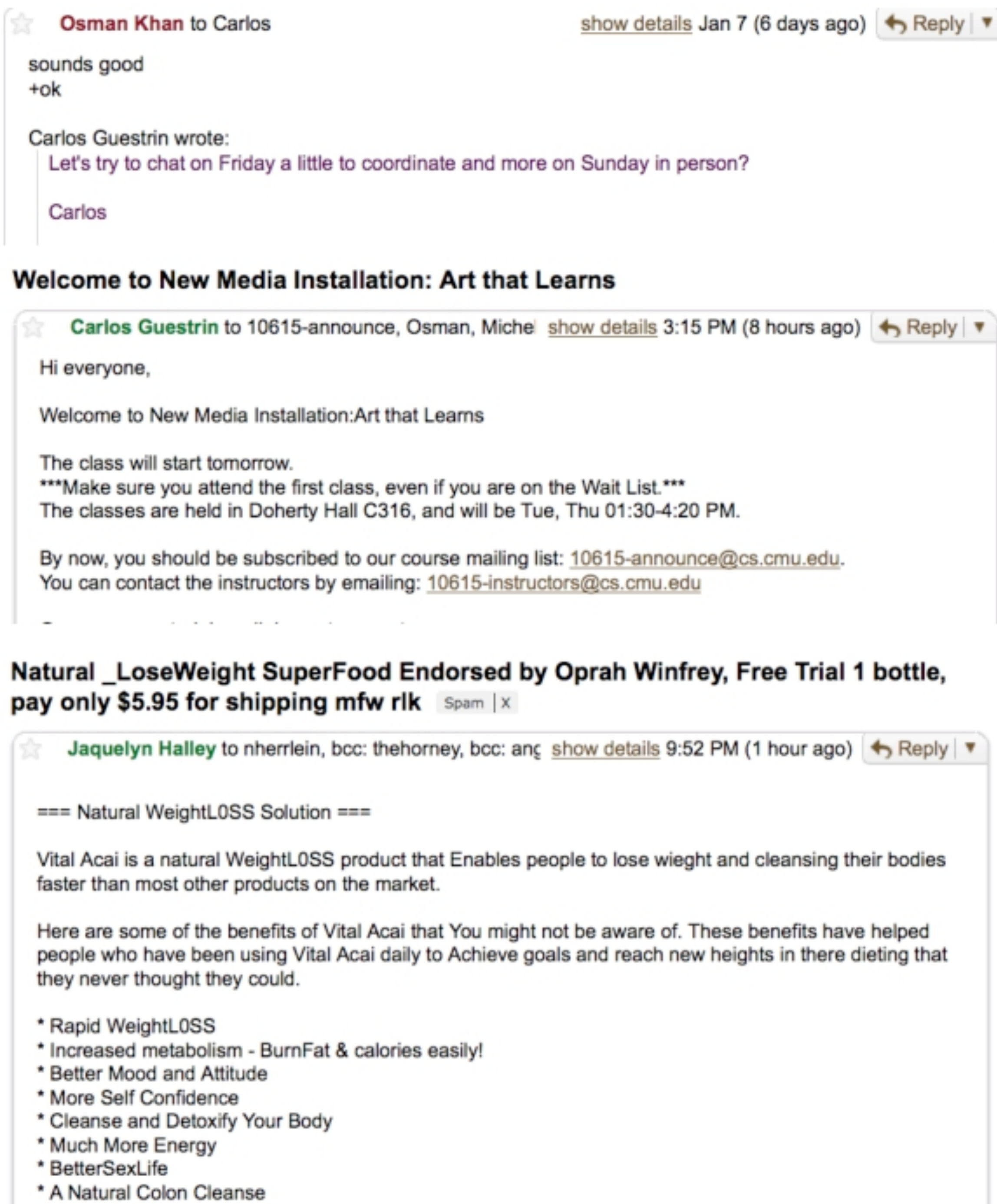
(by examples)

Classification

from data to discrete classes

Spam filtering

data



Osman Khan to Carlos — show details Jan 7 (6 days ago) — Reply

sounds good
+ok

Carlos Guestrin wrote:
Let's try to chat on Friday a little to coordinate and more on Sunday in person?

Carlos

Welcome to New Media Installation: Art that Learns

Carlos Guestrin to 10615-announce, Osman, Miche — show details 3:15 PM (8 hours ago) — Reply

Hi everyone,

Welcome to New Media Installation:Art that Learns

The class will start tomorrow.
Make sure you attend the first class, even if you are on the Wait List.
The classes are held in Doherty Hall C316, and will be Tue, Thu 01:30-4:20 PM.

By now, you should be subscribed to our course mailing list: 10615-announce@cs.cmu.edu.
You can contact the instructors by emailing: 10615-instructors@cs.cmu.edu

Natural _LoseWeight SuperFood Endorsed by Oprah Winfrey, Free Trial 1 bottle, pay only $5.95 for shipping mfw rlk Spam | X

Jaquelyn Halley to nherrlein, bcc: thehorney, bcc: ang — show details 9:52 PM (1 hour ago) — Reply

=== Natural WeightL0SS Solution ===

Vital Acai is a natural WeightL0SS product that Enables people to lose wieght and cleansing their bodies faster than most other products on the market.

Here are some of the benefits of Vital Acai that You might not be aware of. These benefits have helped people who have been using Vital Acai daily to Achieve goals and reach new heights in there dieting that they never thought they could.

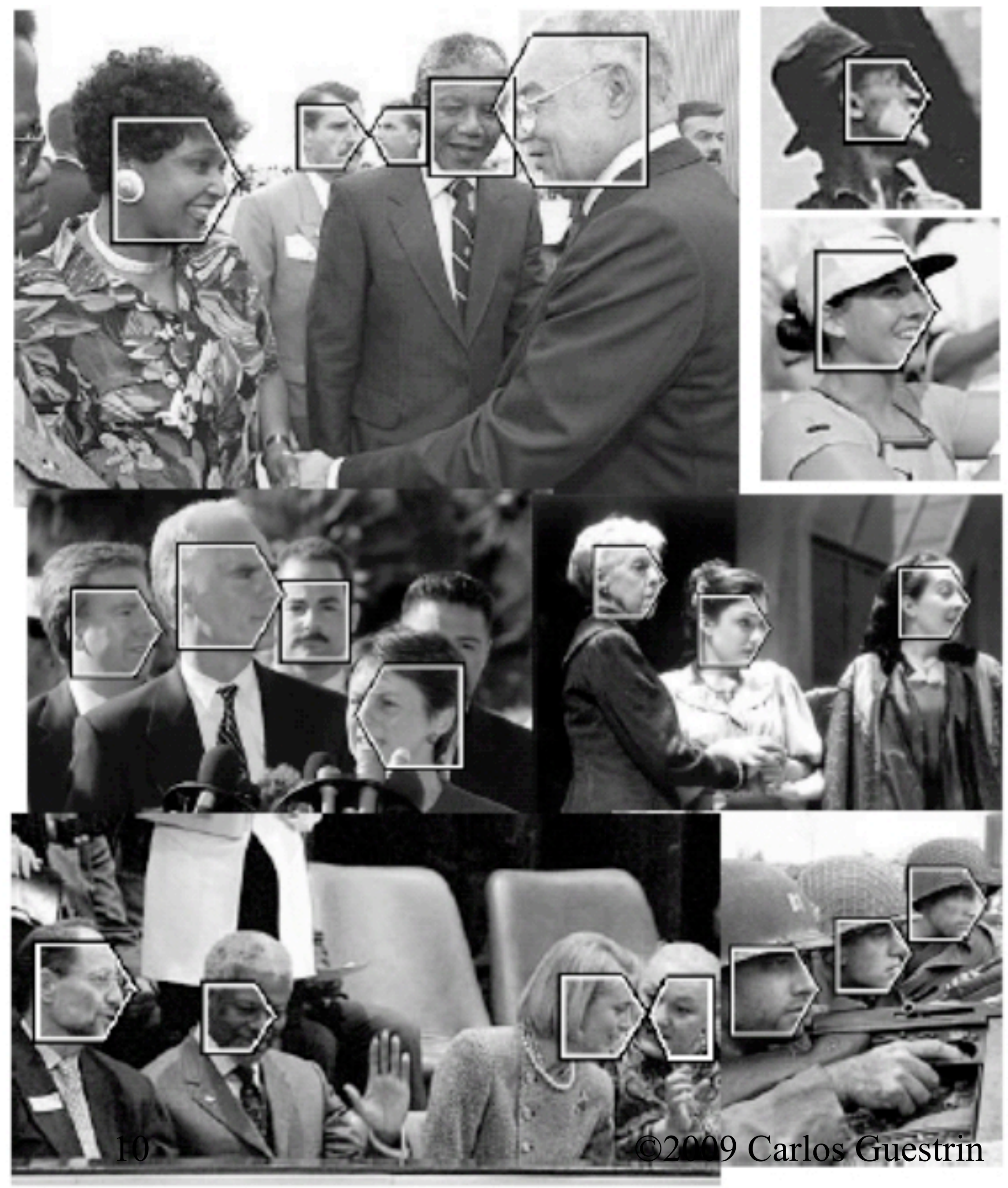
* Rapid WeightL0SS
* Increased metabolism - BurnFat & calories easily!
* Better Mood and Attitude
* More Self Confidence
* Cleanse and Detoxify Your Body
* Much More Energy
* BetterSexLife
* A Natural Colon Cleanse

prediction

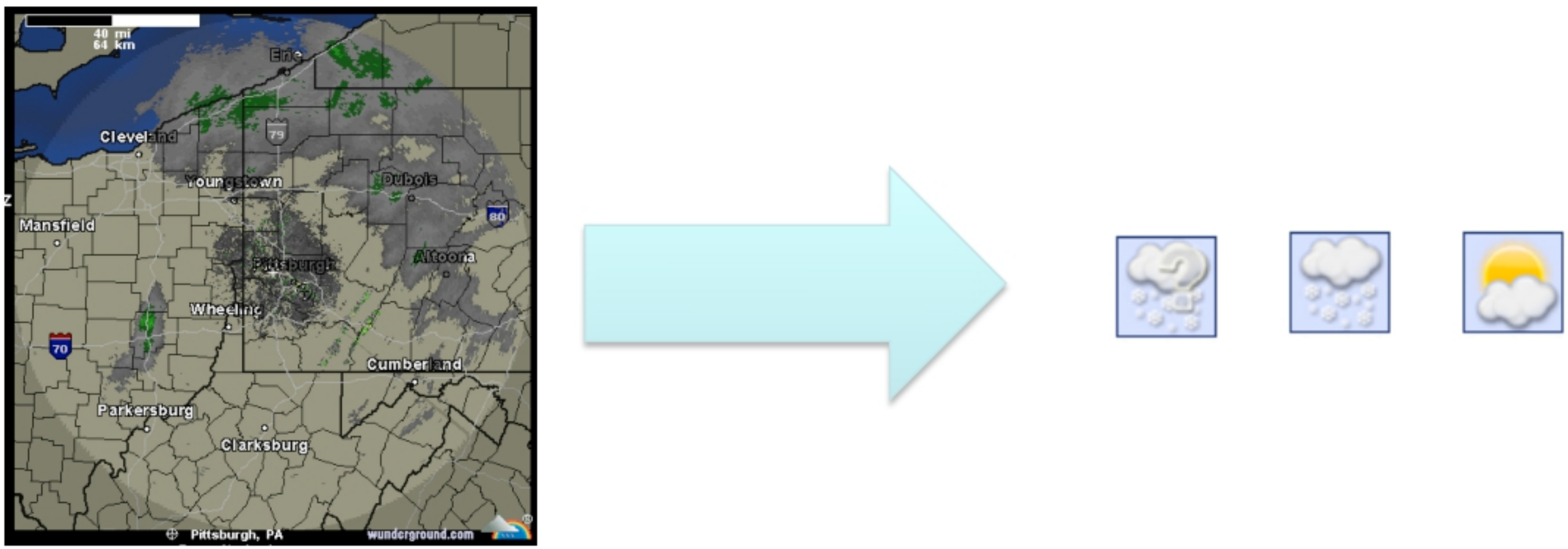
Spam
vs.
Not Spam

Face recognition

Example training images
for each orientation



Weather prediction



Regression

predicting a numeric value

Stock market

Weather prediction revisited

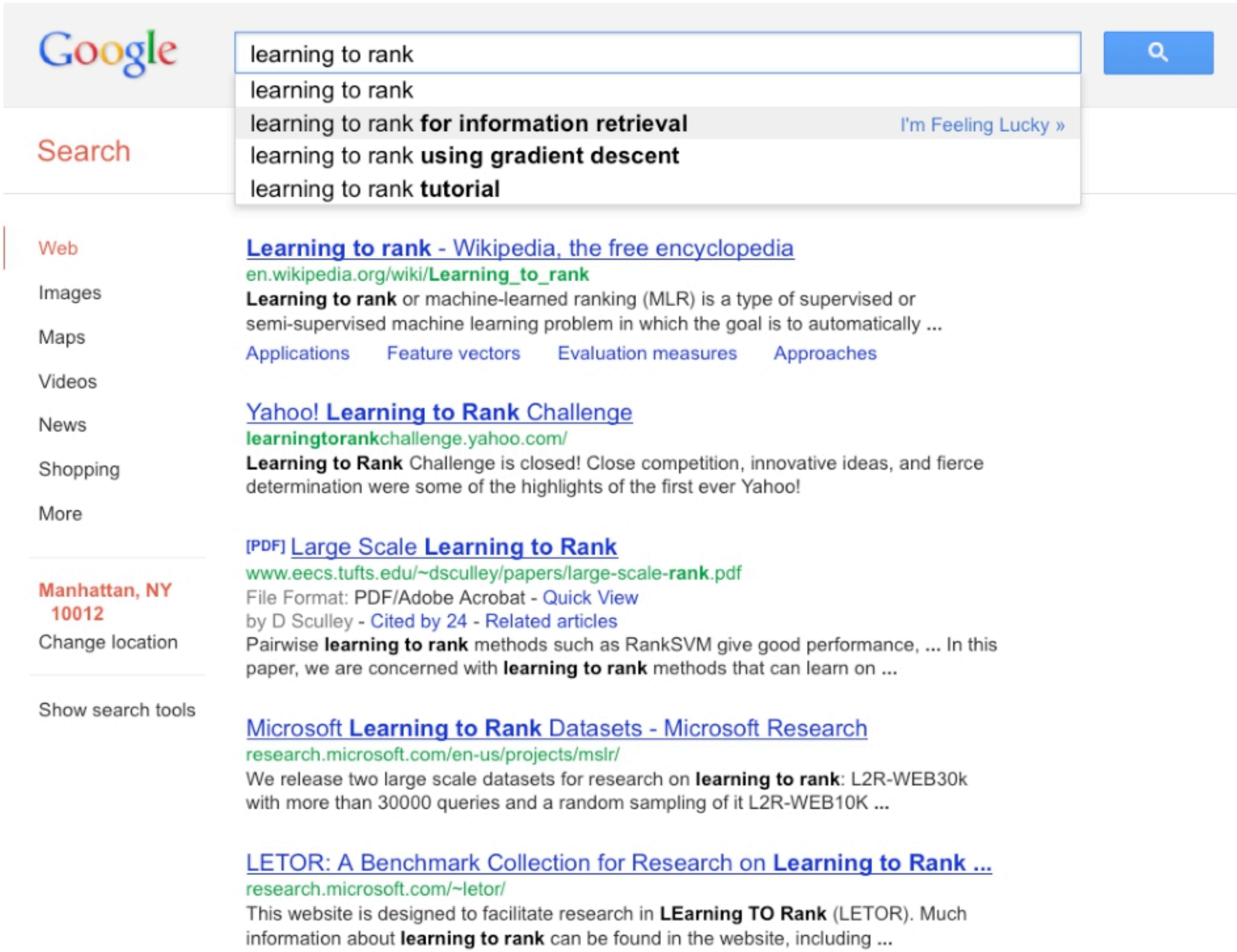
Temperature

72° F

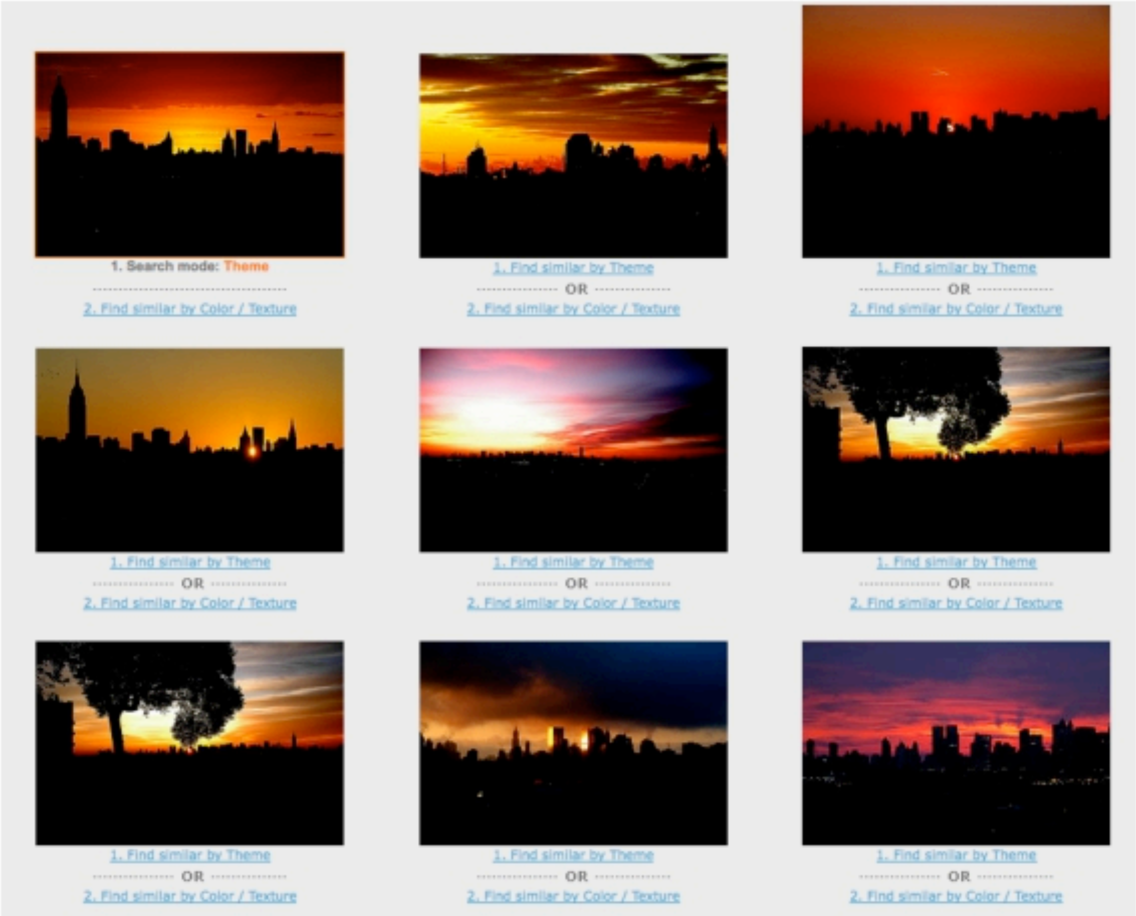
Ranking

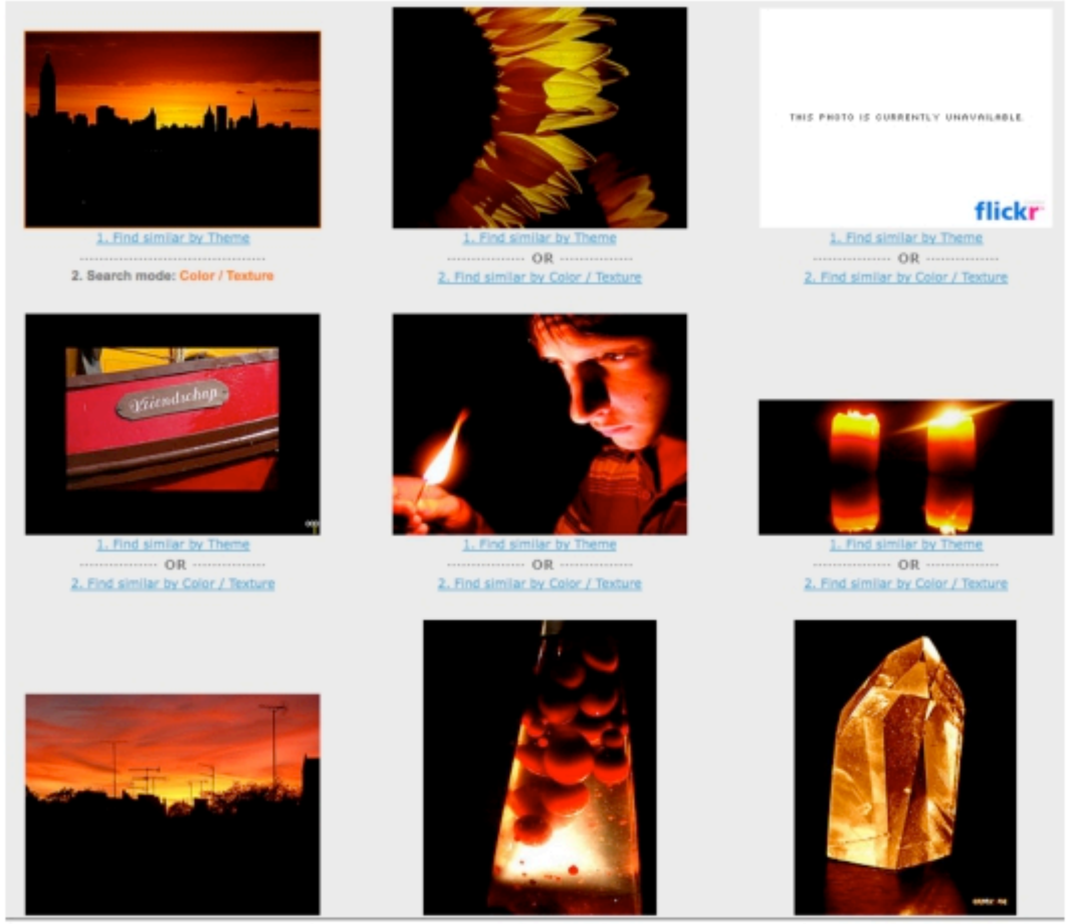
comparing items

Web search


learning to rank
learning to rank
learning to rank for information retrieval
I'm Feeling Lucky »
learning to rank using gradient descent
learning to rank tutorial
Search
Web
Images
Maps
Videos
News
Shopping
More
Manhattan, NY 10012
Change location
Show search tools
Learning to rank - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Learning_to_rank
Learning to rank or machine-learned ranking (MLR) is a type of supervised or semi-supervised machine learning problem in which the goal is to automatically ...
Applications
Feature vectors
Evaluation measures
Approaches
Yahoo! Learning to Rank Challenge
learningtorankchallenge.yahoo.com/
Learning to Rank Challenge is closed! Close competition, innovative ideas, and fierce determination were some of the highlights of the first ever Yahoo!
[PDF] Large Scale Learning to Rank
www.eecs.tufts.edu/~dsculley/papers/large-scale-rank.pdf
File Format: PDF/Adobe Acrobat - Quick View
by D Sculley - Cited by 24 - Related articles
Pairwise learning to rank methods such as RankSVM give good performance, ... In this paper, we are concerned with learning to rank methods that can learn on ...
Microsoft Learning to Rank Datasets - Microsoft Research
research.microsoft.com/en-us/projects/mslr/
We release two large scale datasets for research on learning to rank: L2R-WEB30k with more than 30000 queries and a random sampling of it L2R-WEB10K ...
LETOR: A Benchmark Collection for Research on Learning to Rank ...
research.microsoft.com/~letor/
This website is designed to facilitate research in LEarning TO Rank (LETOR). Much information about learning to rank can be found in the website, including ...

Given image, find similar images

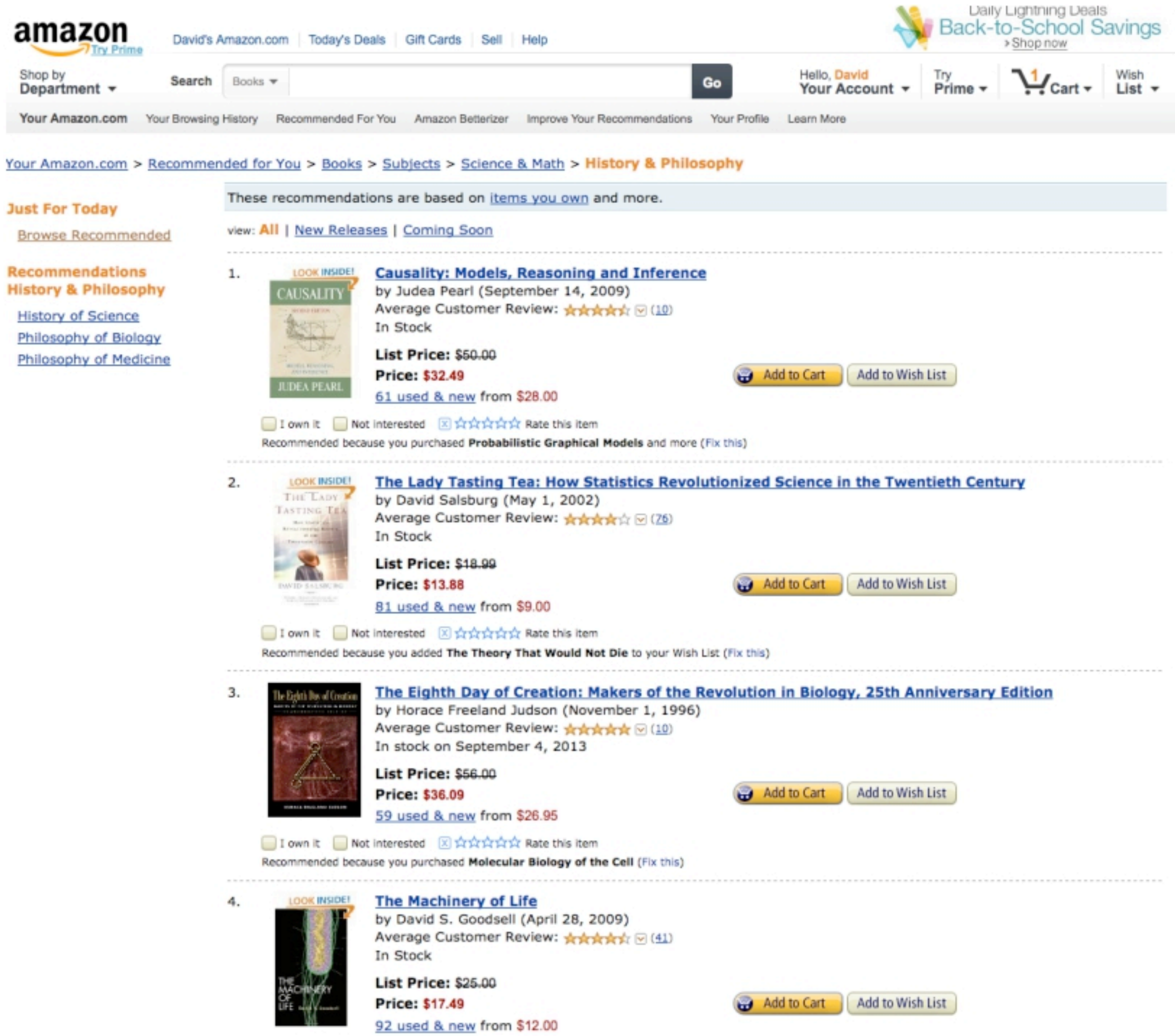




http://www.tiltomo.com/

Collaborative Filtering

Recommendation systems



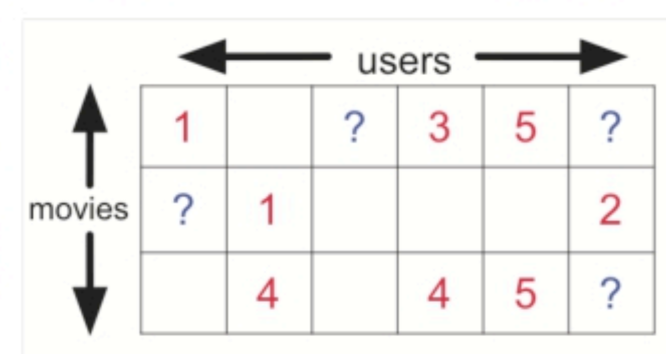
Recommendation systems

Machine learning competition with a $1 million prize

Leaderboard

Display top 20 leaders.

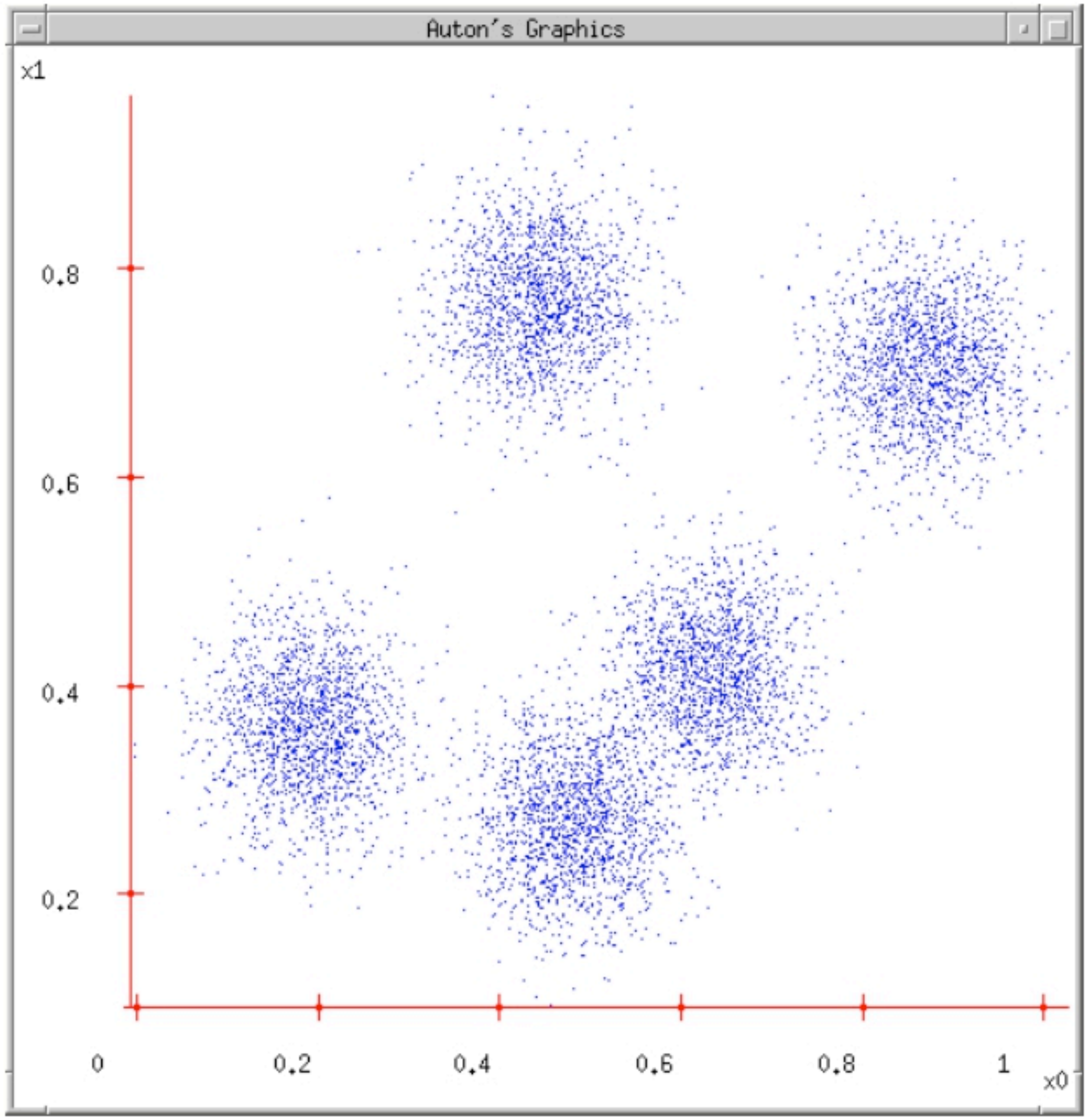
Rank	Team Name	Best Score	% Improvement	Last Submit Time
1	The Ensemble	0.8553	10.10	2009-07-26 18:38:22
2	BellKor's Pragmatic Chaos	0.8554	10.09	2009-07-26 18:18:28
Grand Prize - RMSE <= 0.8563				
3	Grand Prize Team	0.8571	9.91	2009-07-24 13:07:49
4	Opera Solutions and Vandelay United	0.8573	9.89	2009-07-25 20:05:52
5	Vandelay Industries !	0.8579	9.83	2009-07-26 02:49:53
6	PragmaticTheory	0.8582	9.80	2009-07-12 15:09:53
7	BellKor in BigChaos	0.8590	9.71	2009-07-26 12:57:25
8	Dace_	0.8603	9.58	2009-07-24 17:18:43
9	Opera Solutions	0.8611	9.49	2009-07-26 18:02:08
10	BellKor	0.8612	9.48	2009-07-26 17:19:11
11	BigChaos	0.8613	9.47	2009-06-23 23:06:52
12	Feeds2	0.8613	9.47	2009-07-24 20:06:46
Progress Prize 2008 - RMSE = 0.8616 - Winning Team: BellKor in BigChaos				
13	xiangliang	0.8633	9.26	2009-07-21 02:04:40
14	Gravity	0.8634	9.25	2009-07-26 15:58:34
15	Ces	0.8642	9.17	2009-07-25 17:42:38
16	Invisible Ideas	0.8644	9.14	2009-07-20 03:26:12
17	Just a guy in a garage	0.8650	9.08	2009-07-22 14:10:42
18	Craig Carmichael	0.8656	9.02	2009-07-25 16:00:54
19	J Dennis Su	0.8658	9.00	2009-03-11 09:41:54
20	acmehill	0.8659	8.99	2009-04-16 06:29:35
Progress Prize 2007 - RMSE = 0.8712 - Winning Team: KorBell				
Cinematch score on quiz subset - RMSE = 0.9514				



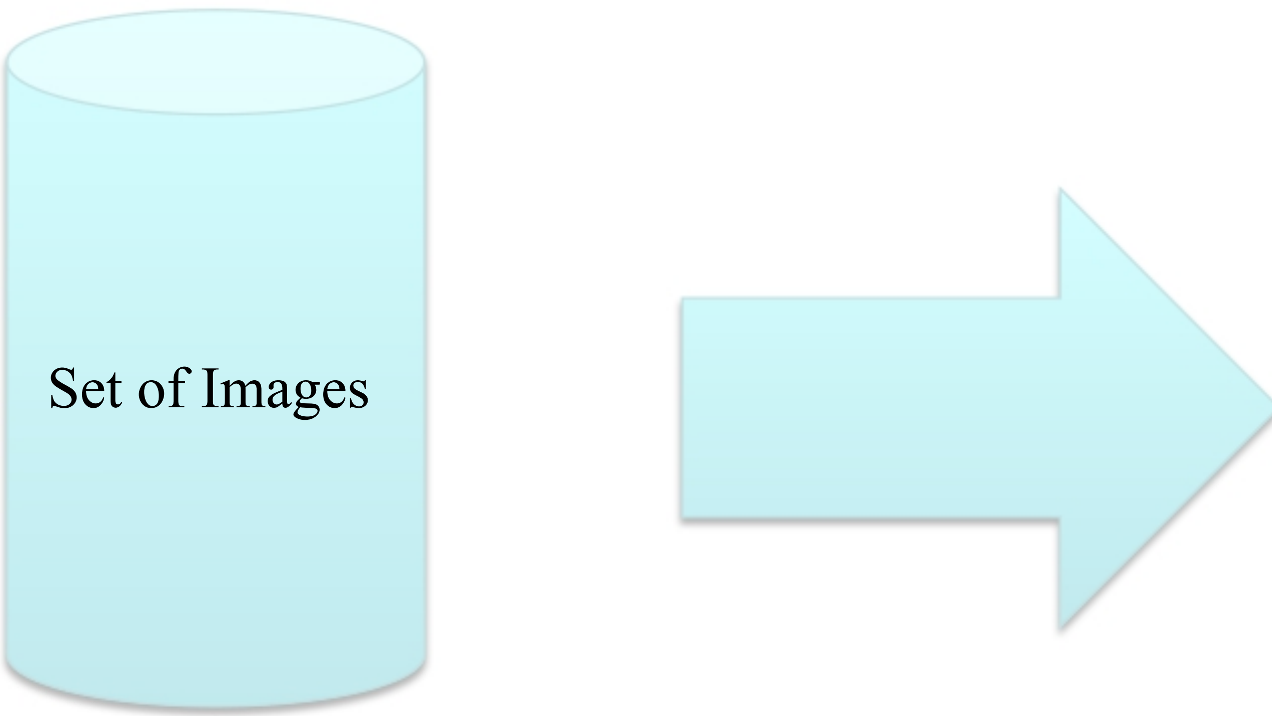
Clustering

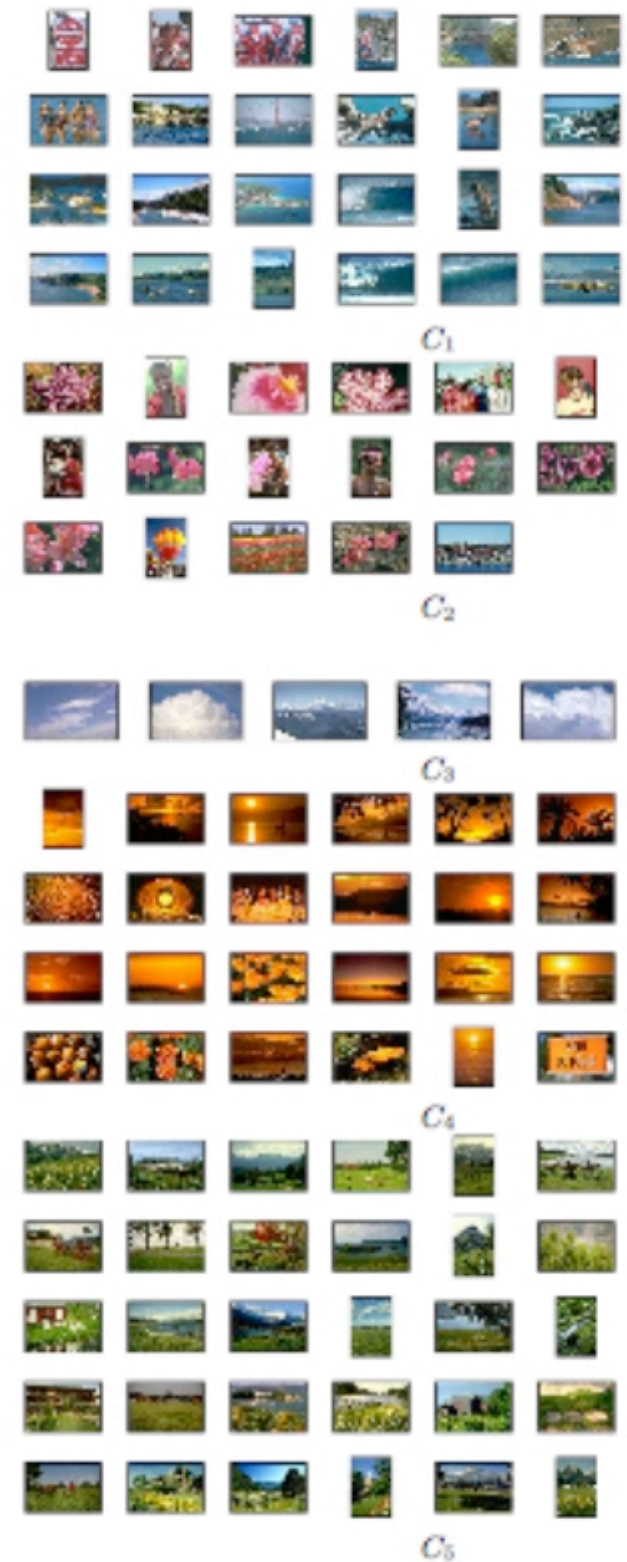
discovering structure in data

Clustering Data: Group similar things



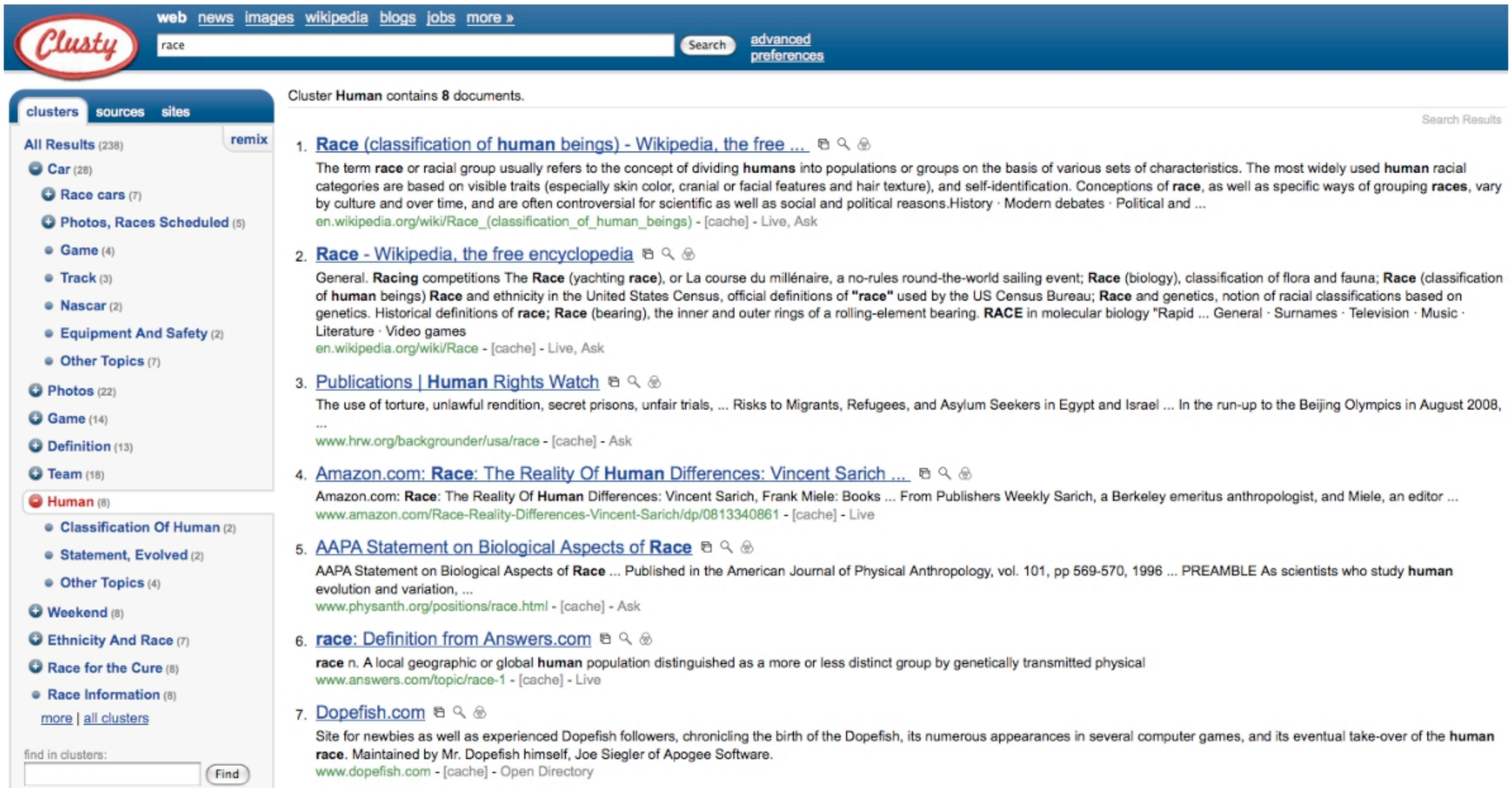
Clustering images





[Goldberger et al.]

Clustering web search results

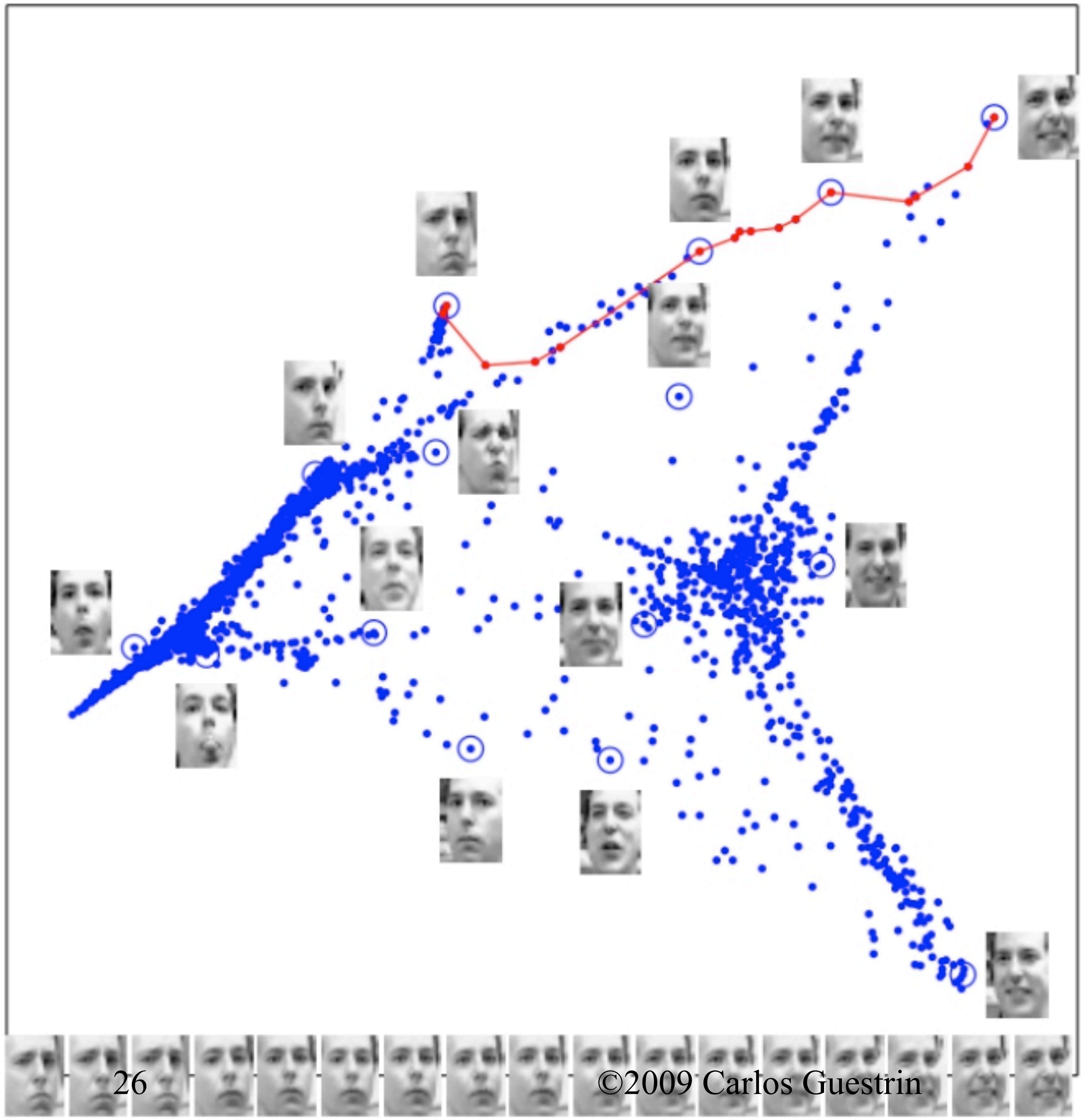


Embedding

visualizing data

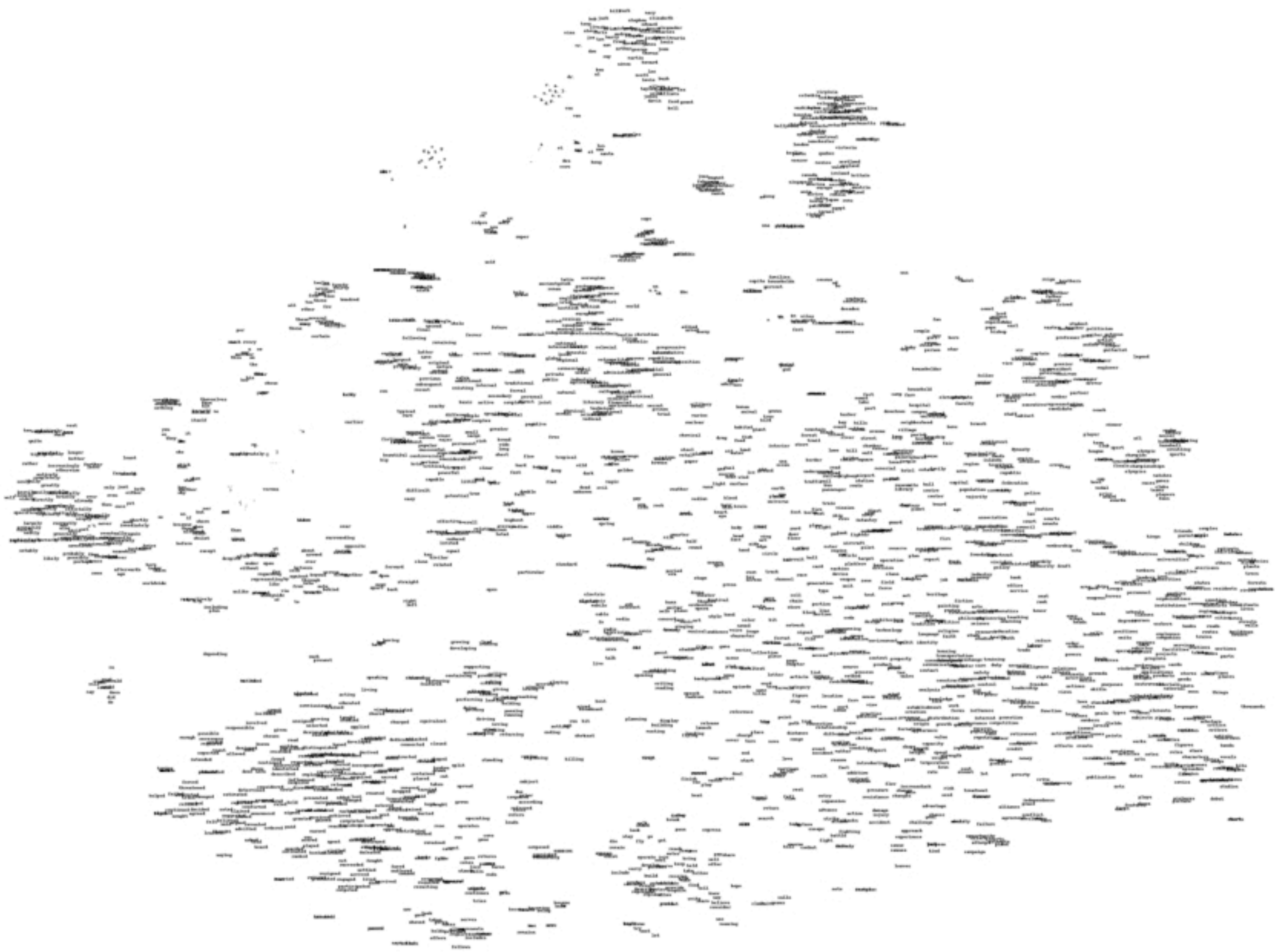
Embedding images

- Images have thousands or millions of pixels.
- Can we give each image a coordinate, such that similar images are near each other?



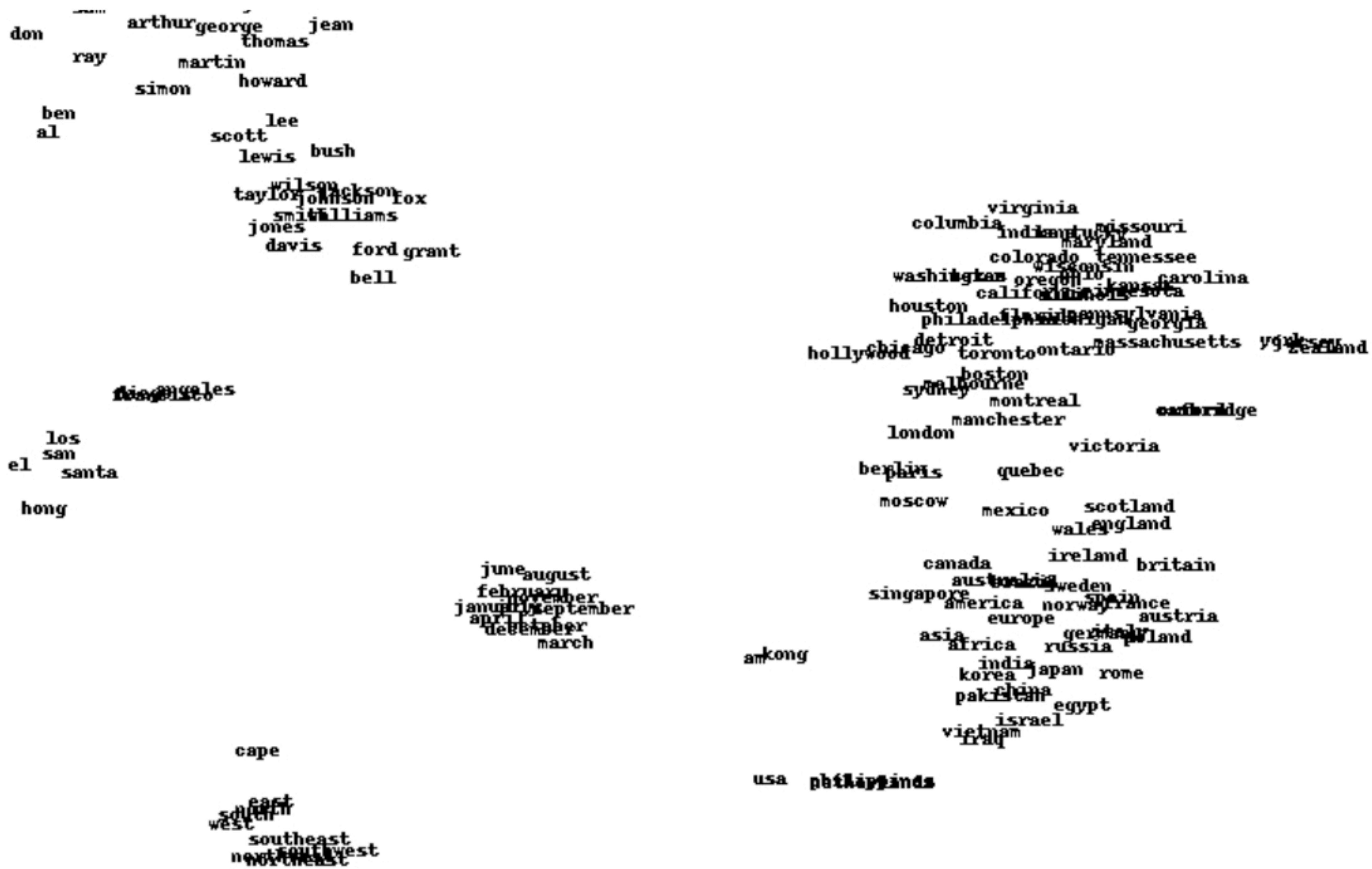
[Saul & Roweis '03]

Embedding words



[Joseph Turian]

Embedding words (zoom in)

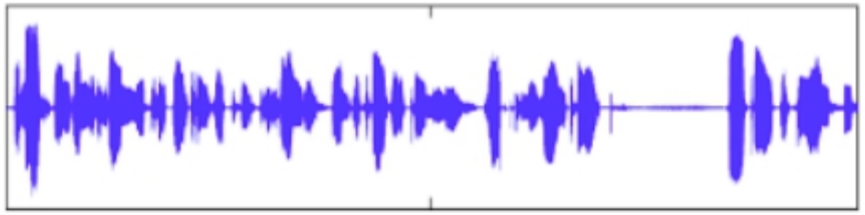


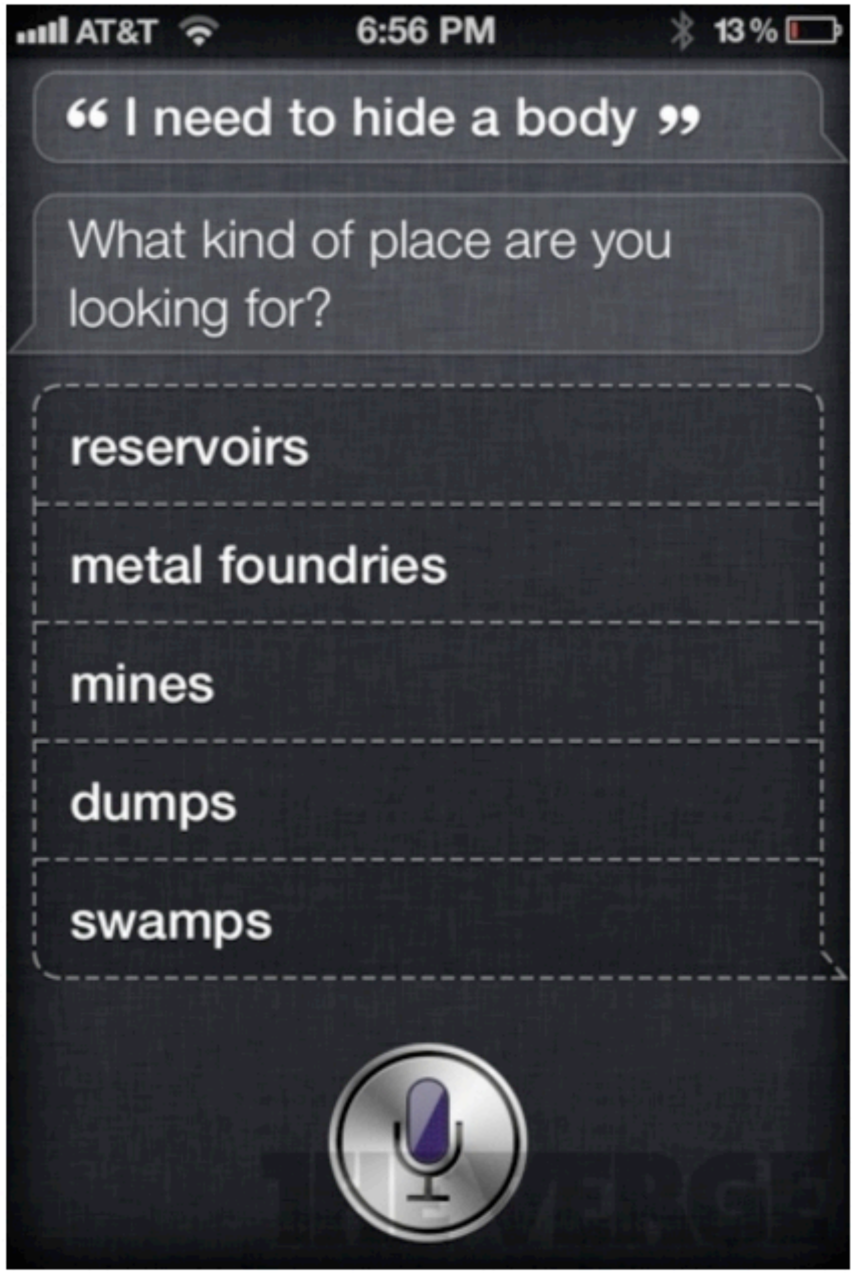
[Joseph Turian]

Structured prediction

from data to discrete classes

Speech recognition

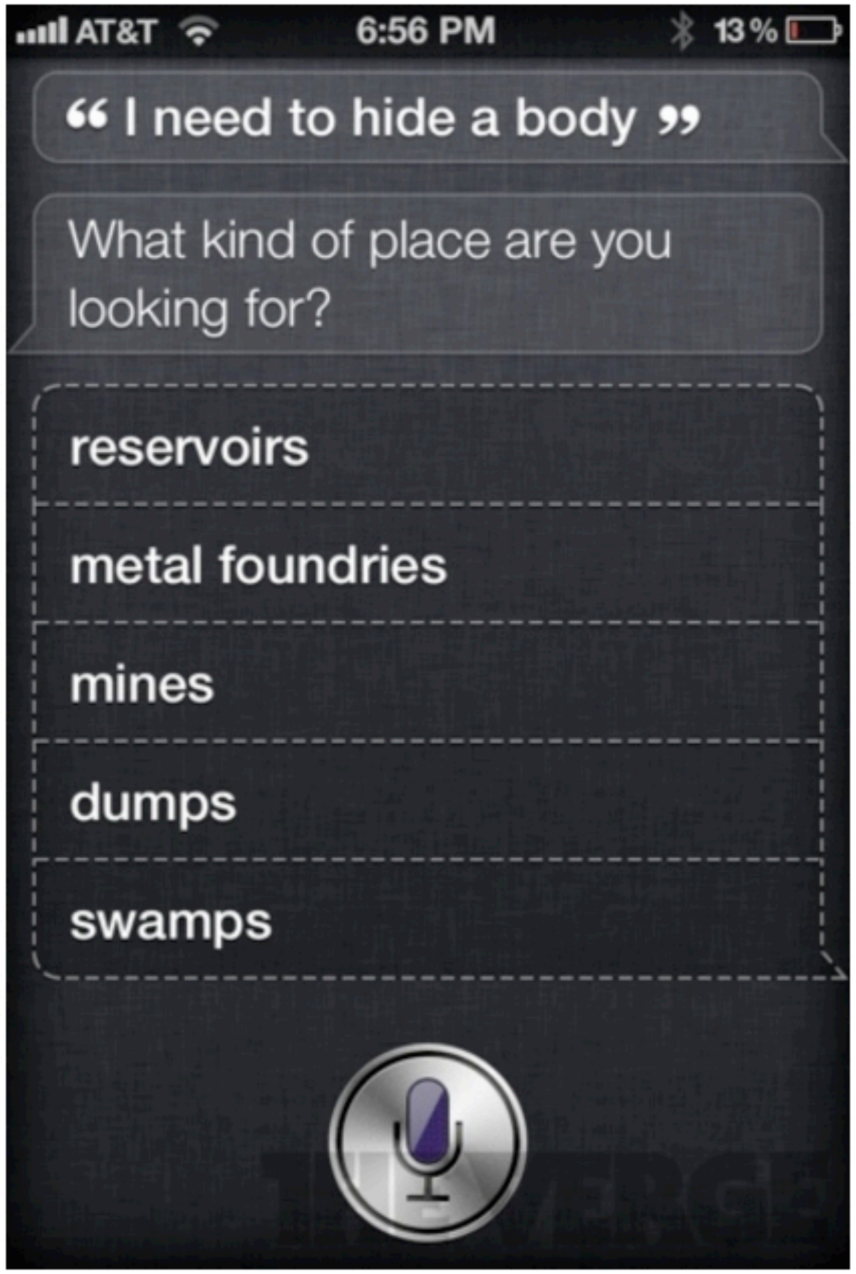




Natural language processing

I need to hide a body

noun, verb, preposition, …



Growth of Machine Learning

- Machine learning is preferred approach to
  - Speech recognition, Natural language processing
  - Computer vision
  - Medical outcomes analysis
  - Robot control
  - Computational biology
  - Sensor networks
  - …
- This trend is accelerating
  - Big data
  - Improved machine learning algorithms
  - Faster computers
  - Good open-source software

ML roadmap

- **First half: supervised learning**
  - SVMs, kernel methods
  - Learning theory
  - Decision trees, boosting, deep learning

- **Second half: data science**
  - Unsupervised learning, EM algorithm
  - Dimensionality reduction
  - Topic models

Supervised Learning: find f

- Given: Training set $\{(x_i, y_i) \mid i = 1 \ldots N\}$
- Find: A good approximation to $f : X \to Y$

Examples: what are X and Y?

- Spam Detection
  - Map email to {Spam, Not Spam}
- Digit recognition
  - Map pixels to {0,1,2,3,4,5,6,7,8,9}
- Stock Prediction
  - Map new, historic prices, etc. to $\Re$ (the real numbers)

A Supervised Learning Problem

Dataset:

Example	x_1	x_2	x_3	x_4	y
1	0	0	1	0	0
2	0	1	0	0	0
3	0	0	1	1	1
4	1	0	0	1	1
5	0	1	1	0	0
6	1	1	0	0	0
7	0	1	0	1	0

- Our goal is to find a function $f : X\text{->}Y$
  - $X = \{0,1\}$^4
  - $Y = \{0,1\}$
- Question 1: How should we pick the *hypothesis space*, the set of possible functions f ?
- Question 2: How do we find the best f in the hypothesis space?

Most General Hypothesis Space

Consider all possible boolean functions over four input features!

- 2^{16} possible hypotheses
- 2^{9} are consistent with our dataset
- How do we choose the best one?

x_1	x_2	x_3	x_4	y
0	0	0	0	?
0	0	0	1	?
0	0	1	0	0
0	0	1	1	1
0	1	0	0	0
0	1	0	1	0
0	1	1	0	0
0	1	1	1	?
1	0	0	0	?
1	0	0	1	1
1	0	1	0	?
1	0	1	1	?
1	1	0	0	0
1	1	0	1	?
1	1	1	0	?
1	1	1	1	?

Dataset:

Example	x_1	x_2	x_3	x_4	y
1	0	0	1	0	0
2	0	1	0	0	0
3	0	0	1	1	1
4	1	0	0	1	1
5	0	1	1	0	0
6	1	1	0	0	0
7	0	1	0	1	0

A Restricted Hypothesis Space

Consider all conjunctive boolean functions.

- 16 possible hypotheses
- None are consistent with our dataset
- How do we choose the best one?

Rule	Counterexample
$\Rightarrow y$	1
$x_1 \Rightarrow y$	3
$x_2 \Rightarrow y$	2
$x_3 \Rightarrow y$	1
$x_4 \Rightarrow y$	7
$x_1 \wedge x_2 \Rightarrow y$	3
$x_1 \wedge x_3 \Rightarrow y$	3
$x_1 \wedge x_4 \Rightarrow y$	3
$x_2 \wedge x_3 \Rightarrow y$	3
$x_2 \wedge x_4 \Rightarrow y$	3
$x_3 \wedge x_4 \Rightarrow y$	4
$x_1 \wedge x_2 \wedge x_3 \Rightarrow y$	3
$x_1 \wedge x_2 \wedge x_4 \Rightarrow y$	3
$x_1 \wedge x_3 \wedge x_4 \Rightarrow y$	3
$x_2 \wedge x_3 \wedge x_4 \Rightarrow y$	3
$x_1 \wedge x_2 \wedge x_3 \wedge x_4 \Rightarrow y$	3

Dataset:

Example	x_1	x_2	x_3	x_4	y
1	0	0	1	0	0
2	0	1	0	0	0
3	0	0	1	1	1
4	1	0	0	1	1
5	0	1	1	0	0
6	1	1	0	0	0
7	0	1	0	1	0

Occam's Razor Principle

- William of Occam: Monk living in the 14th century
- Principle of parsimony:

"One should not increase, beyond what is necessary, the number of entities required to explain anything"

- When many solutions are available for a given problem, we should select the simplest one
- But what do we mean by simple?
- We will use prior knowledge of the problem to solve to define what is a simple solution

Example of a prior: smoothness

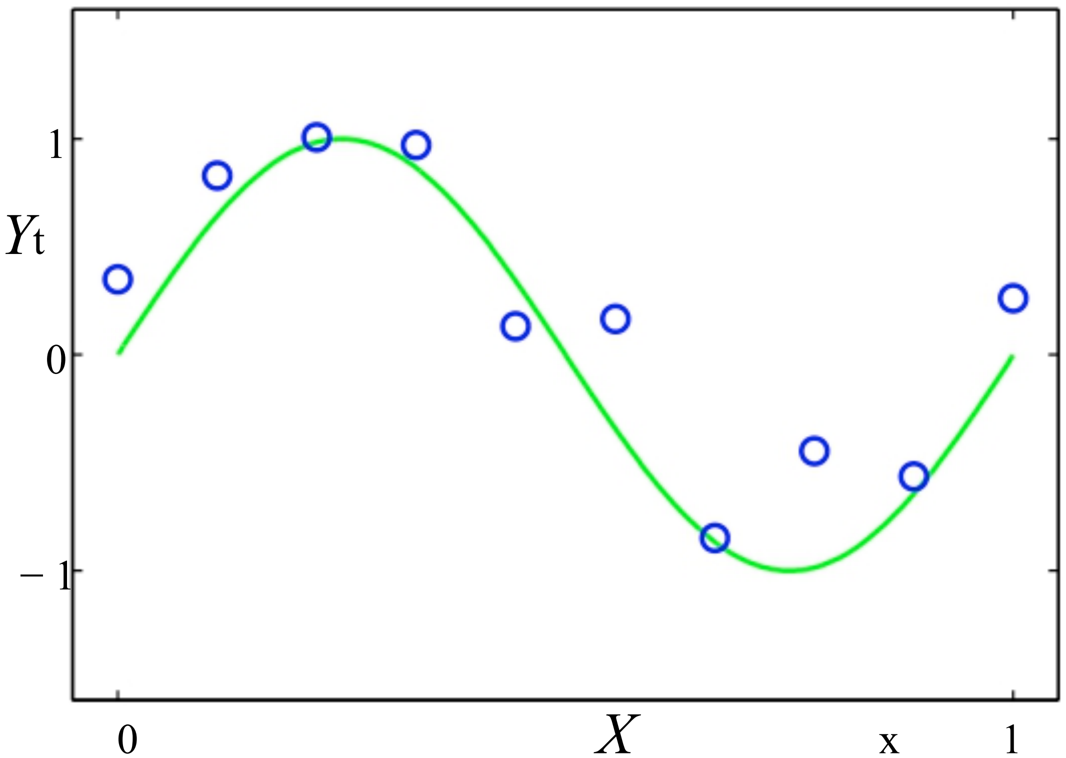
[Samy Bengio]

Key Issues in Machine Learning

- How do we choose a hypothesis space?
  - Often we use **prior knowledge** to guide this choice
- How can we gauge the accuracy of a hypothesis on unseen data?
  - **Occam's razor:** use the *simplest* hypothesis consistent with data! This will help us avoid overfitting.
  - ***Learning theory*** will help us quantify our ability to ***generalize*** as a function of the amount of training data and the hypothesis space
- How do we find the best hypothesis?
  - This is an **algorithmic** question, the main topic of computer science
- How to model applications as machine learning problems? (engineering challenge)

Second example: Regression

Dataset: 10 (X,Y) points generated from a sin function, with noise



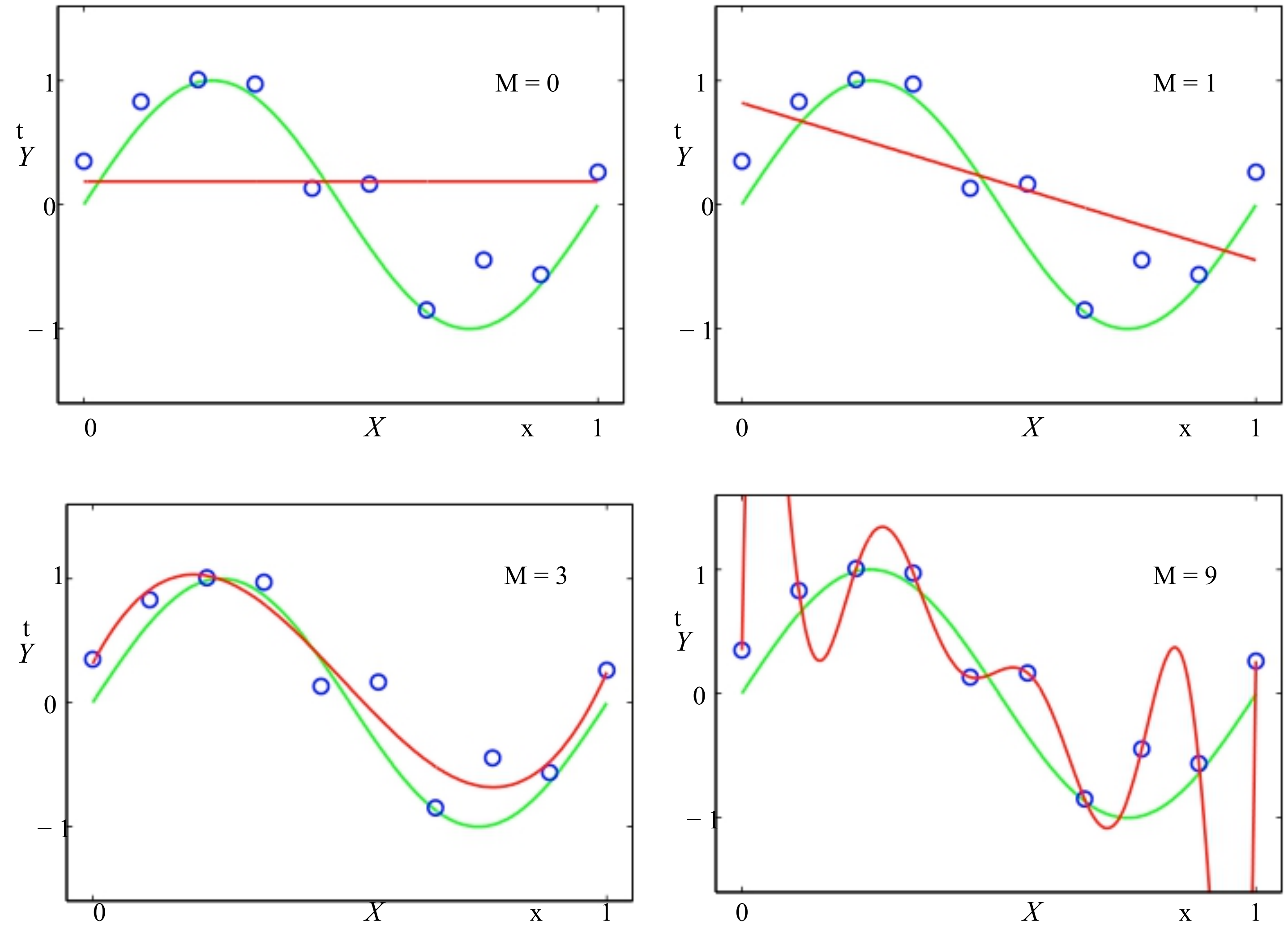
- Regression:
  - $f : X\text{->}Y$
  - $X = \Re$
  - $Y = \Re$

[Bishop]

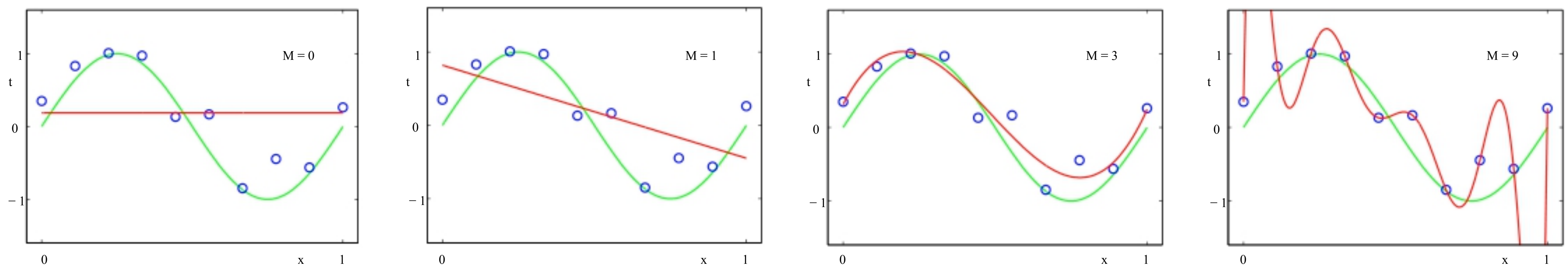
Degree-M Polynomials

How about letting *f* be a degree M polynomial?

- Which one is **best** ?



Hypo. Space: Degree-N Polynomials



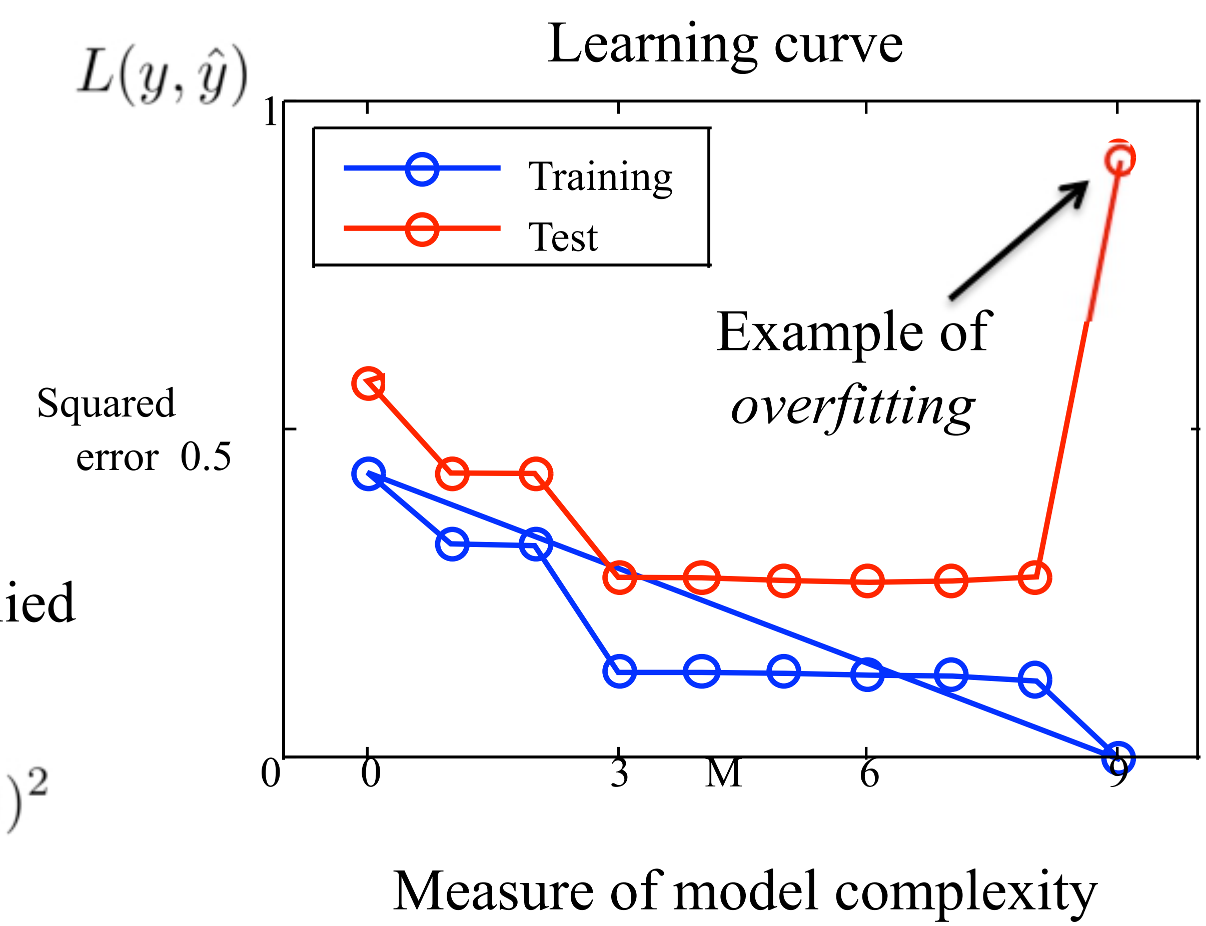
We measure error using a *loss function* $L(y, \hat{y})$

For regression, a common choice is squared loss:

$$L(y_i, f(x_i)) = (y_i - f(x_i))^2$$

The *empirical loss* of the function *f* applied to the training data is then:

$$\frac{1}{N}\sum_{i=1}^{N} L(y_i, f(x_i)) = \frac{1}{N}\sum_{i=1}^{N}(y_i - f(x_i))^2$$



Binary classification

- Input: email
- Output: spam/ham
- Setup:
  - Get a large collection of example emails, each labeled “spam” or “ ham”
  - Note: someone has to hand label all this data!
  - Want to learn to predict labels of new, future emails

- Features: The attributes used to make the ham / spam decision
  - Words: FREE!
  - Text Patterns: $dd, CAPS
  - Non-text: SenderInContacts
  - …

Dear Sir.

First, I must solicit your confidence in this transaction, this is by virture of its nature as being utterly confidencial and top secret. …

TO BE REMOVED FROM FUTURE MAILINGS, SIMPLY REPLY TO THIS MESSAGE AND PUT "REMOVE" IN THE SUBJECT.

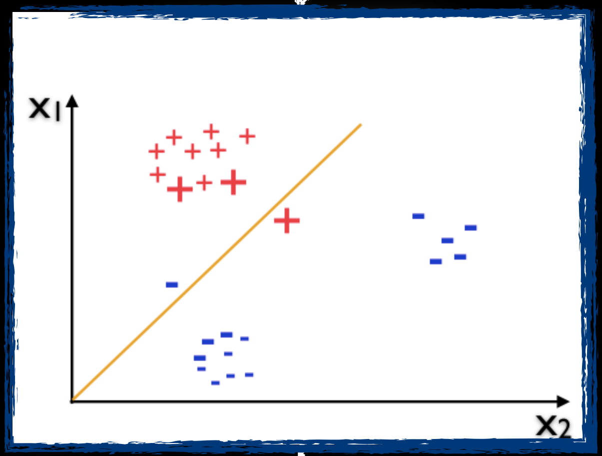
99 MILLION EMAIL ADDRESSES FOR ONLY $99

Ok, Iknow this is blatantly OT but I'm beginning to go insane. Had an old Dell Dimension XPS sitting in the corner and decided to put it to use, I know it was working pre being stuck in the corner, but when I plugged it in, hit the power nothing happened.

workflow of machine learning

instance → (features, charge or not) , (feature, charge or not) , ... ← class

pool of observations



instance → classifier → prediction or classification

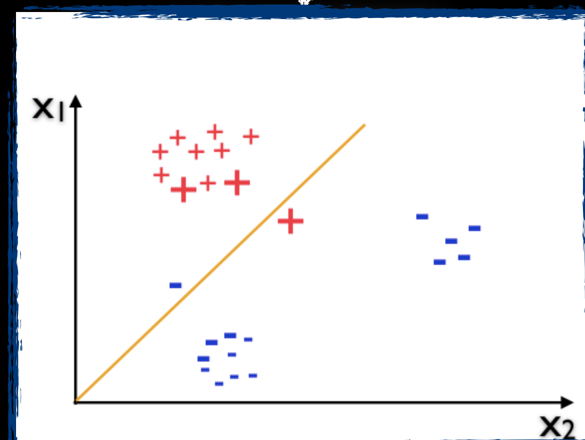
another algorithm constructed by the learner

workflow of machine learning

instance → (features, charge or not) , (feature, charge or not) , ... ← class

pool of observations

hyperplane



instance → classifier → prediction or classification

another algorithm constructed by the learner