Machine learning

Image source: https://www.coursera.org/course/ml

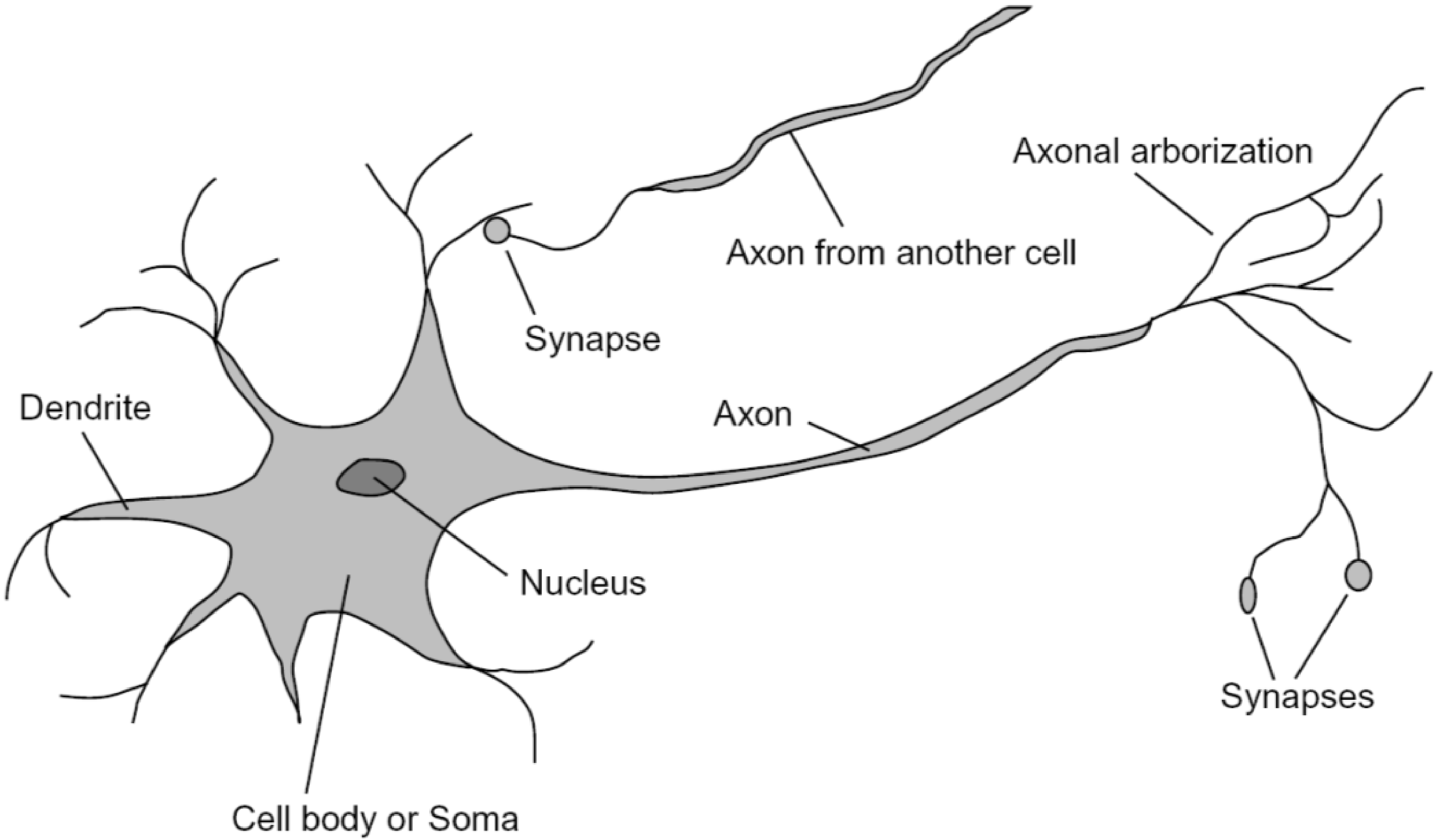
Machine learning

- Definition
  - Getting a computer to do well on a task without explicitly programming it
  - Improving performance on a task based on experience

Learning for episodic tasks

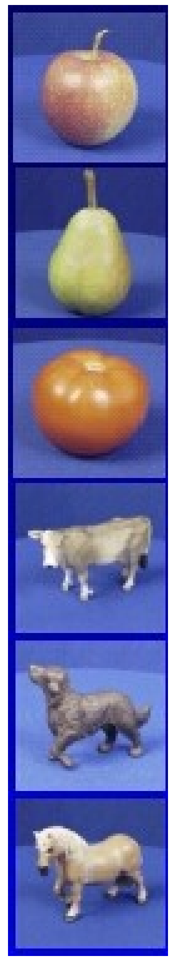
- We have just looked at learning in sequential environments
- Now let’s consider the “easier” problem of episodic environments
  - The agent gets a series of unrelated problem instances and has to make some decision or inference about each of them
  - In this case, “experience” comes in the form of *training data*

Loose inspiration: Human neurons



Example: Image classification

input **desired output**



apple

pear

tomato

cow

dog

horse

Training data

apple

pear

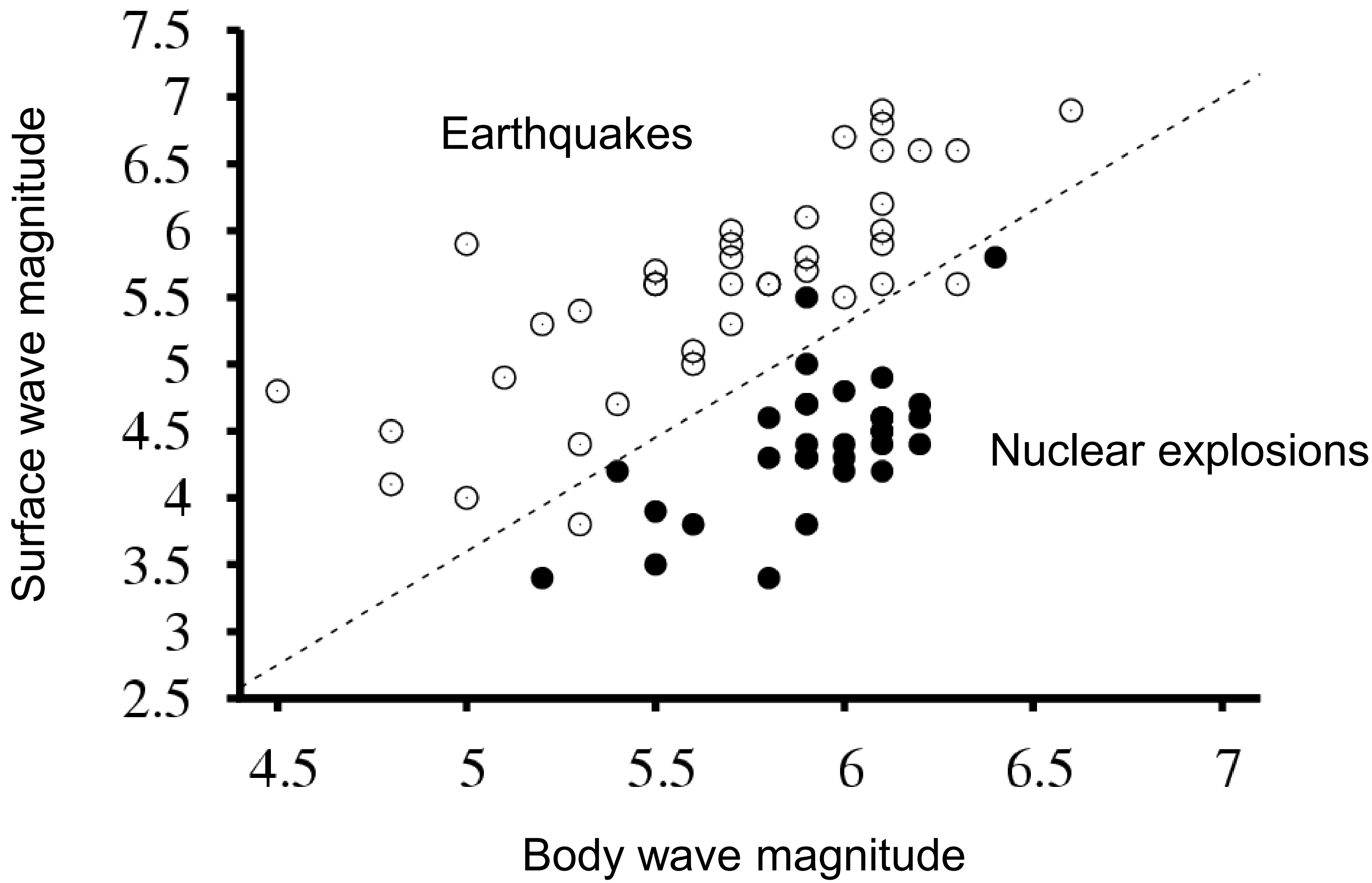
tomato

cow

dog

horse

Example 2: Seismic data classification



Example 3: Spam filter

Dear Sir.

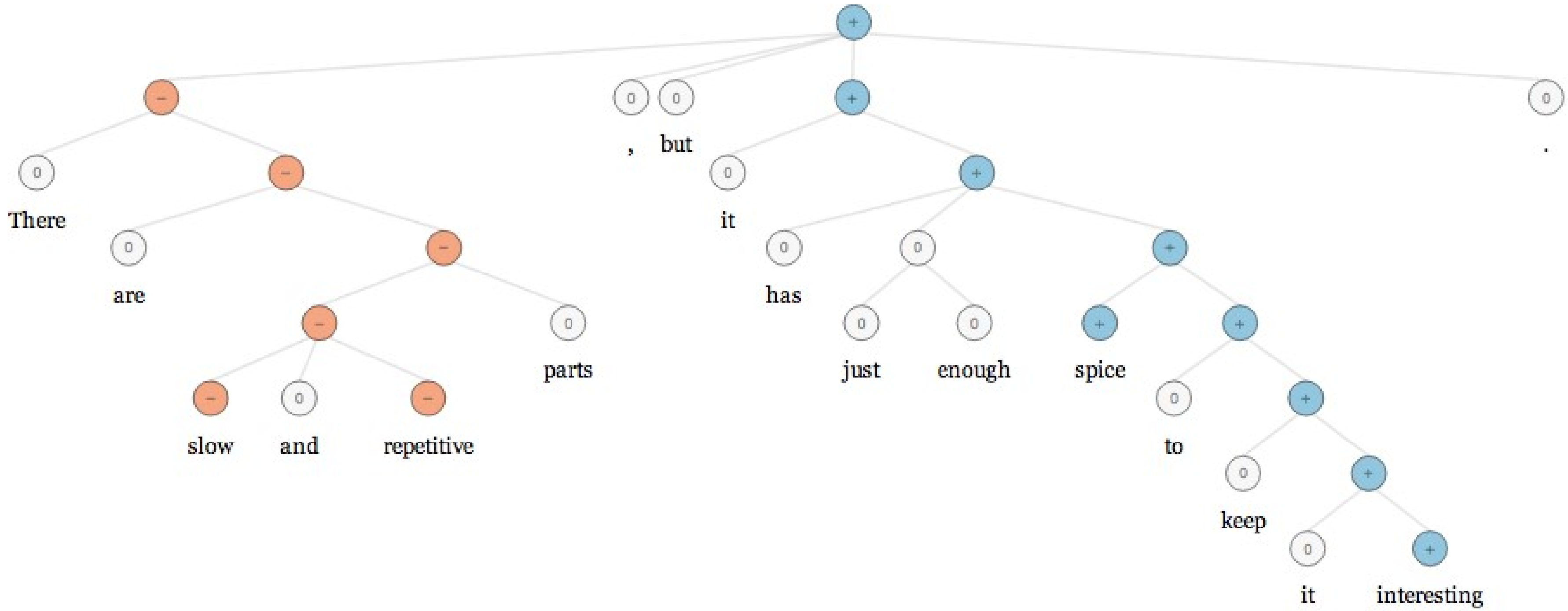
First, I must solicit your confidence in this transaction, this is by virture of its nature as being utterly confidencial and top secret. …

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.

Example 4: Sentiment analysis



http://gigaom.com/2013/10/03/stanford-researchers-to-open-source-model-they-say-has-nailed-sentiment-analysis/

http://nlp.stanford.edu:8080/sentiment/rntnDemo.html

Example 5: Robot grasping

L. Pinto and A. Gupta, Supersizing self-supervision: Learning to grasp from 50K tries and 700 robot hours," arXiv.org/abs/1509.06825

[YouTube video]()

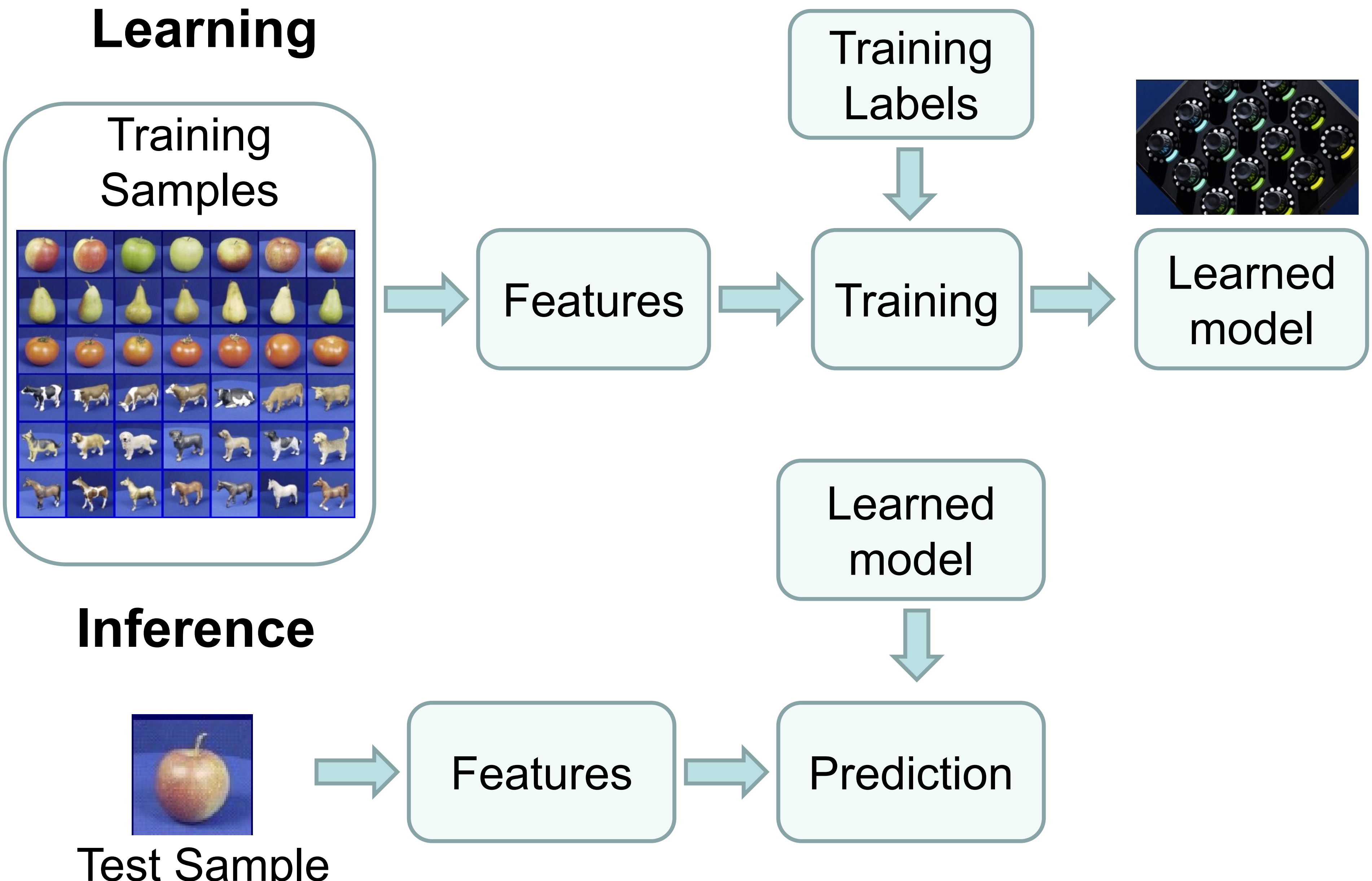
The basic *supervised learning* framework

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

output   classification function   input

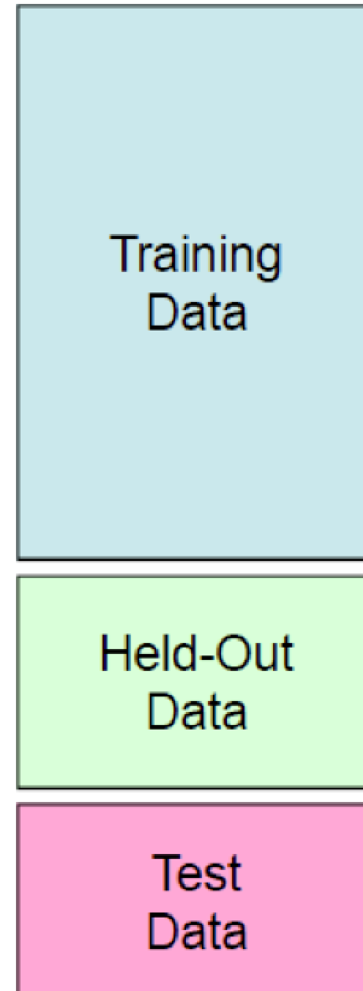
- **Learning:** given a *training set* of labeled examples $\{(\mathbf{x}_1,y_1), \ldots, (\mathbf{x}_N,y_N)\}$, estimate the parameters of the prediction function f
- **Inference:** apply f to a never before seen *test example* $\mathbf{x}$ and output the predicted value $y = f(\mathbf{x})$
- The key challenge is *generalization*

Learning and inference pipeline



Experimentation cycle

- Learn *parameters* on the *training set*
- Tune *hyperparameters* (implementation choices) on the *held out validation set*
- Evaluate performance on the *test set*
- **Very important**: do not peek at the test set!
- *Generalization* and *overfitting*
  - Want classifier that does well on never before seen data
  - Overfitting: good performance on the training/validation set, poor performance on test set



What’s the big deal?

Baidu admits cheating in international supercomputer competition

ZDNet

Baidu recently apologised for violating the rules of an international supercomputer test in May, when the Chinese search engine giant claimed to beat both Google and Microsoft on the ImageNet image-recognition test.

By Cyrus Lee | June 10, 2015 -- 00:15 GMT (17:15 PDT) | Topic: China

TECHNOLOGY

The New York Times

Computer Scientists Are Astir After Baidu Team Is Barred From A.I. Competition

By JOHN MARKOFF JUNE 3, 2015

Baidu caught gaming recent supercomputer performance test

by Andrew Tarantola | @terrortola | June 3rd 2015 At 11:09pm

engadget

IMAGENET Large Scale Visual Recognition Challenge (ILSVRC)

Date: June 2, 2015

Dear ILSVRC community,

This is a follow up to the announcement on May 19, 2015 with some more details and the status of the test server.

During the period of November 28th, 2014 to May 13th, 2015, there were at least 30 accounts used by a team from Baidu to submit to the test server at least 200 times, far exceeding the specified limit of two submissions per week. This includes short periods of very high usage, for example with more than 40 submissions over 5 days from March 15th, 2015 to March 19th, 2015. Figure A below shows submissions from ImageNet accounts known to be associated with the team in question. Figure B shows a comparison to the activity from all other accounts.

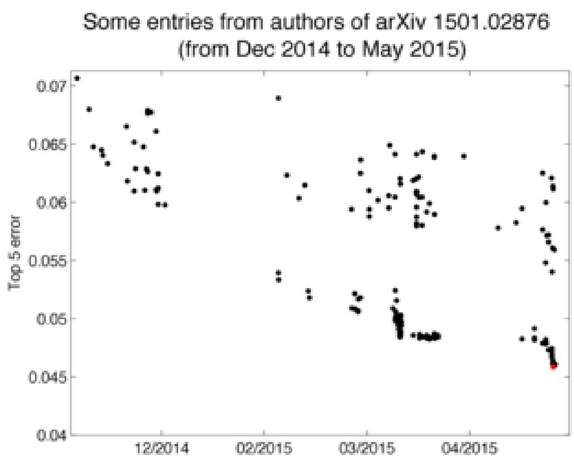


Figure A

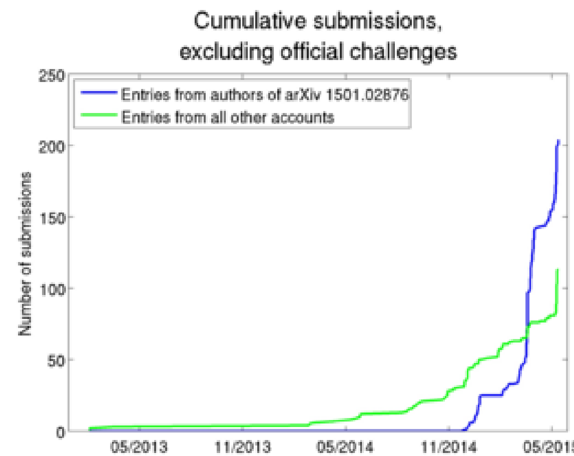


Figure B

The results obtained during this period are reported in a recent arXiv paper. Because of the violation of the regulations of the test server, these results may not be directly comparable to results obtained and reported by other teams. To make this clear, by exploiting the ability to test many slightly different solutions on the test server it is possible to 1) select the best out of a set of very similar solutions based on test performance and achieve a small but potentially significant advantage and 2) choose methods for further research and development based directly on the test data instead of using only the training and validation data for such choices.

http://www.image-net.org/challenges/LSVRC/announcement-June-2-2015

Naïve Bayes classifier

$$
\begin{aligned}
f(\boldsymbol{x}) &= \arg\max_y P(y \mid \boldsymbol{x}) \\
&\propto \arg\max_y P(y) P(\boldsymbol{x} \mid y) \\
&= \arg\max_y P(y) \prod_d P(x_d \mid y)
\end{aligned}
$$

A single
dimension or
attribute of **x**

Decision tree classifier

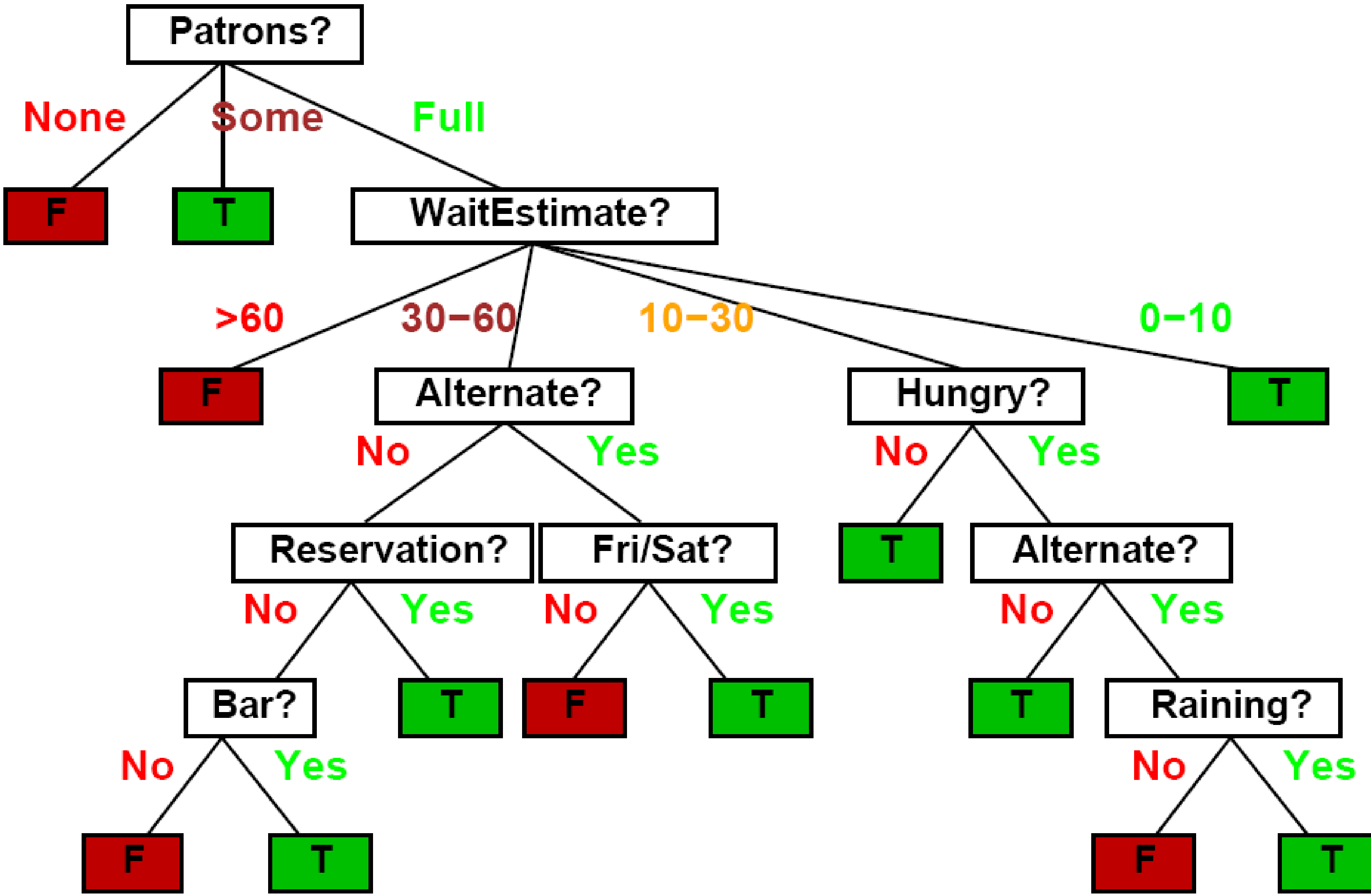
Example problem: decide whether to wait for a table at a restaurant, based on the following attributes:

1. **Alternate:** is there an alternative restaurant nearby?
2. **Bar:** is there a comfortable bar area to wait in?
3. **Fri/Sat:** is today Friday or Saturday?
4. **Hungry:** are we hungry?
5. **Patrons:** number of people in the restaurant (None, Some, Full)
6. **Price:** price range ($, $$, $$$)
7. **Raining:** is it raining outside?
8. **Reservation:** have we made a reservation?
9. **Type:** kind of restaurant (French, Italian, Thai, Burger)
10. **WaitEstimate:** estimated waiting time (0-10, 10-30, 30-60, >60)

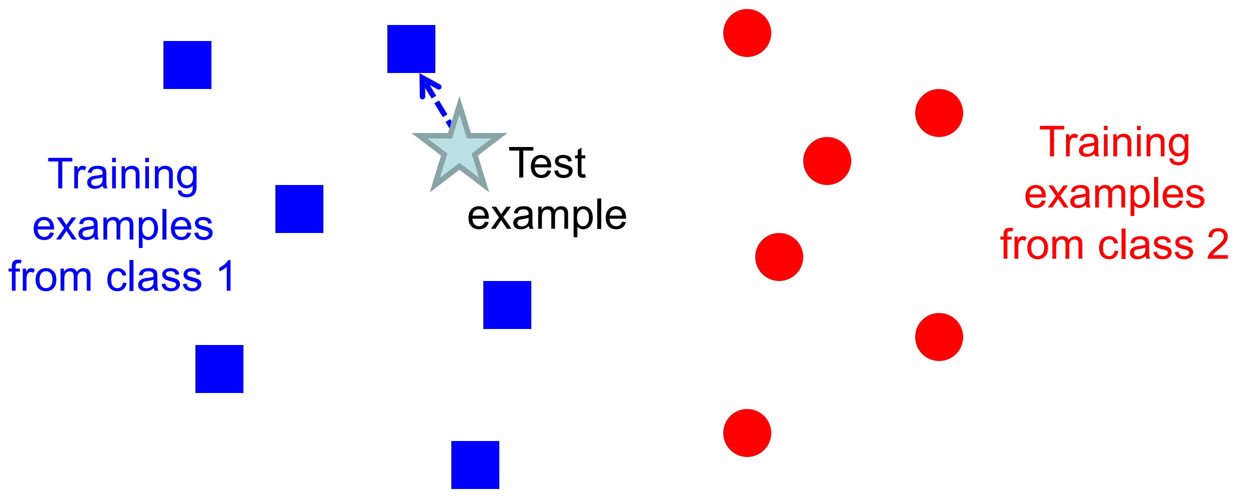
Decision tree classifier

Example	Attributes										Target
	Alt	*Bar*	*Fri*	*Hun*	*Pat*	*Price*	*Rain*	*Res*	*Type*	*Est*	*Wait*
X_1	T	F	F	T	Some	$$$	F	T	French	0–10	T
X_2	T	F	F	T	Full	$	F	F	Thai	30–60	F
X_3	F	T	F	F	Some	$	F	F	Burger	0–10	T
X_4	T	F	T	T	Full	$	F	F	Thai	10–30	T
X_5	T	F	T	F	Full	$$$	F	T	French	>60	F
X_6	F	T	F	T	Some	$$	T	T	Italian	0–10	T
X_7	F	T	F	F	None	$	T	F	Burger	0–10	F
X_8	F	F	F	T	Some	$$	T	T	Thai	0–10	T
X_9	F	T	T	F	Full	$	T	F	Burger	>60	F
X_{10}	T	T	T	T	Full	$$$	F	T	Italian	10–30	F
X_{11}	F	F	F	F	None	$	F	F	Thai	0–10	F
X_{12}	T	T	T	T	Full	$	F	F	Burger	30–60	T

Decision tree classifier



Nearest neighbor classifier

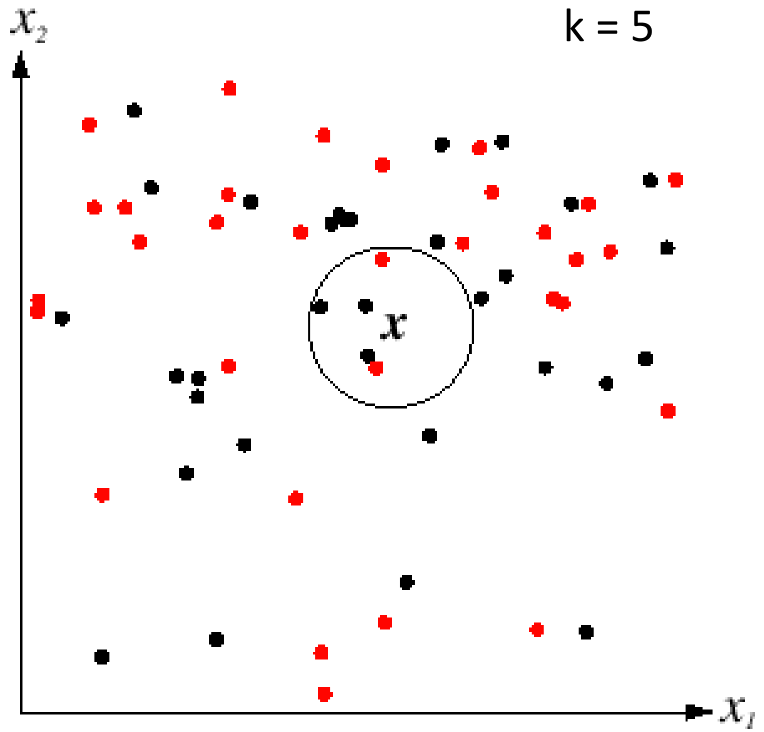


f(**x**) = label of the training example nearest to **x**

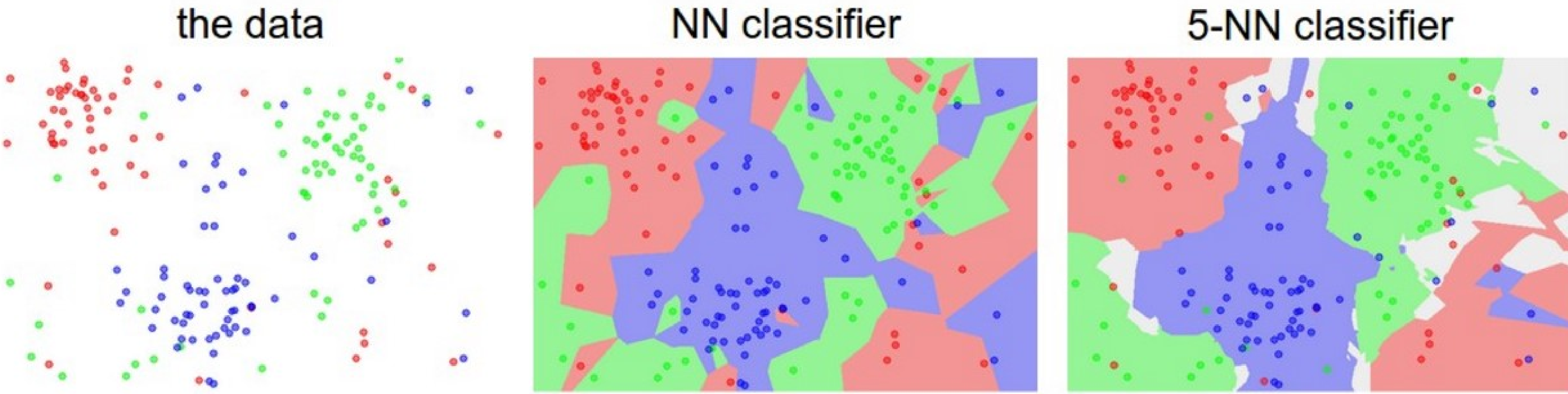
- All we need is a distance function for our inputs
- No training required!

K-nearest neighbor classifier

- For a new point, find the k closest points from training data
- Vote for class label with labels of the k points



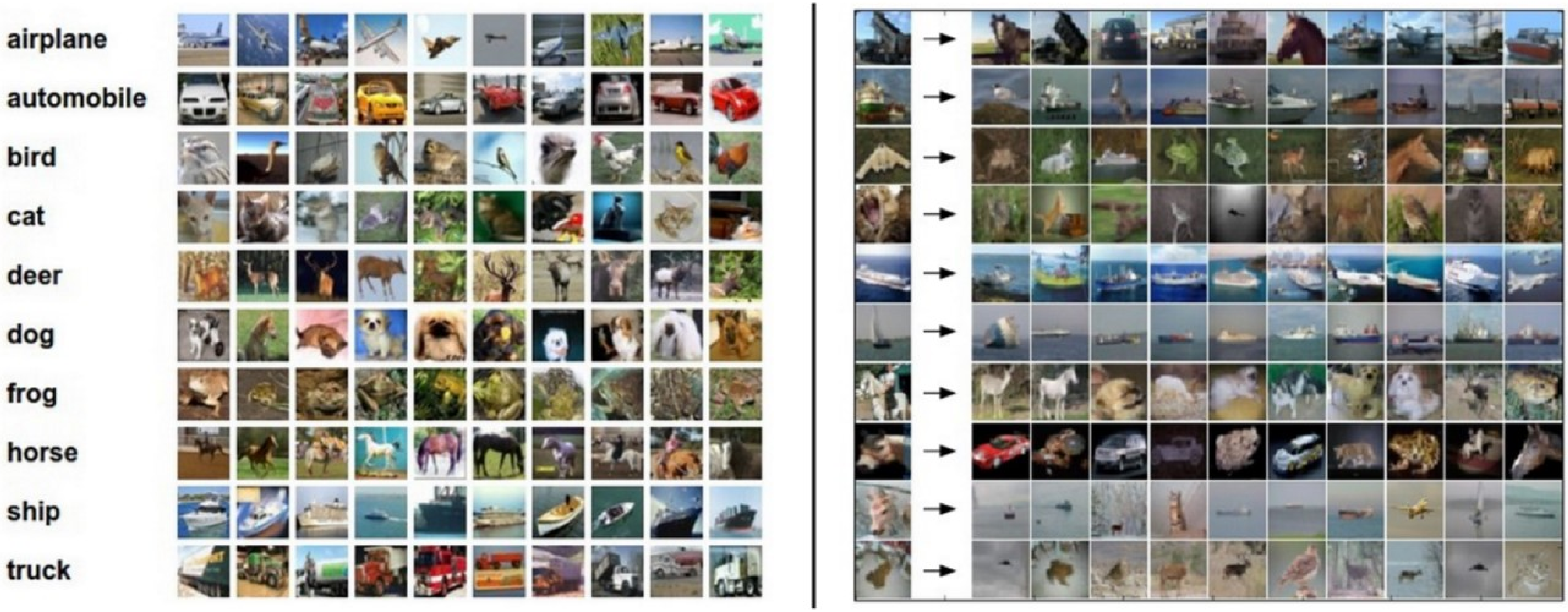
K-nearest neighbor classifier



- Which classifier is more robust to *outliers*?

Credit: Andrej Karpathy, http://cs231n.github.io/classification/

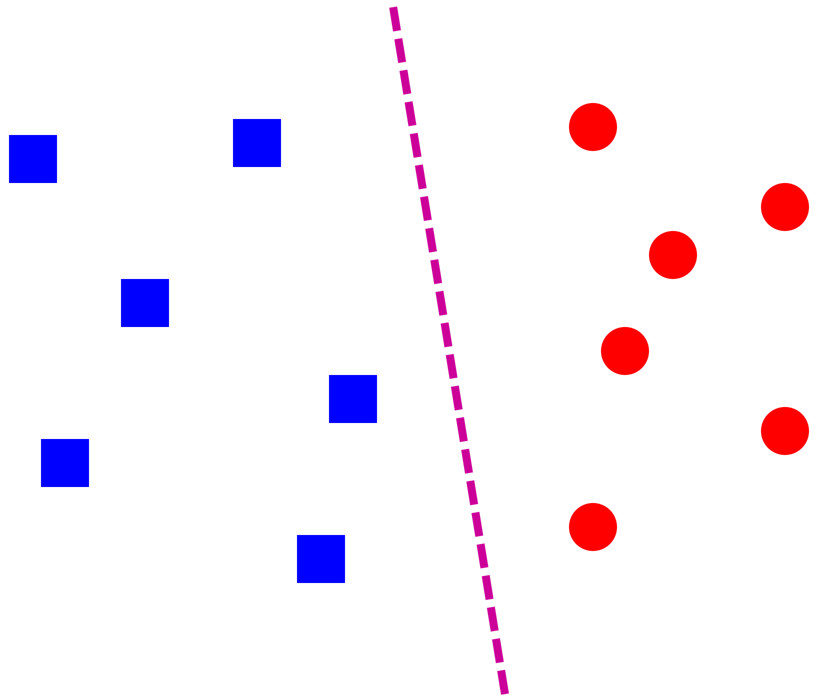
K-nearest neighbor classifier



Left: Example images from the CIFAR-10 dataset. Right: first column shows a few test images and next to each we show the top 10 nearest neighbors in the training set according to pixel-wise difference.

Credit: Andrej Karpathy, http://cs231n.github.io/classification/

Linear classifier



- Find a *linear function* to separate the classes

$$f(\mathbf{x}) = \text{sgn}(w_1x_1 + w_2x_2 + \ldots + w_Dx_D + b) = \text{sgn}(\mathbf{w} \cdot \mathbf{x} + b)$$

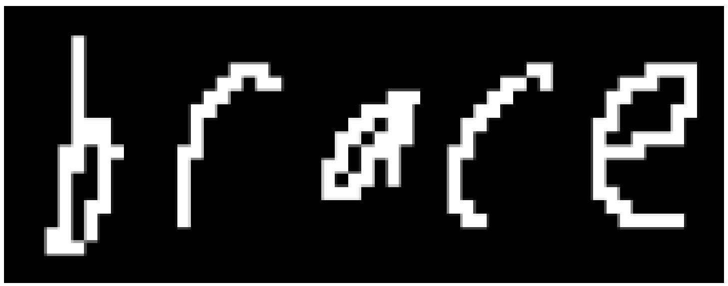
NN vs. linear classifiers

- NN pros:
  - \+ Simple to implement
  - \+ Decision boundaries not necessarily linear
  - \+ Works for any number of classes
  - \+ *Nonparametric* method
- NN cons:
  - – Need good distance function
  - – Slow at test time
- Linear pros:
  - \+ Low-dimensional *parametric* representation
  - \+ Very fast at test time
- Linear cons:
  - – Works for two classes
  - – How to train the linear function?
  - – What if data is not linearly separable?

Other machine learning scenarios

- Other prediction scenarios
  - Regression
  - Structured prediction
- Other supervision scenarios
  - Unsupervised learning
  - Semi-supervised learning
  - Active learning
  - Lifelong learning

Beyond simple classification: Structured prediction



brace

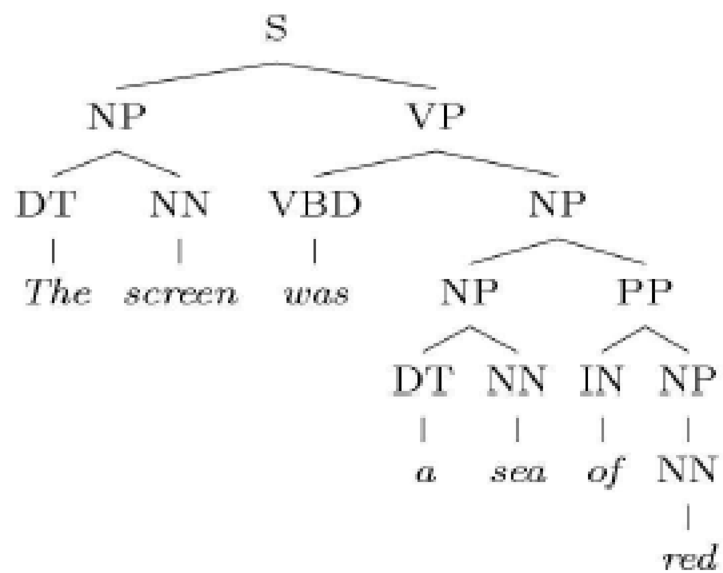
Image

Word

Source: B. Taskar

Structured Prediction

The screen was
a sea of red



Sentence

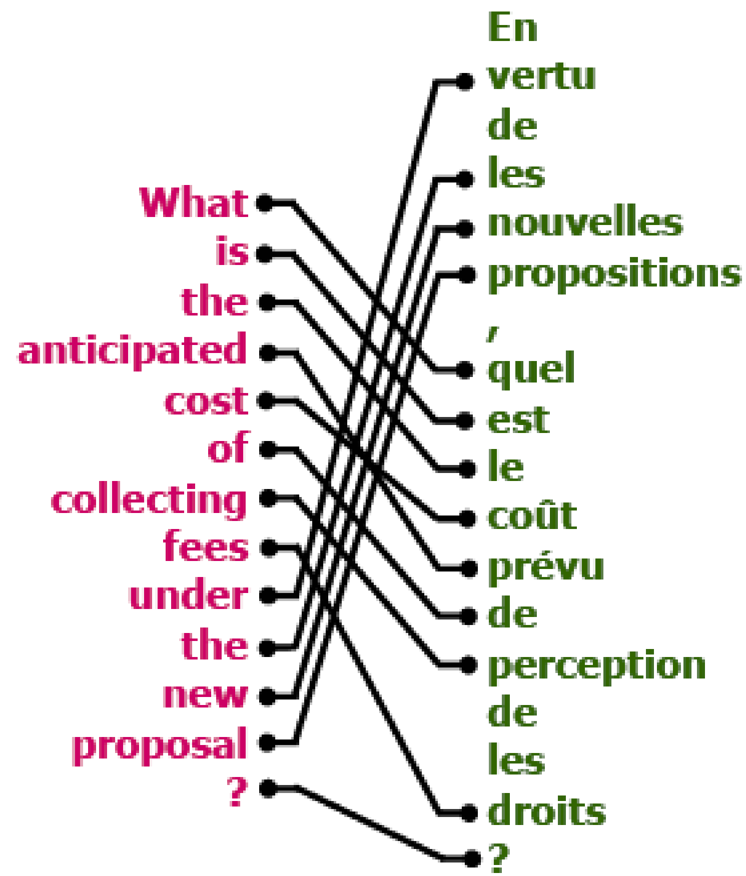
Parse tree

Source: B. Taskar

Structured Prediction

What is the anticipated cost of collecting fees under the new proposal?

En vertu des nouvelles propositions, quel est le coût prévu de perception des droits?



Sentence in two languages

Word alignment

Source: B. Taskar

Structured Prediction

RSCCPCYWGGCPW
GQNCYPEGCSGPKV

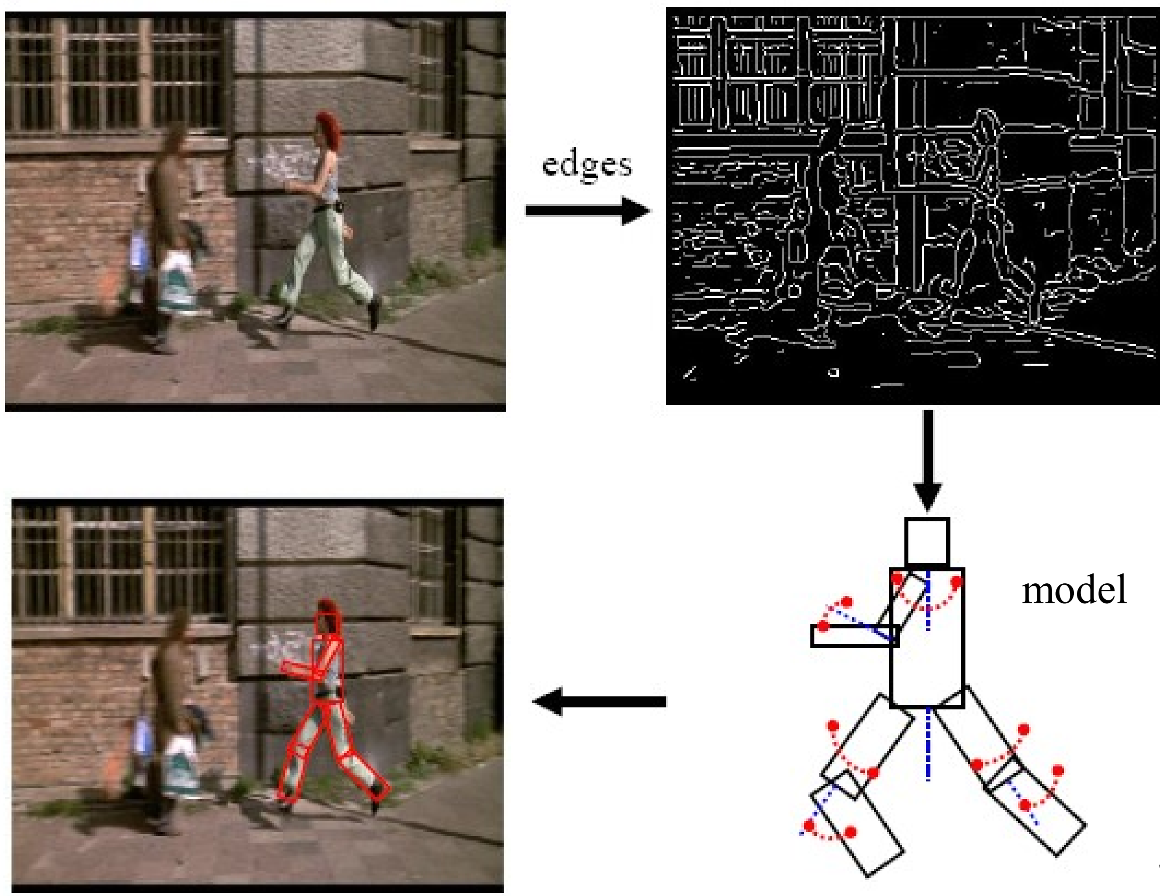
Amino-acid sequence

Bond structure

Source: B. Taskar

Structured Prediction

- Many image-based inference tasks can loosely be thought of as "structured prediction"



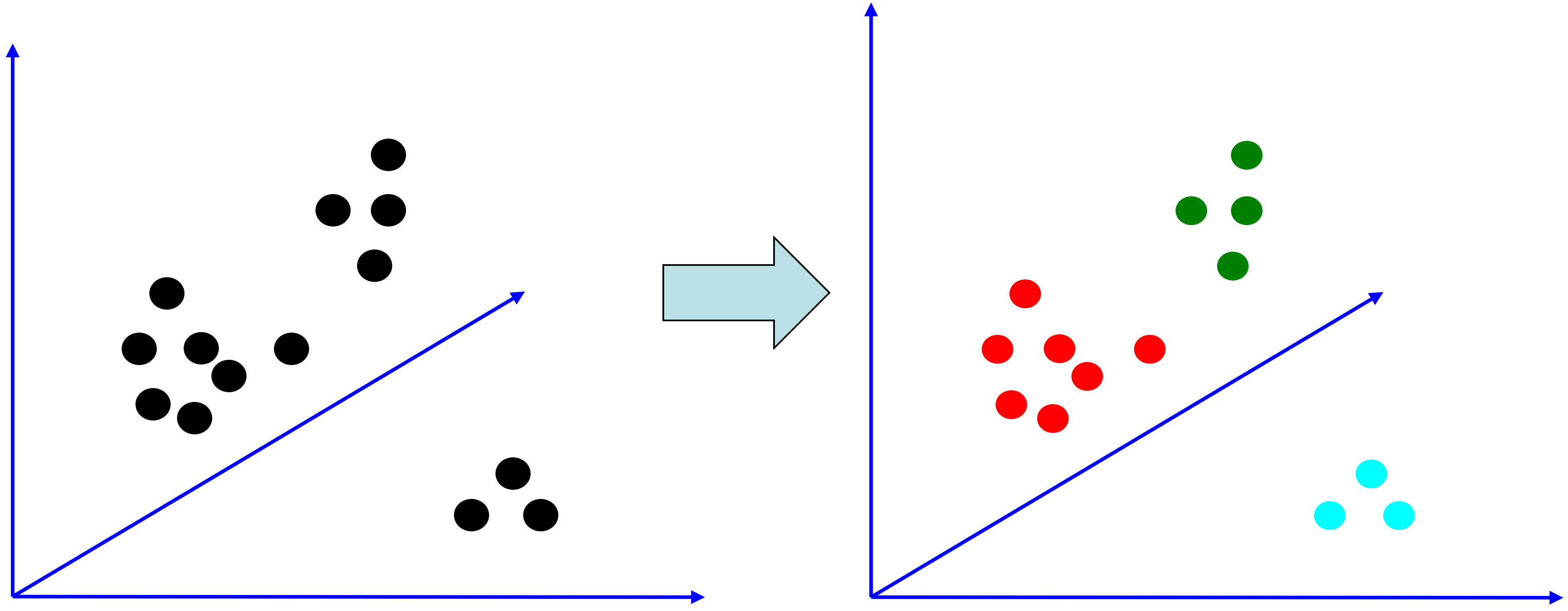
Source: D. Ramanan

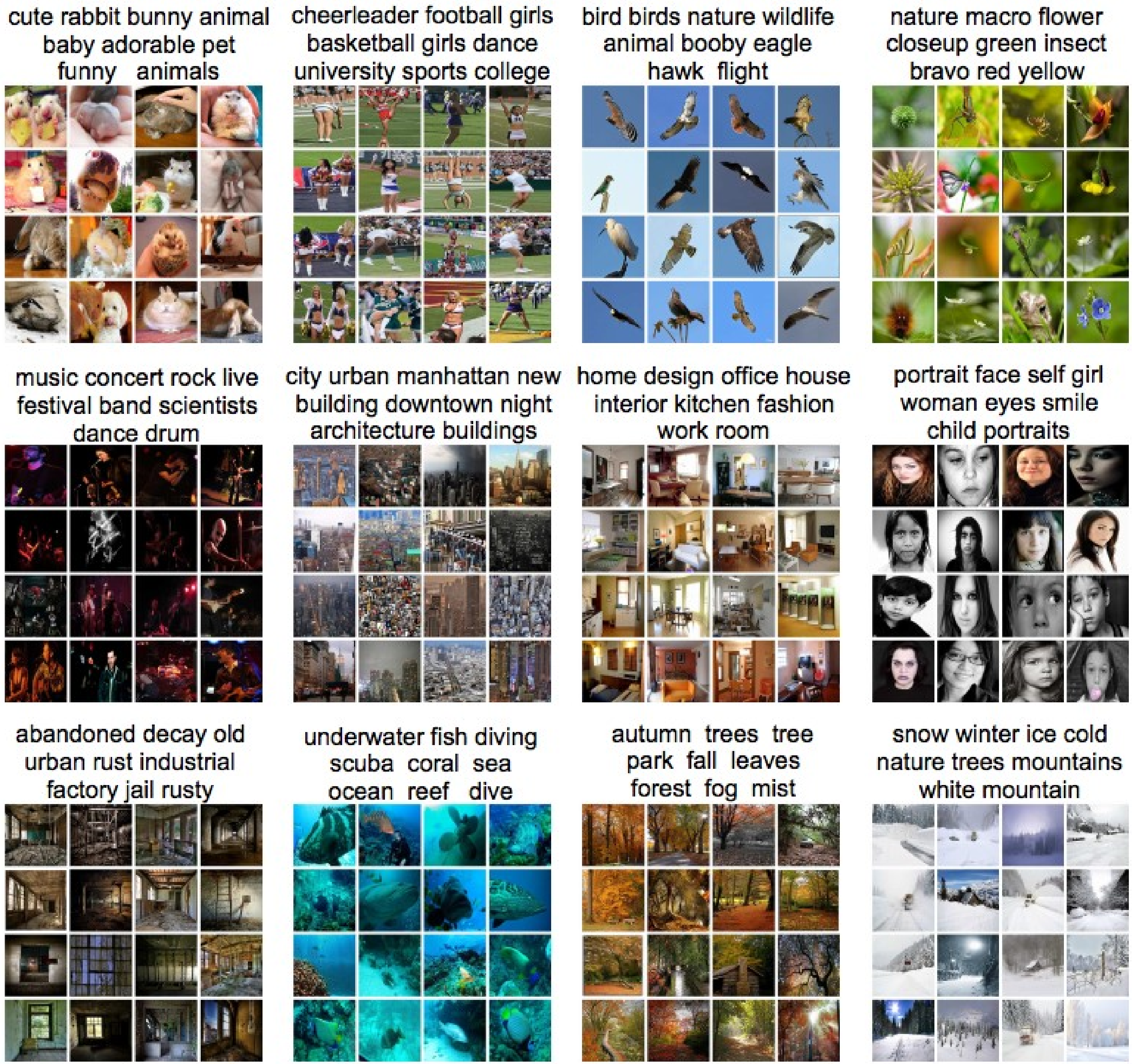
Unsupervised Learning

- Idea: Given only *unlabeled* data as input, learn some sort of structure
- The objective is often more vague or subjective than in supervised learning
- This is more of an exploratory/descriptive data analysis

Unsupervised Learning

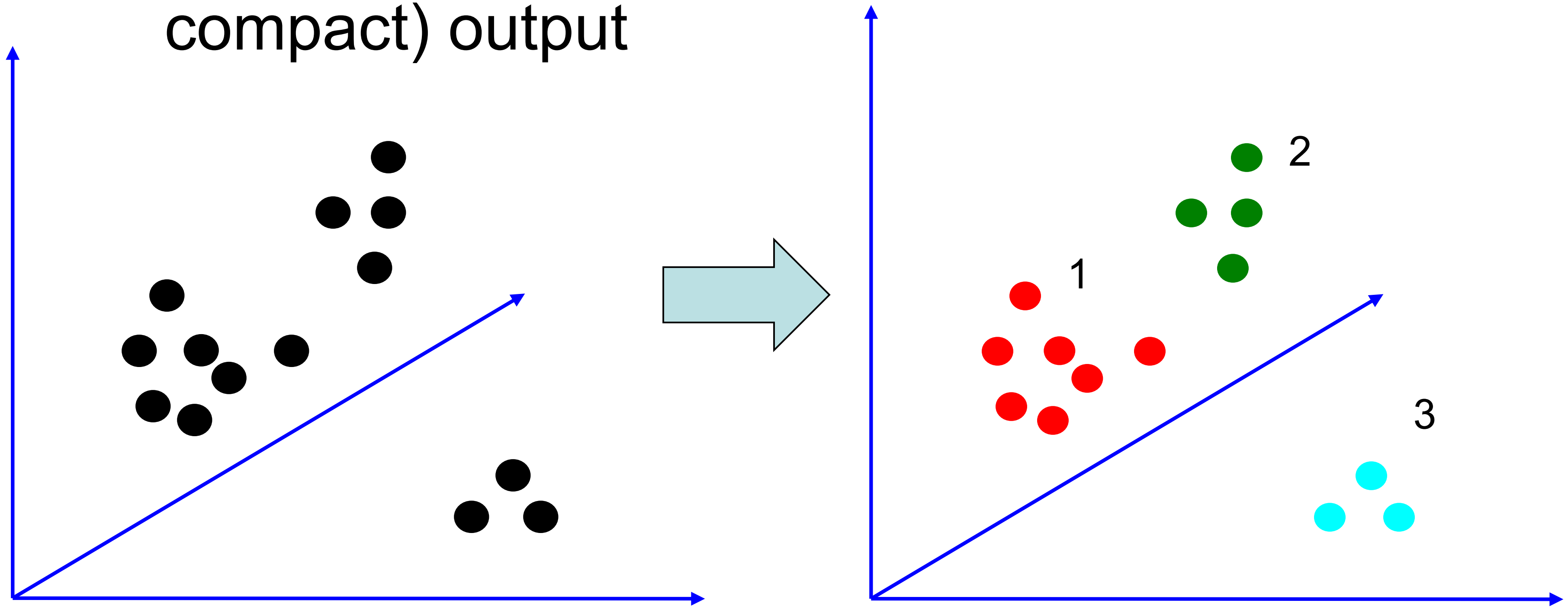
- **Clustering**
  – Discover groups of “similar” data points




cute rabbit bunny animal baby adorable pet funny animals
cheerleader football girls basketball girls dance university sports college
bird birds nature wildlife animal booby eagle hawk flight
nature macro flower closeup green insect bravo red yellow
music concert rock live festival band scientists dance drum
city urban manhattan new building downtown night architecture buildings
home design office house interior kitchen fashion work room
portrait face self girl woman eyes smile child portraits
abandoned decay old urban rust industrial factory jail rusty
underwater fish diving scuba coral sea ocean reef dive
autumn trees tree park fall leaves forest fog mist
snow winter ice cold nature trees mountains white mountain

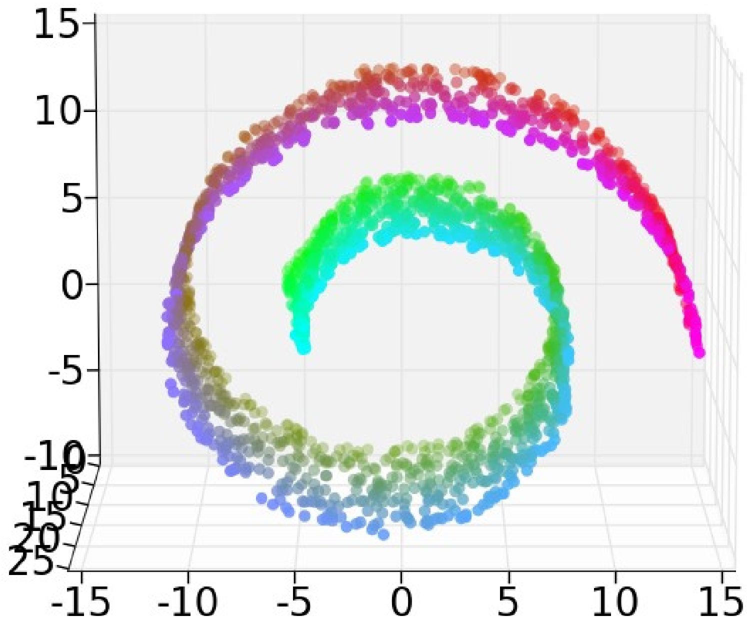
Unsupervised Learning

- **Quantization**
  - Map a continuous input to a discrete (more compact) output



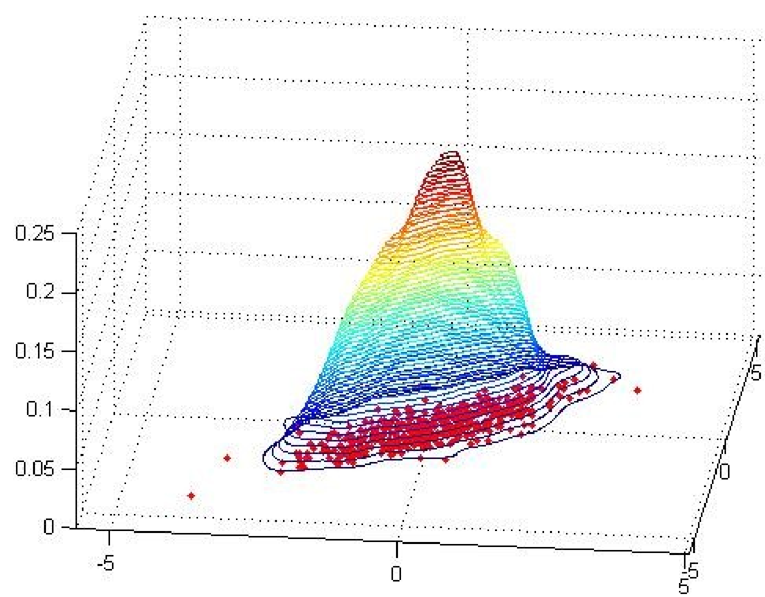
Unsupervised Learning

- **Dimensionality reduction, manifold learning**
  - Discover a lower-dimensional surface on which the data lives



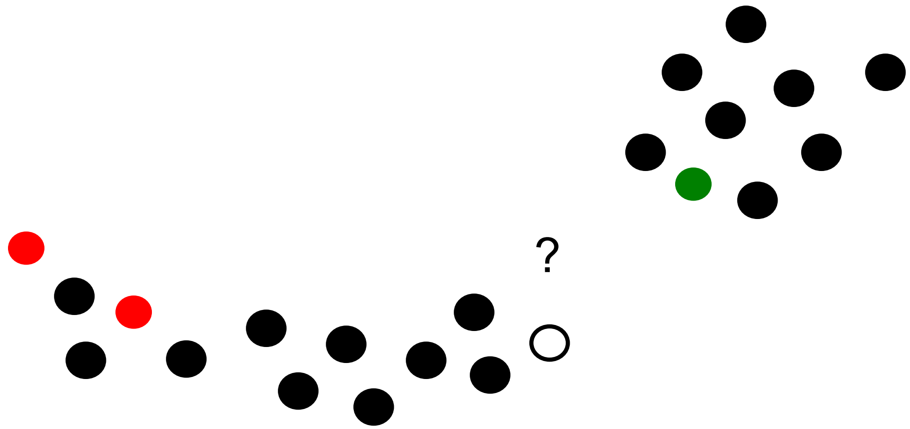
Unsupervised Learning

- **Density estimation**
  - Find a function that approximates the probability density of the data (i.e., value of the function is high for “typical” points and low for “atypical” points)
  - Can be used for **anomaly detection**



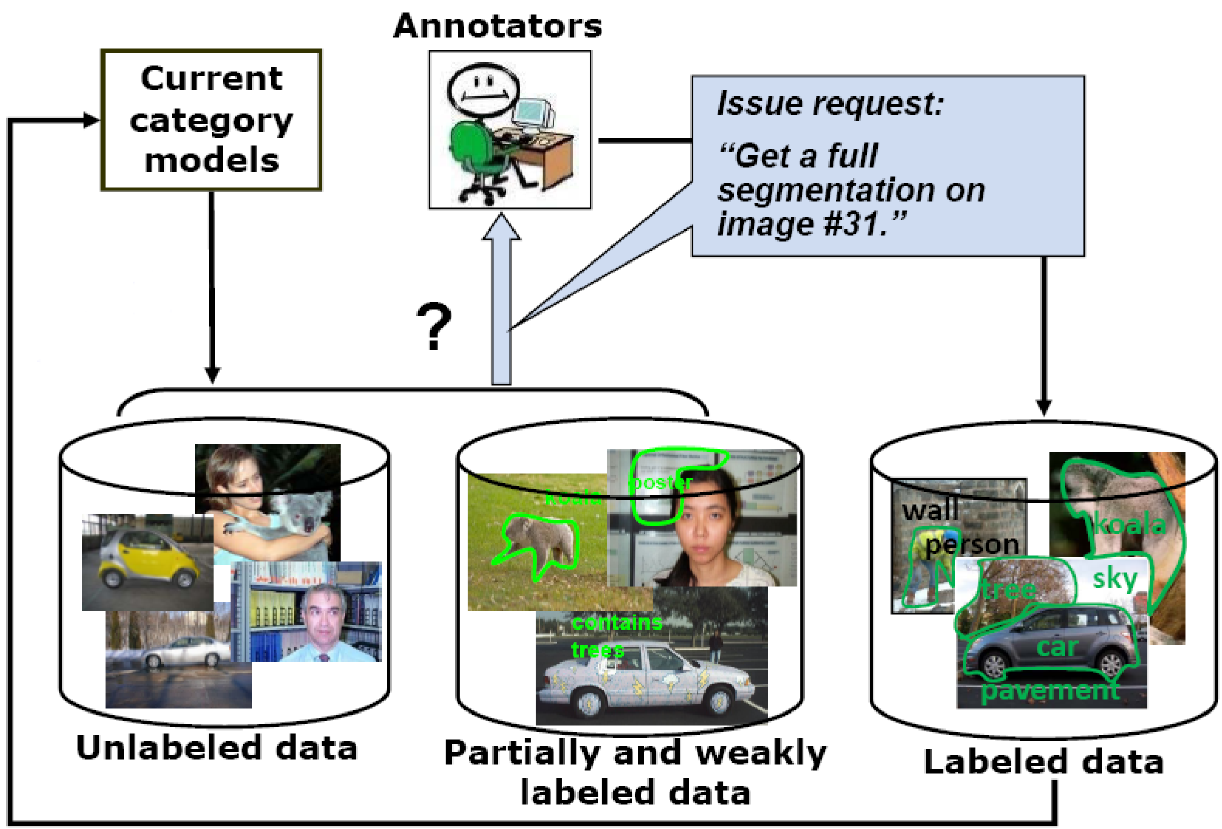
Semi-supervised learning

- Lots of data is available, but only small portion is labeled (e.g. since labeling is expensive)
  - Why is learning from labeled and unlabeled data better than learning from labeled data alone?



Active learning

- The learning algorithm can choose its own training examples, or ask a "teacher" for an answer on selected inputs



S. Vijayanarasimhan and K. Grauman, "Cost-Sensitive Active Visual Category Learning," 2009

Lifelong learning

Read the Web

Research Project at Carnegie Mellon University

Home | Project Overview | Resources & Data | Publications | People

NELL: Never-Ending Language Learning

Can computers learn to read? We think so. "Read the Web" is a research project that attempts to create a computer system that learns over time to read the web. Since January 2010, our computer system called NELL (Never-Ending Language Learner) has been running continuously, attempting to perform two tasks each day:

- First, it attempts to "read," or extract facts from text found in hundreds of millions of web pages (e.g., `playsInstrument(George_Harrison, guitar)`).
- Second, it attempts to improve its reading competence, so that tomorrow it can extract more facts from the web, more accurately.

Browse the Knowledge Base!

So far, NELL has accumulated over 50 million candidate beliefs by reading the web, and it is considering these at different levels of confidence. NELL has high confidence in 2,033,557 of these beliefs — these are displayed on this website. It is not perfect, but NELL is learning. You can track NELL's progress below or @cmunell on Twitter, browse and download its knowledge base, read more about our technical approach, or join the discussion group.

http://rtw.ml.cmu.edu/rtw/

Lifelong learning

Recently-Learned Facts twitter

Refresh

instance	iteration	date learned	confidence
goose_gossage is an athlete	787	16-nov-2013	100.0
fitchburg_state_college is a building	788	19-nov-2013	98.7
kirk_gibson is an actor	787	16-nov-2013	99.0
alex_turner ia a celebrity	787	16-nov-2013	97.5
anthony_r__birley is a criminal	788	19-nov-2013	92.2
the final score of the sports game semi_finals was 6-1	792	01-dec-2013	100.0
national_museum is a museum in the city tokyo	792	01-dec-2013	100.0
w__bush is a U.S. politician endorsed by the U.S. politician john_ashcroft	788	19-nov-2013	93.8
frank004 is a person who graduated from the university state_university	790	24-nov-2013	99.6
mississippi_state_university is a sports team also known as state_university	787	16-nov-2013	99.2

http://rtw.ml.cmu.edu/rtw/

NEIL: Never Ending Image Learner

I Crawl, I See, I Learn.

WHAT COMMON SENSE FACTS HAVE NEIL LEARNED?

Here are a few examples:

```
Airbus_330 can be a kind of / look similar to Airplane.
```

```
Deer can be a kind of / look similar to Antelope.
```

```
Car can have a part Wheel.
```

```
Airbus_330 can have a part Airplane_nose.
```

```
Leaning_tower can be found in Pisa.
```

```
Zebra can be found in Savanna.
```

Xinlei Chen, Abhinav Shrivastava and Abhinav Gupta. NEIL: Extracting Visual Knowledge from Web Data. In ICCV 2013

Thank You