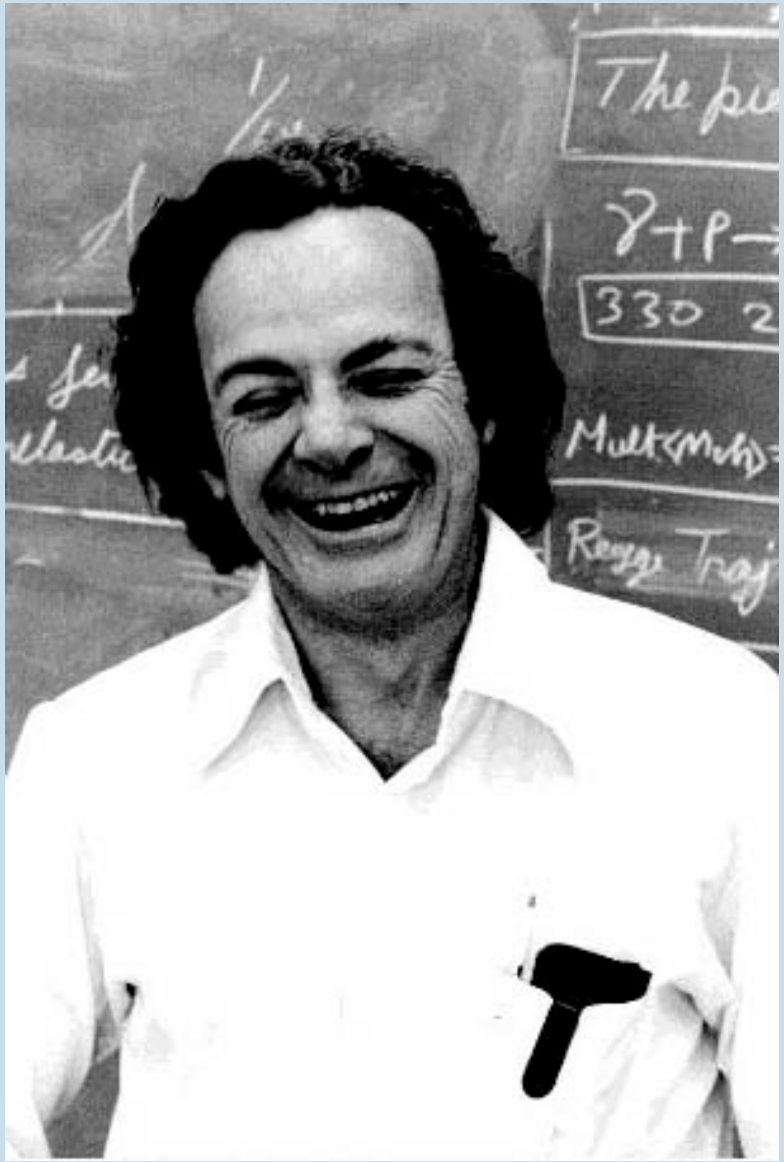
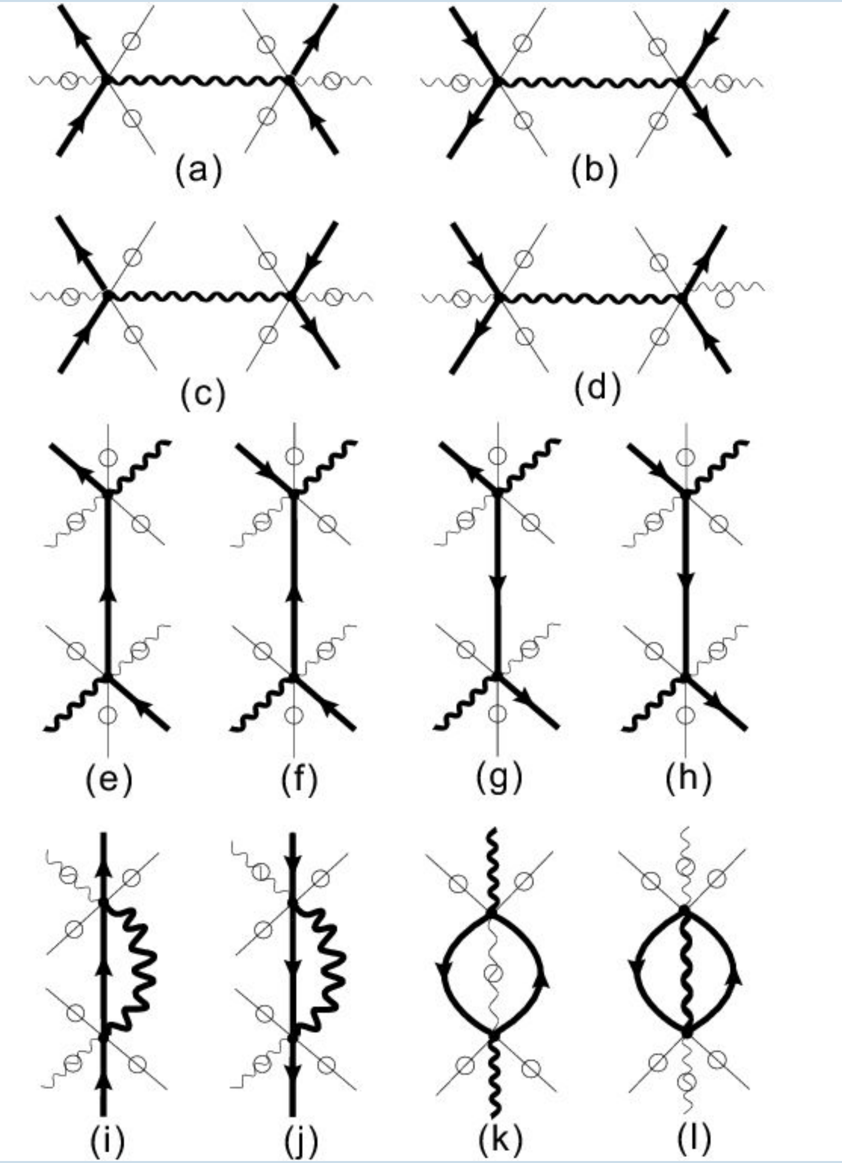
LEARN LIKE A MACHINE




(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)

A few AI applications today

A LOT OF NUMBER CRUNCHING

- BUSINESS INTELLIGENCE
- IOT PREDICTIVE MAINTENANCE
- SEARCH RECOMMENDATIONS
- FORECASTING MODELS

VISION

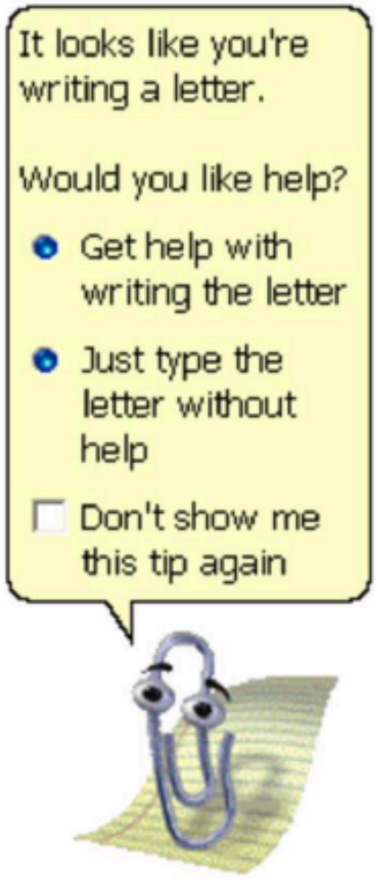
- AUTO TECH AND DRONE COLLISION AVOIDANCE
- E-COMMERCE SEARCH
- PICK AND PLACE ROBOTS
- HEALTHCARE DIAGNOSTICS

LANGUAGE PROCESSING

- CHATBOTS
- NEWS & MEDIA CONTENT CREATION
- SMART HOME VOICE INTERFACES
- TEXT ANALYTICS

Source: www.cbinsights.com

FROM CLIPPY TO ALEXA…



"Alexa, ask Uber to request a ride."

"Alexa, ask Fitbit how I slept last night.

"Alexa, ask the bartender, what's in a Tom Collins?"

"Alexa, tell Tide I have a juice stain."

Source: www.cbinsights.com

A computer program is said to learn from **experience** *E* with respect to some class of **tasks** *T* and **performance measure** *P*, if its performance at tasks in *T*, as measured by *P*, improves with experience *E*.

Tom Mitchell, *Machine Learning* (1997)

Images/Videos

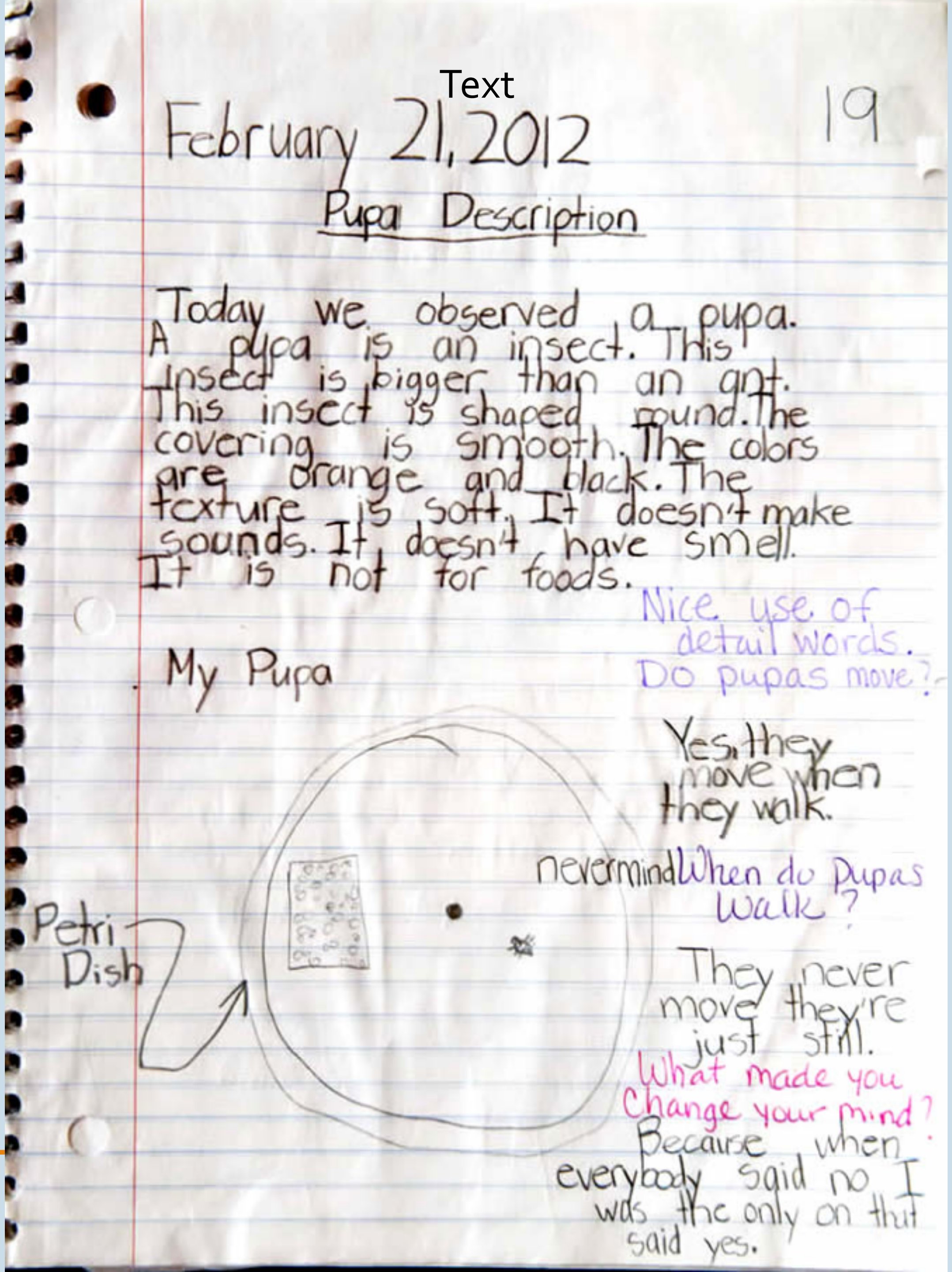
Tasks

Label: “Motorcycle”

Summarize: “Motorcyclist turning right”

Search: Is there a helmet in this picture?

Text


February 21, 2012

19

Pupa Description

Today we observed a pupa. A pupa is an insect. This insect is bigger than an ant. This insect is shaped round. The covering is smooth. The colors are orange and black. The texture is soft. It doesn't make sounds. It doesn't have smell. It is not for foods.

Nice use of detail words. Do pupas move?

My Pupa

Yes, they move when they walk.

nevermind When do Pupas walk?

They never move they're just still.

What made you change your mind?

Because when everybody said no I was the only on that said yes.

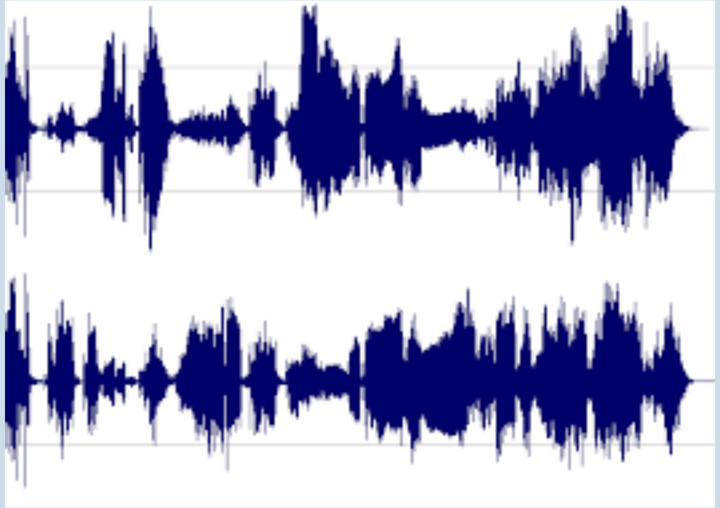
Tasks

Summarize: "Evidence leads to change of mind."

Search: "What is a pupa?"

Generate: Write text in this style

Speech/Audio



Tasks

Recognize: Who’s speaking?

Identify: What’s being said?

Translate: Speak this sentence in Italian.

Numerical

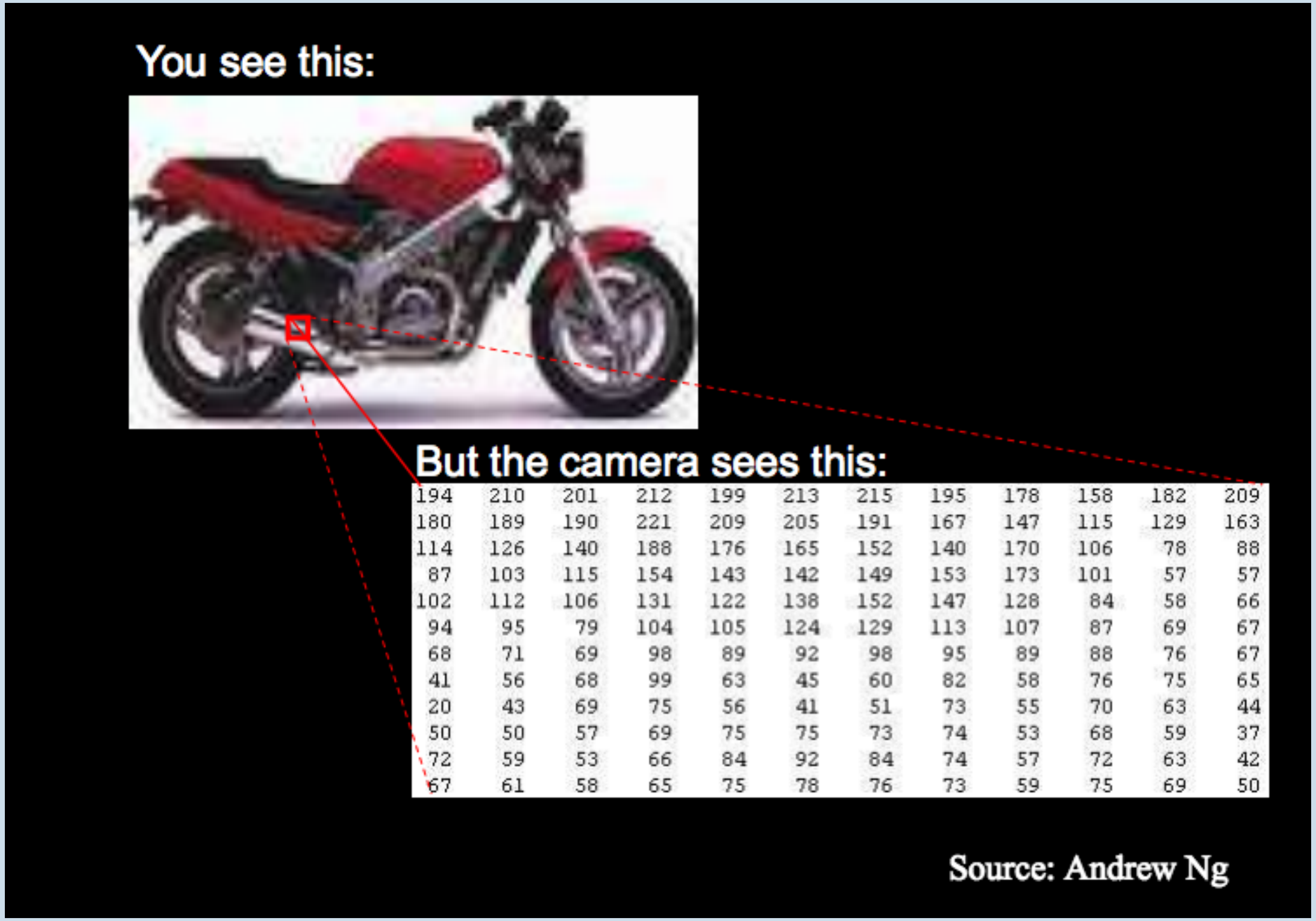
1 LastName	2 Sex	3 Age	4 Weight	5 Smoker	6 BloodPressure		7 Trials
'SMITH'	Male	38	176	1	124	93	18
'JOHNSON'	Male	43	163	0	109	77	[11,13,22]
'WILLIAMS'	Female	38	131	0	125	83	[]
'JONES'	Female	40	133	0	117	75	[6,12]
'BROWN'	Female	49	119	0	122	80	[14,23]
'DAVIS'	Female	46	142	0	121	70	19
'MILLER'	Female	33	142	1	130	88	13
'WILSON'	Male	40	180	0	115	82	[]
'MOORE'	Male	28	183	0	115	78	2
'TAYLOR'	Female	31	132	0	118	86	11

Tasks

Classify: Is person X (who is not in the data set) a smoker?

Predict: What is person X's blood pressure?

To do these tasks well requires a large amount of data – but not necessarily big data.



Data = A row of numbers

Dataset = A table of numbers

Inputs to a machine learning system are just tables of numbers.

$$a_1 * 26 + a_2 * 3.8 = 62$$

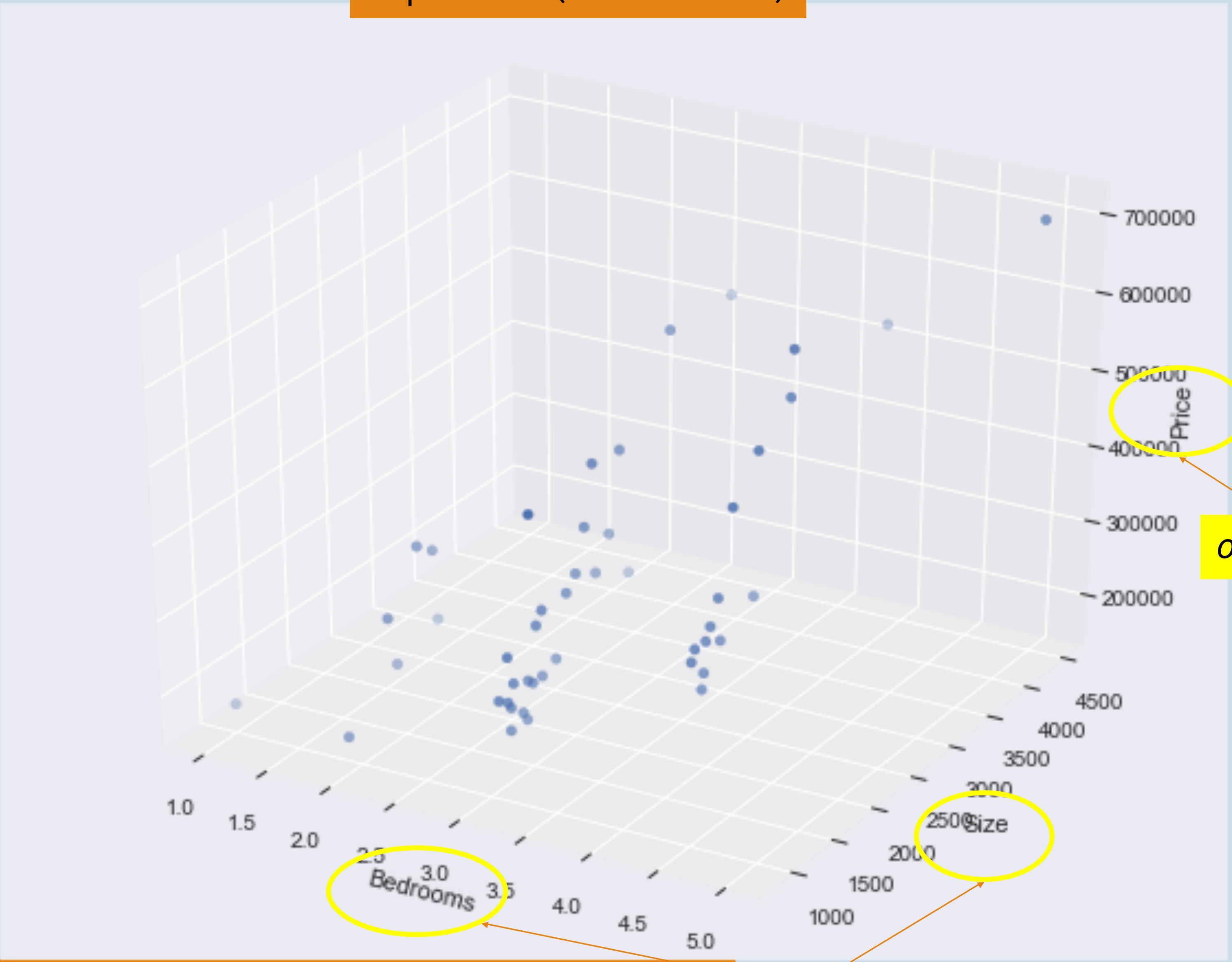
One row of the dataset

$$a_1 * 26 + a_2 * 3.8 = 62$$

Parameters

Model

Experience (aka the Data)



output

inputs/features

Model + Data + Parameters

$$a_1 * s_1 + a_2 * b_1 = p'_1$$
$$a_1 * s_2 + a_2 * b_2 = p'_2$$
$$\dots$$
$$a_1 * s_m + a_2 * b_m = p'_m$$

parameters

predictions

Task

Given the number of bedrooms and the size of a house, predict the price of the house.

Key Machine Learning Question

What are the right parameter values?

The total cost of getting it wrong – a measure of learning performance

$$a_1 * b_1 + a_2 * s_1$$

cost of getting it wrong

$$(prediction - price)^2$$

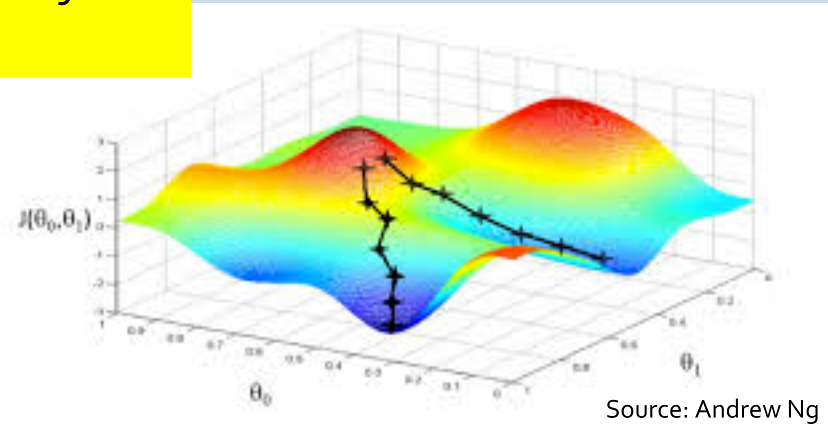
s_1 b_1

Size	Bedrooms	Price	Prediction	Cost
2104	3	399,900	789,498	151,786,601,604
1600	3	329,900	100,022	52,834,894,884
2400	3	369,000	564,300	38,142,090,000
1416	2	232,000	-345,390	333,379,212,100
			Total Cost	**424,507,983,585**

Goal: Find the parameters a1 and a2 that minimize *total* cost over *the entire dataset*.

Machine learning problems are giant numerical optimization problems.

The cost of getting it wrong as a function of the parameter values



Source: Andrew Ng

To learn is to get to the lowest point in the terrain. To do this consistently we need a reliable method for finding this lowest point. A common technique is Gradient Descent.

Image Credit: Jason Brownlee

Type of Data

Type of Prediction

	Labeled	Unlabeled
Regression: Predict a Real number	$a_1 * s + a_2 * b$	The emphasis is on the learning algorithm rather than the model
Classification: Predict it belongs in X rather than Y	$\frac{1}{1 + e^{-(a_1 * s + a_2 * b)}}$	

models

Models are determined by the type of data you have and the type of prediction you're looking to make.

parameters

features

Use to predict a real number: regression

$$a_1 * s + a_2 * b$$

Linear combination by multiplication and addition; 2 unknown parameters; number of features are usually ~ 100s or more; # parameters = # features

Structure

Use to classify

$$\frac{1}{1 + e^{-(a_1 * s + a_2 * b)}}$$

Non-linear combination by exponentiation, multiplication and addition; 2 unknown parameters; number of features are usually ~ 100s or more; # parameters = # features

$$a_1 * s + a_2 * s^2 + a_3 * sb + a_4 * b + a_5 * b^3$$

$$\frac{1}{1 + e^{-(a_1 * s + a_2 * s^2 + a_3 * sb + a_4 * b + a_5 * b^3)}}$$

It's easy to make non-linear models using the same basic structure; these are 3rd order models with 5 features, hence 5 parameters

Predict a number (continuous) or predict a class (discrete). That's what prediction is in machine learning.

Note: There is actually one additional parameter (and a corresponding feature that's always equal to 1) in the models above. I've left it out to keep the expressions simpler.

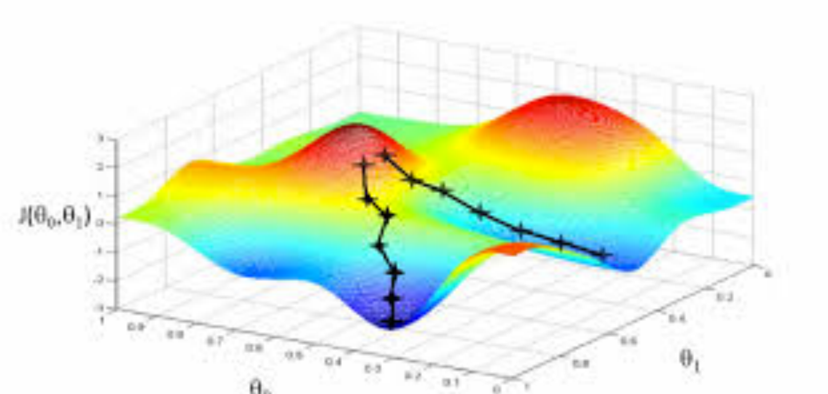
Fifth-order Feature Set

of bedrooms

	1	b	b^2	b^3	b^4	b^5
1	1	b	b^2	b^3	b^4	b^5
s	s	sb	sb^2	sb^3	sb^4	sb^5
s^2	s^2	s^2b	s^2b^2	s^2b^3	s^2b^4	s^2b^5
s^3	s^3	s^3b	s^3b^2	s^3b^3	s^3b^4	s^3b^5
s^4	s^4	s^4b	s^4b^2	s^4b^3	s^4b^4	s^4b^5
s^5	s^5	s^5b	s^5b^2	s^5b^3	s^5b^4	s^5b^5

Learning System = Model + Cost Function + Method for Finding Optimal Cost

$$a_1 * s + a_2 * b$$

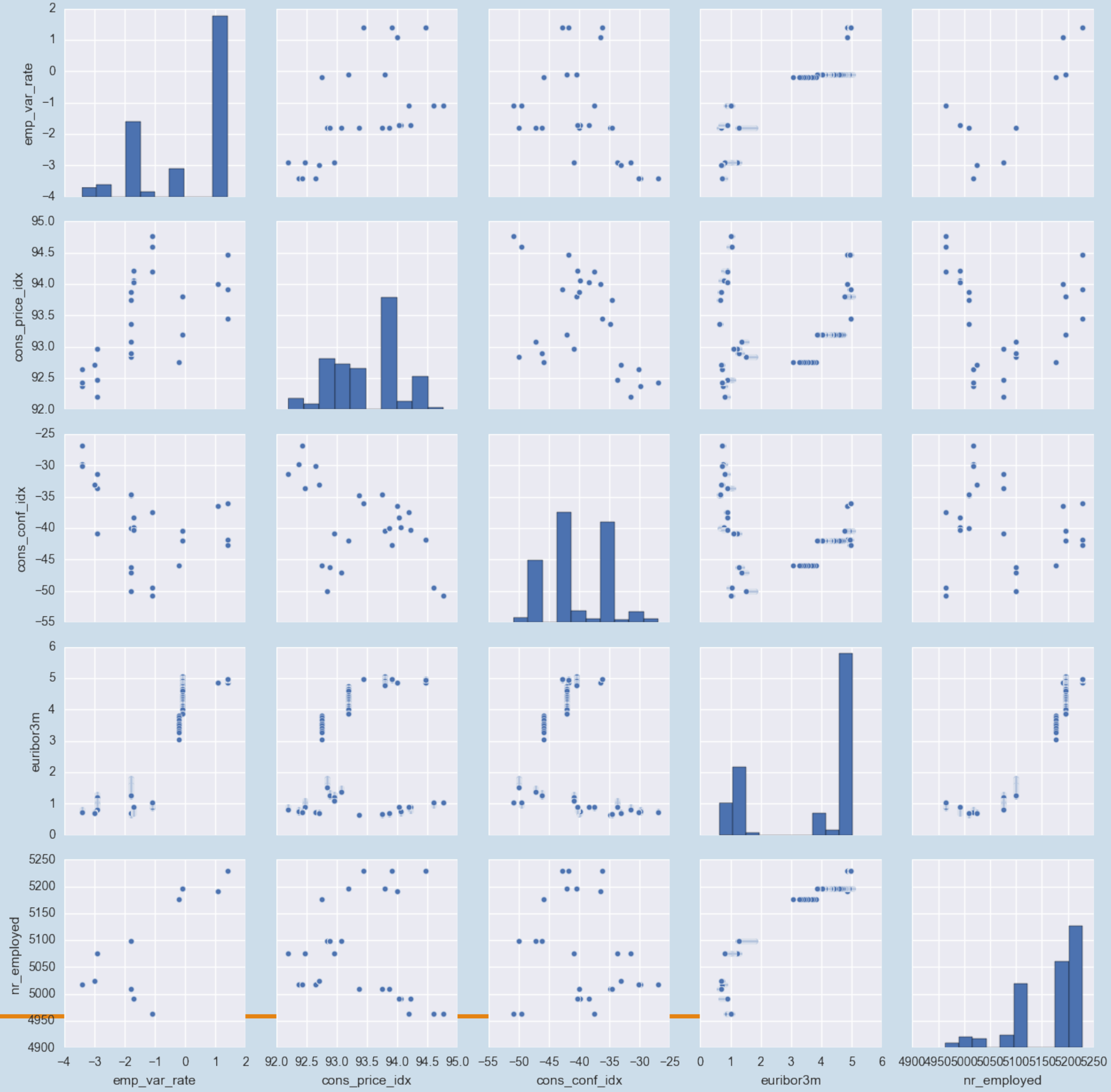


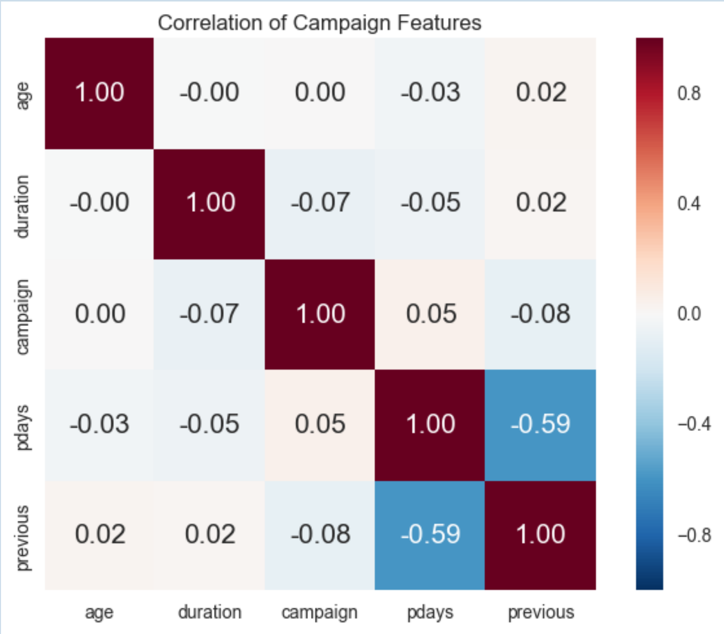
Source: Andrew Ng

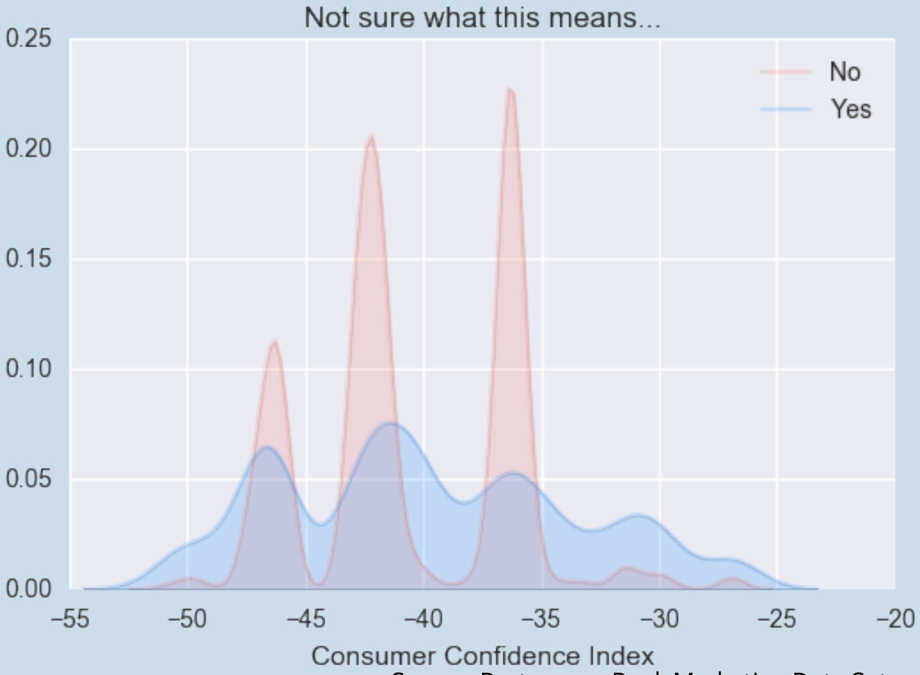
- What is the structure of the model?
- What kinds of features should we use?
- How many features?
- Linear or non-linear?

- What's the right cost function?
- What's the right algorithm for finding the optimal cost?
- How much data should be used on each iteration?
- To what degree should the features be discounted (regularized)?
- What should the learning rate be?
- How many learning iterations?

What's the right measure of *system* performance?







Source: Portuguese Bank Marketing Data Set,
https://archive.ics.uci.edu/ml/datasets/Bank+Marketing

Per capita consumption of mozzarella cheese (US)

correlates with

Civil engineering doctorates awarded (US)

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Per capita consumption of mozzarella cheese (US) *Pounds (USDA)*	9.3	9.7	9.7	9.7	9.9	10.2	10.5	11	10.6	10.6
Civil engineering doctorates awarded (US) *Degrees awarded (National Science Foundation)*	480	501	540	552	547	622	655	701	712	708

Correlation: 0.958648

Source: tylervigen.com/old-version.html

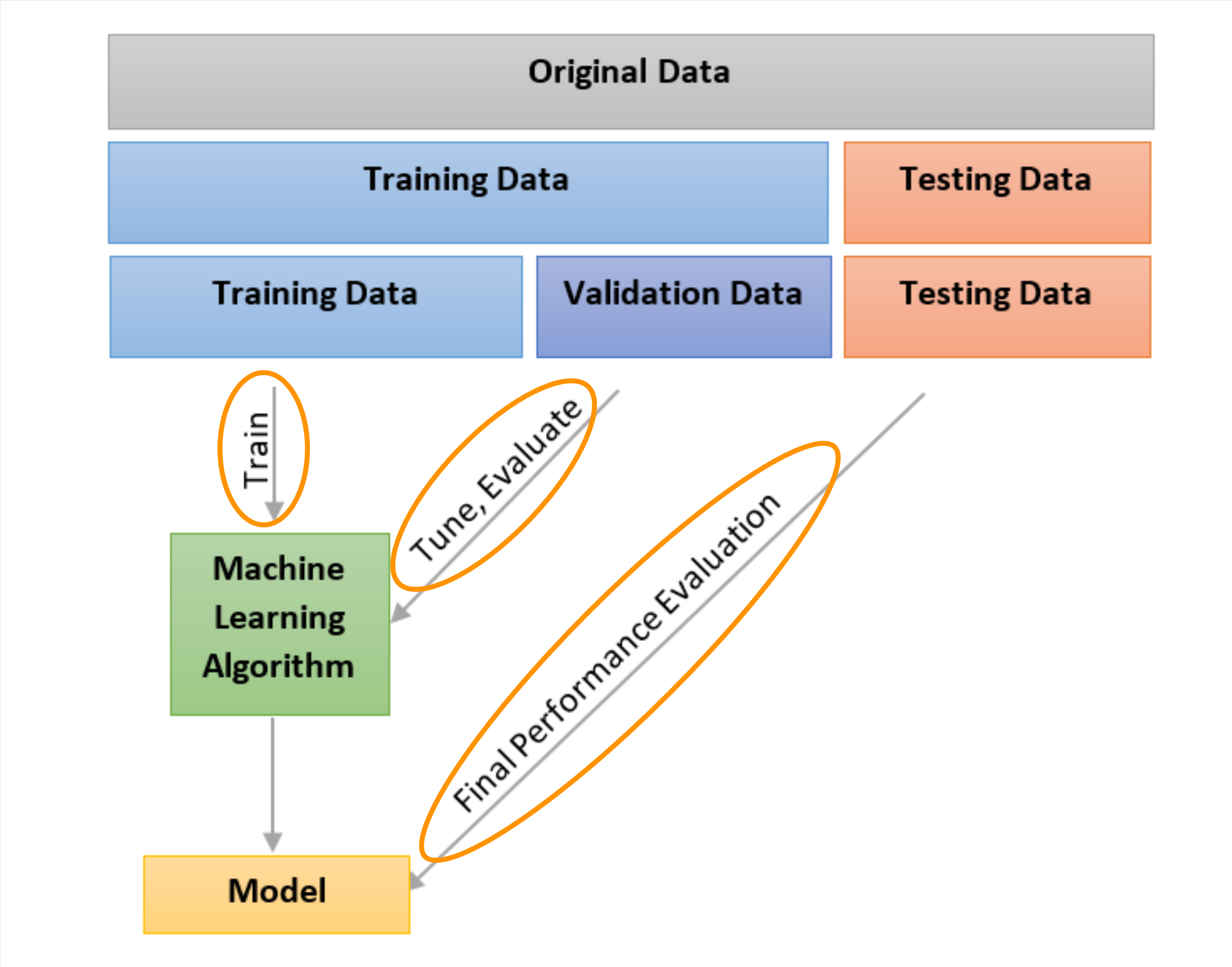
- Are we predicting a number or a class?
- What kinds of features should we use?
- How many features?
- Linear or non-linear?

- What's the right cost function?
- What's the right algorithm for finding the optimal cost?
- How much data should be used on each iteration?
- To what degree should the features be discounted (regularized)?
- What should the learning rate be?
- How many learning iterations?

- **How do you take the available data and translate it into a form that can be fed into a machine learning system?**

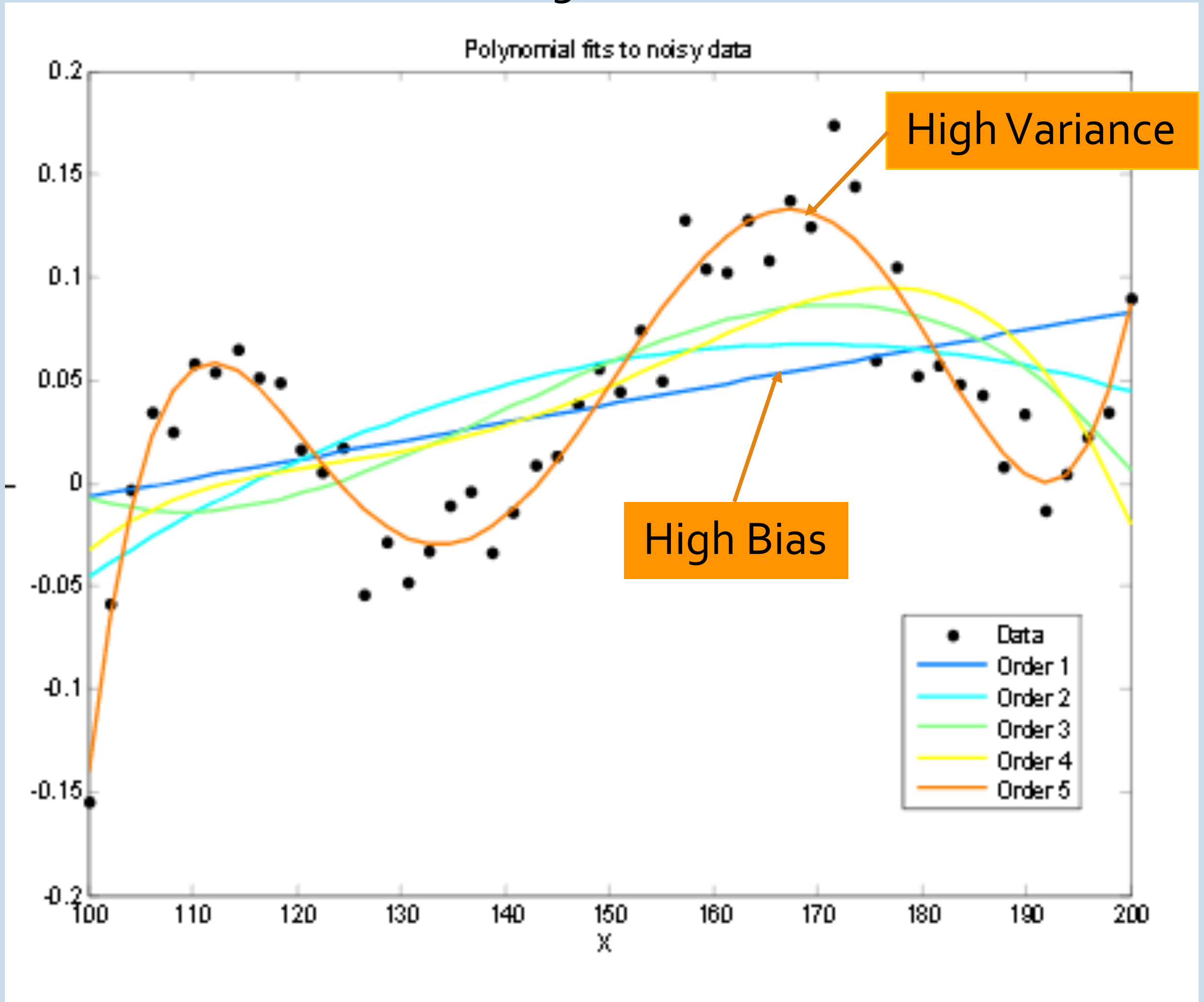
- **What's the right measure of my learning system's performance?**

This is where you can exercise creativity even if you're not knee deep in the model building and manipulation.

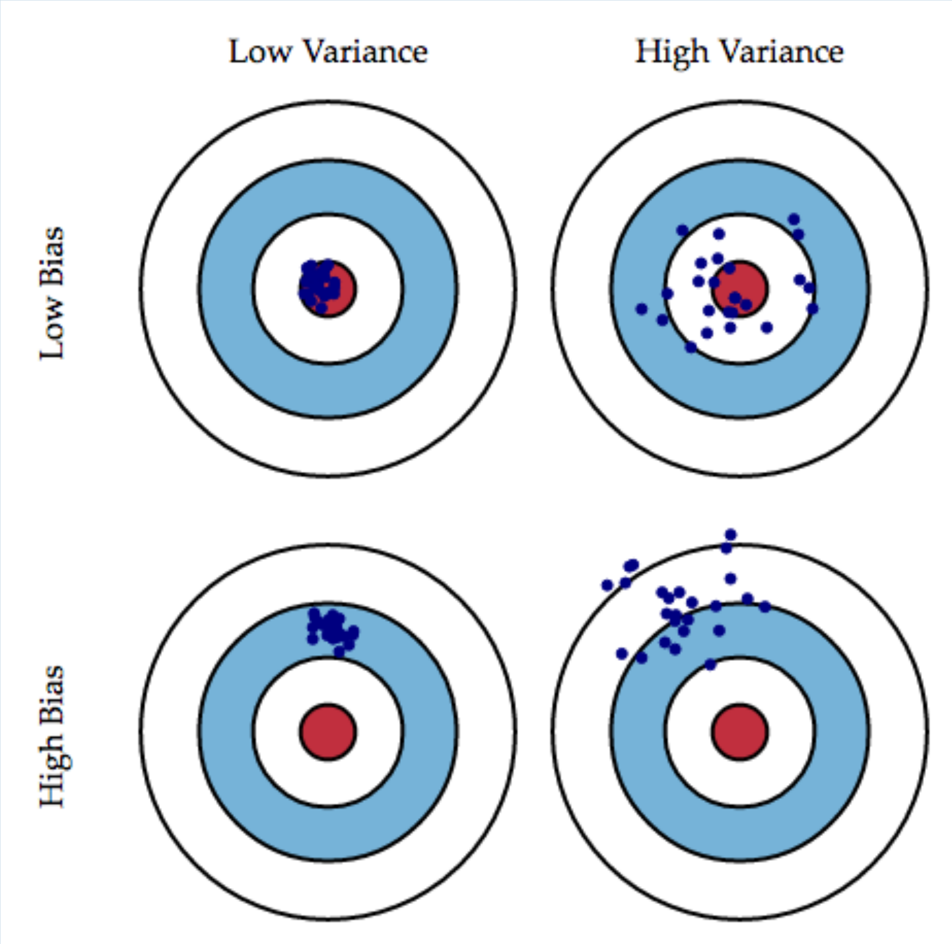


Original Data
Training Data
Testing Data
Training Data
Validation Data
Testing Data
Train
Tune, Evaluate
Final Performance Evaluation
Machine Learning Algorithm
Model

Bias and Variance for Regression Problems

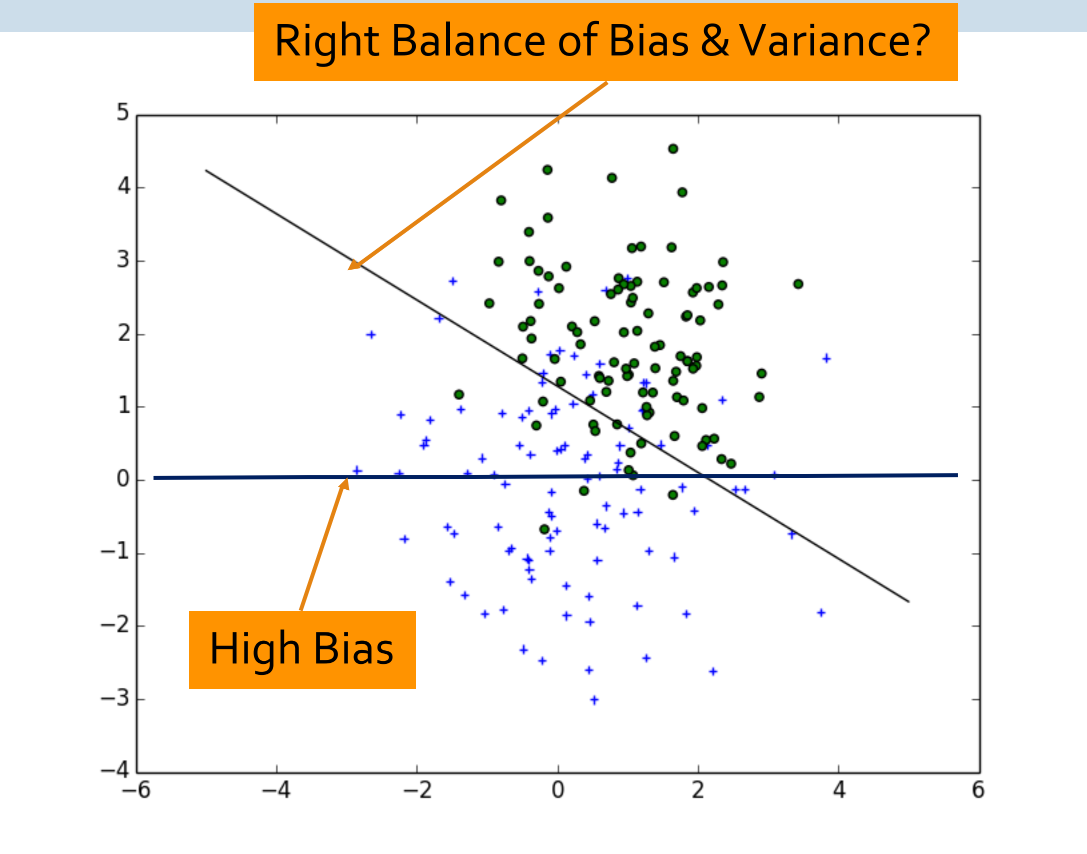


Source: http://stackoverflow.com/questions/17128847/plotting-polynomials-of-best-fit

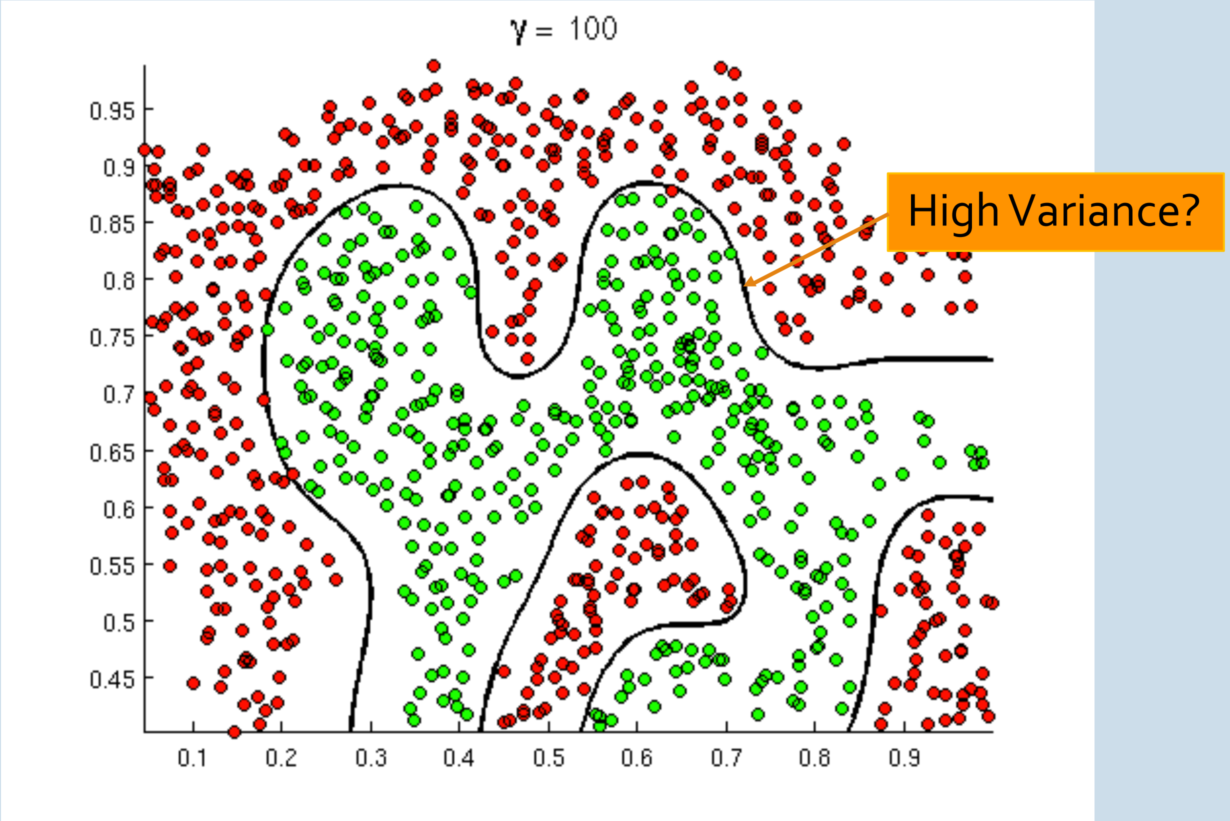


Source: Scott Fortman-Roe, http://scott.fortmann-roe.com/docs/BiasVariance.html

Bias and Variance for Classification Problems

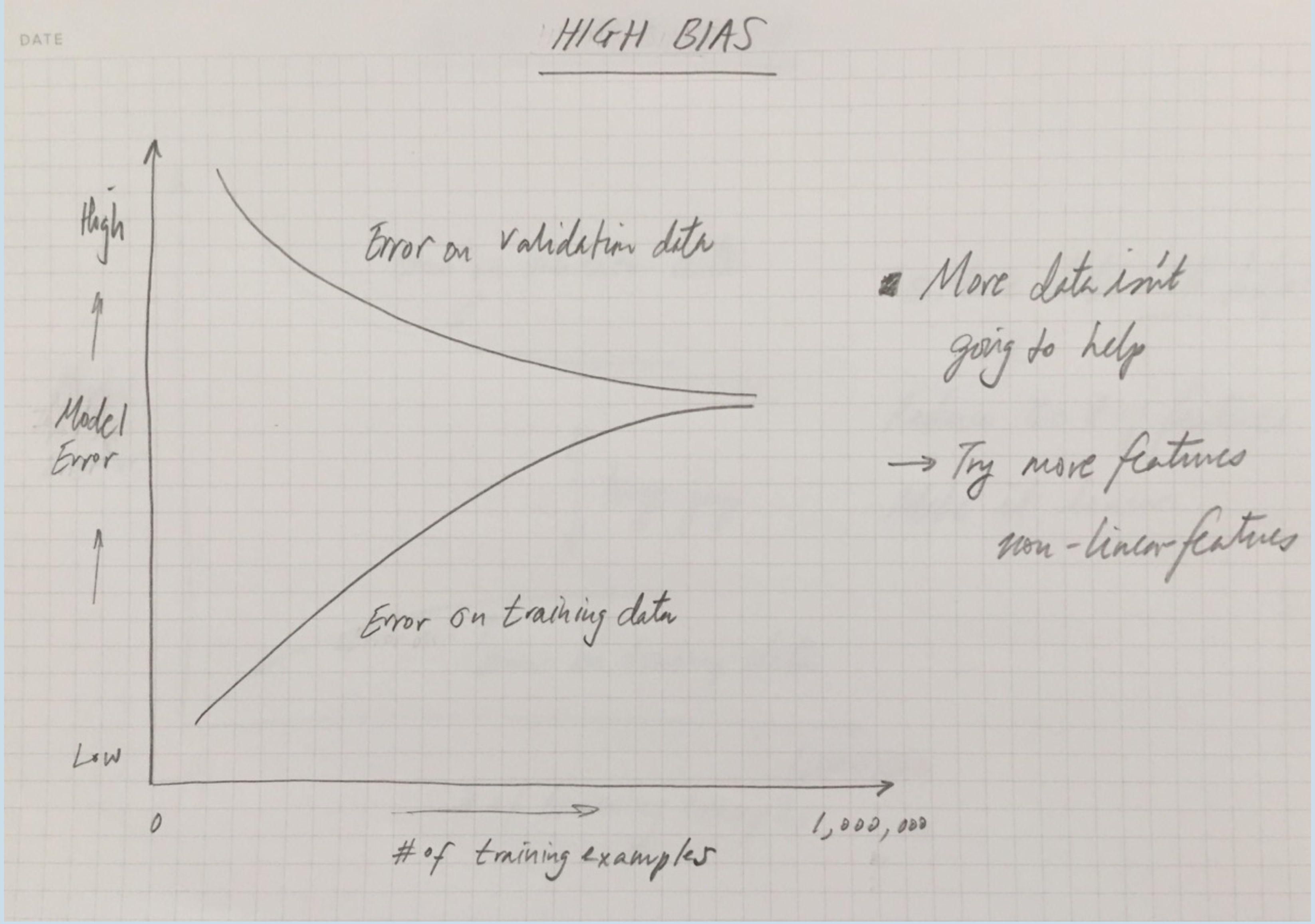


Source: http://stackoverflow.com/questions/22294241/plotting-a-decision-boundary-separating-2-classes-using-matplotlibs-pyplot



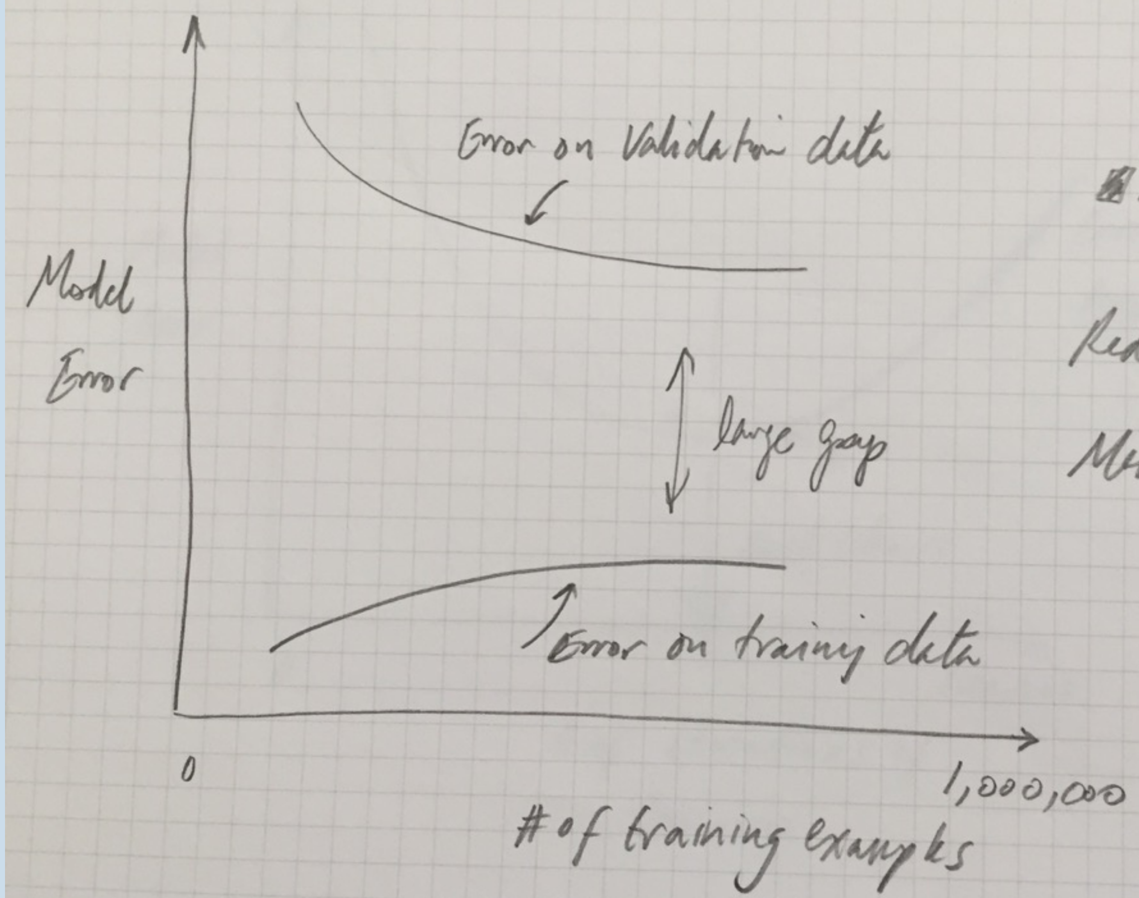
Source: http://openclassroom.stanford.edu/MainFolder/DocumentPage.php?course=MachineLearning&doc=exercises/ex8/ex8.html

"Prediction is hard, especially about the future." – you know who


HIGH BIAS
DATE
High
Model Error
Low
Error on validation data
Error on training data
0
1,000,000
# of training examples
More data isn't going to help
→ Try more features
non-linear features


HOW NON-LINEAR?
High Bias Regime
High Variance Regime
Error on Validation data
Model Error
Error on training data
Best Choice
Low
High
Non Linearity

HIGH VARIANCE

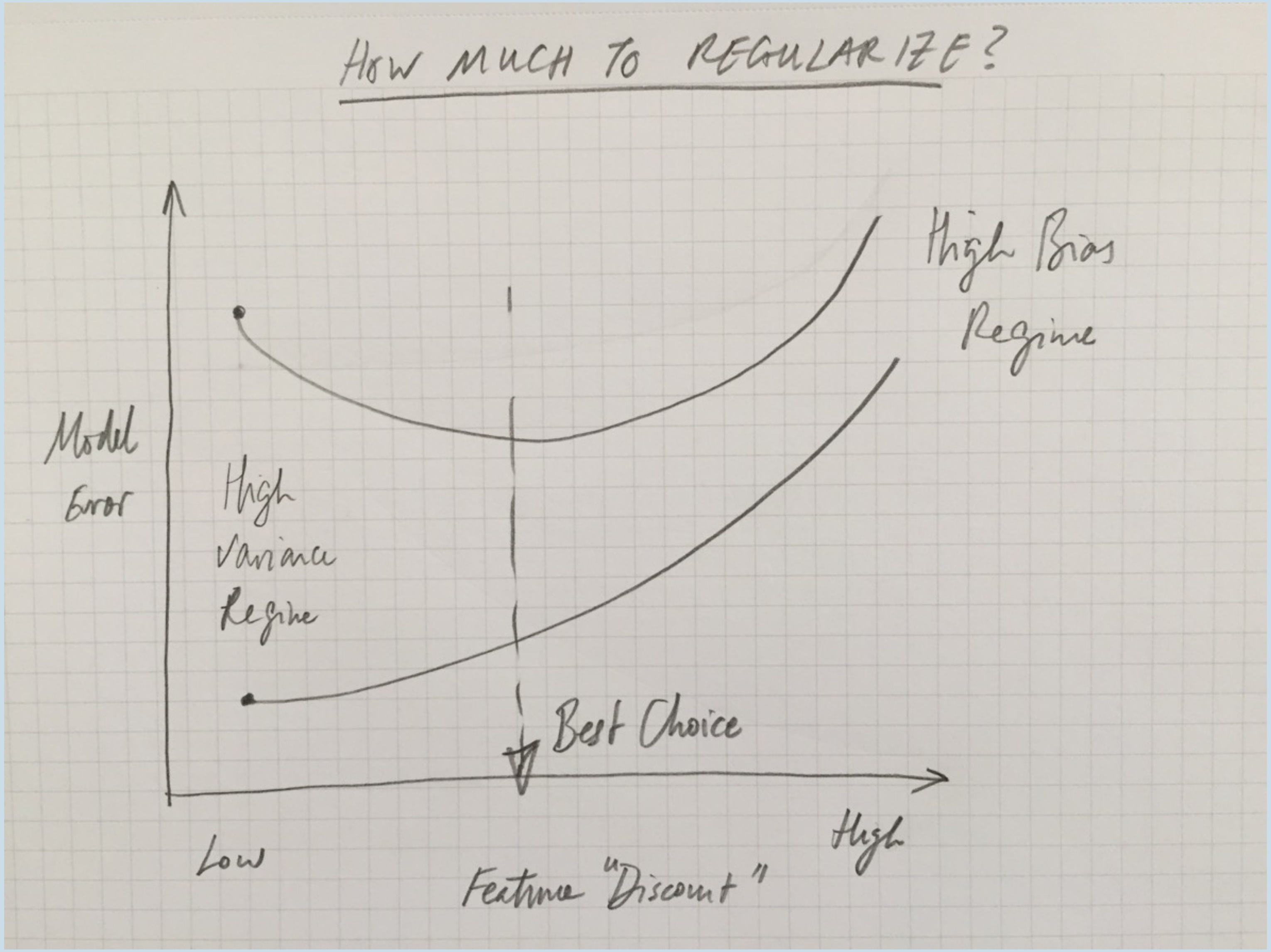


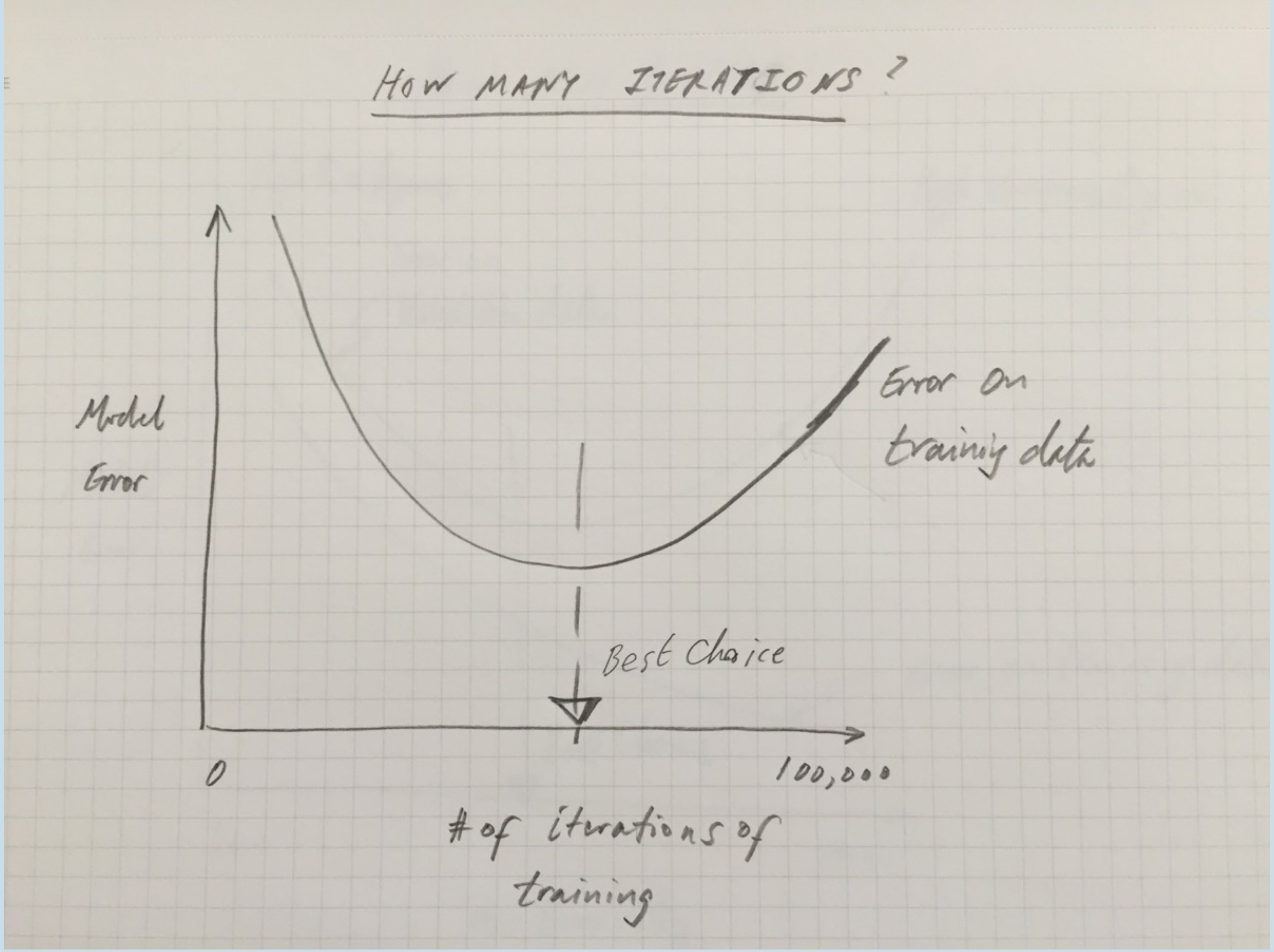
More data will help!

OR

Reduce the # of features

Make it linear


HOW MUCH TO REGULARIZE?
Model Error
High Variance Regime
High Bias Regime
Best Choice
Low
High
Feature "Discount"


HOW MANY ITERATIONS?
Model Error
Error on training data
Best Choice
0
100,000
# of iterations of training

Imbalanced Data

"Learned Model": Predict terrorist if person is named Tim

Assumptions:
- 1 in 1,000 people are named Tim
- 1 in 100,000 people are terrorists
- 1,000,000 people screened

	Terrorist	Not Terrorist	Total
Tim	1 (True Positive)	999 (False Positive)	1,000
Not Tim	9 (False Negative)	998,991 (True Negative)	999,000
Total	10	999,990	1,000,000

$$Accuracy = \frac{t_p + t_n}{t_p + t_n + f_p + f_n}$$

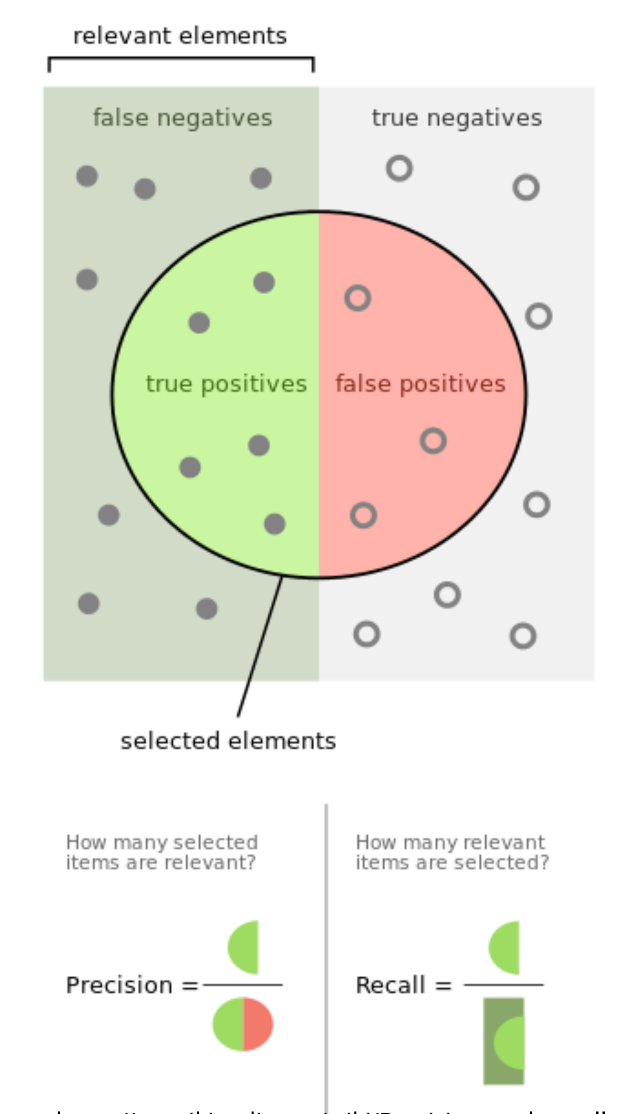
(1 + 998,991)/1,000,000 = 0.998992 99.89% accurate!

$$Precision = \frac{t_p}{t_p + f_p}$$

1 /999 = 0.001001 0.1% precision

$$Recall = \frac{t_p}{t_p + f_n}$$

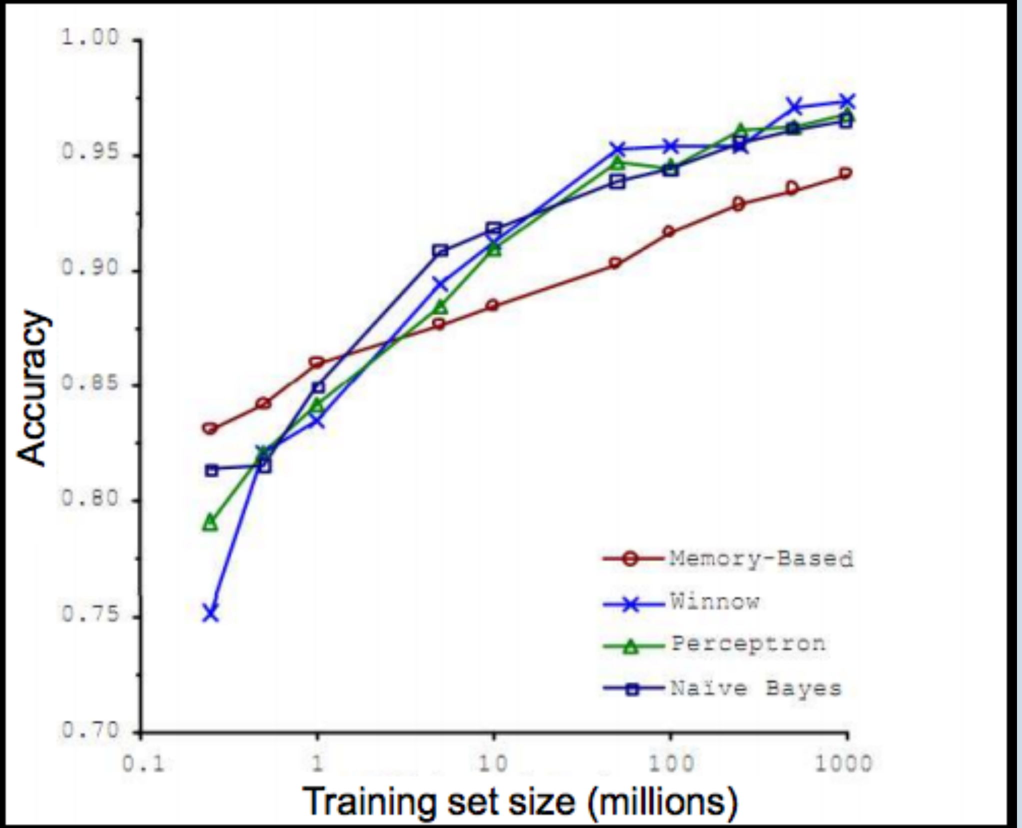
1 /(1 + 9) = 0.1 10% recall



Source: https://en.wikipedia.org/wiki/Precision_and_recall

Prediction accuracy is not always the right measure of system performance. Watch out for imbalanced datasets!

- Choices of learning algorithm:
  - Memory based
  - Winnow
  - Perceptron
  - Naïve Bayes
  - SVM
  - ....
- What matters the most?

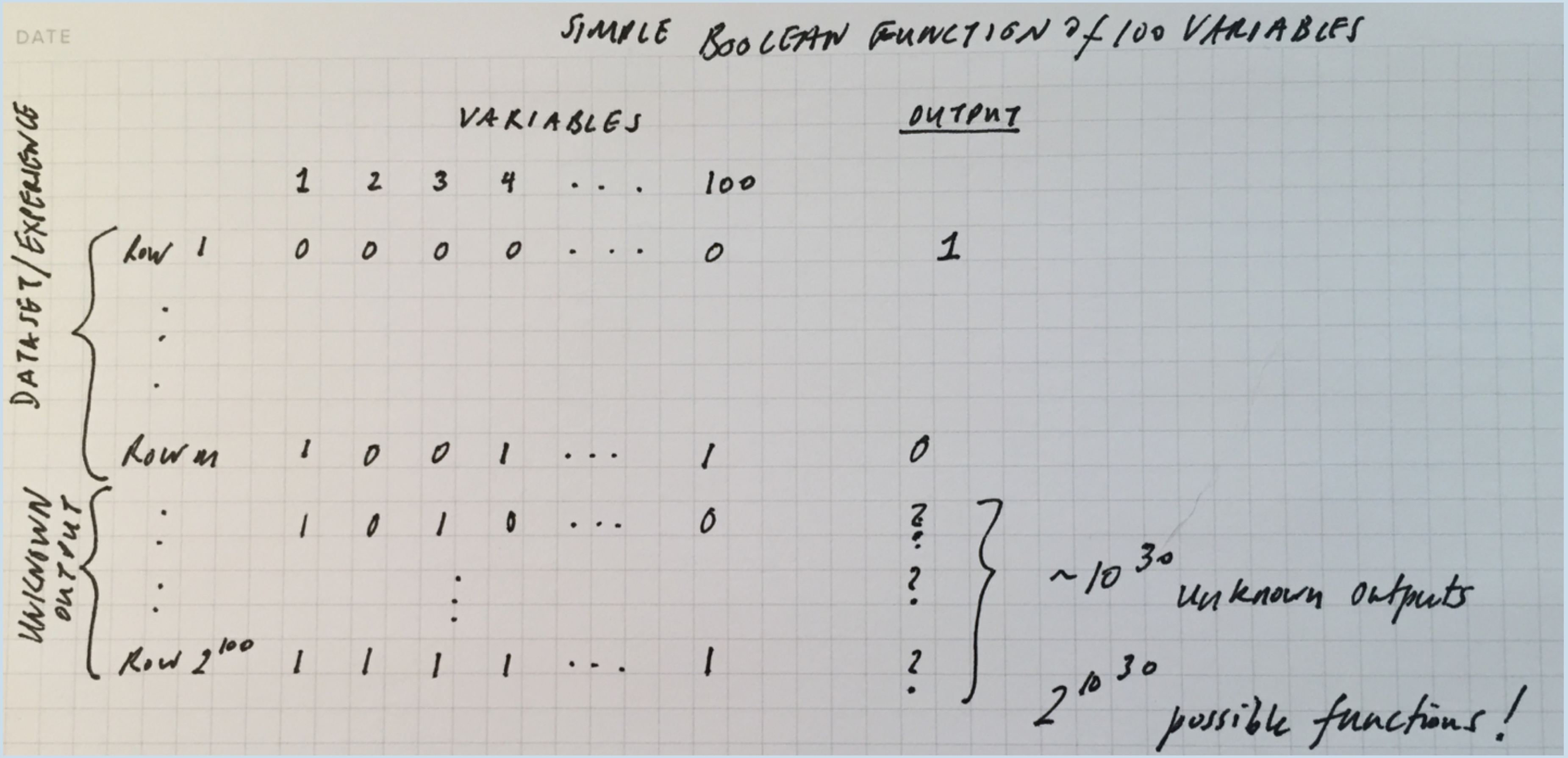


[Banko & Brill, 2001]

"It's not who has the best algorithm that wins.
It's who has the most data."

Source: Andrew Ng

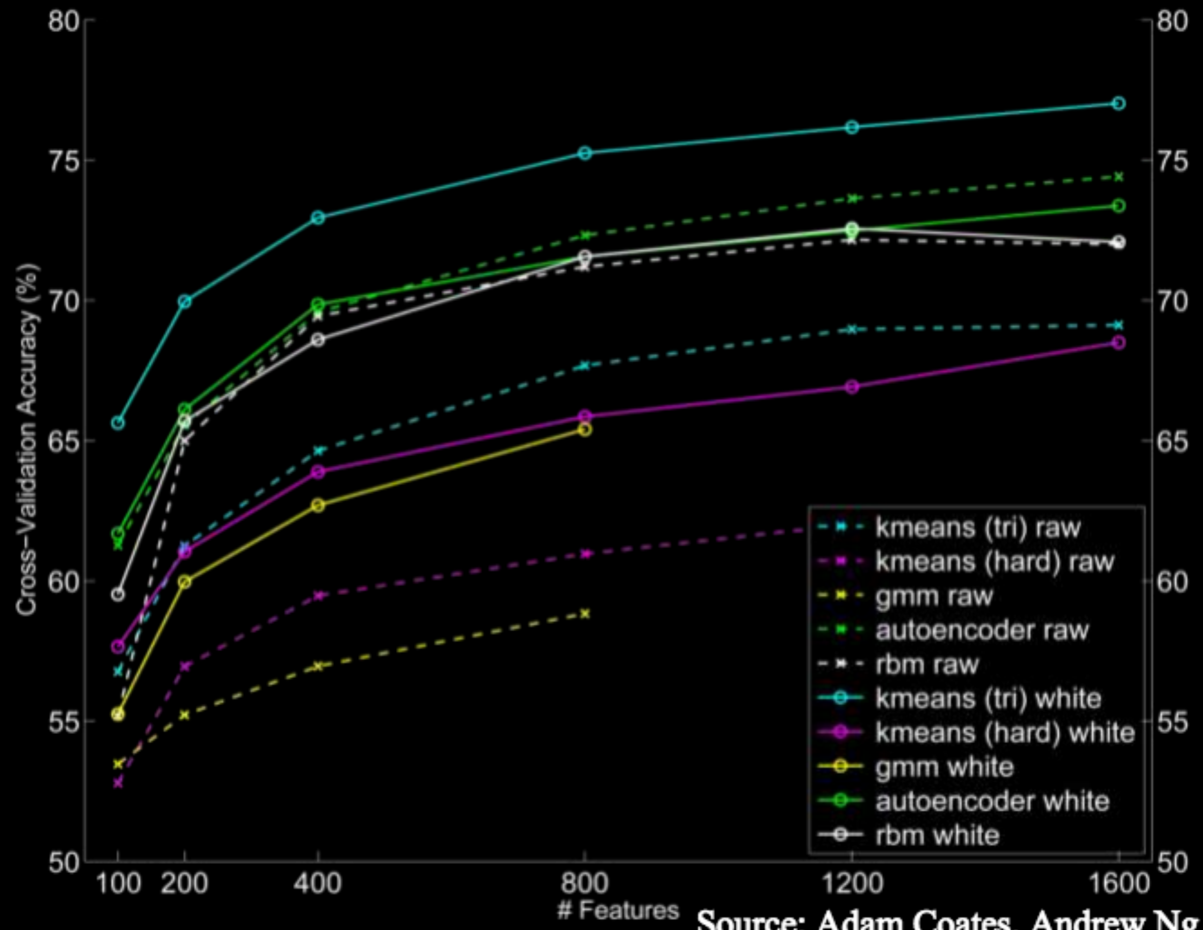
Data alone is not enough



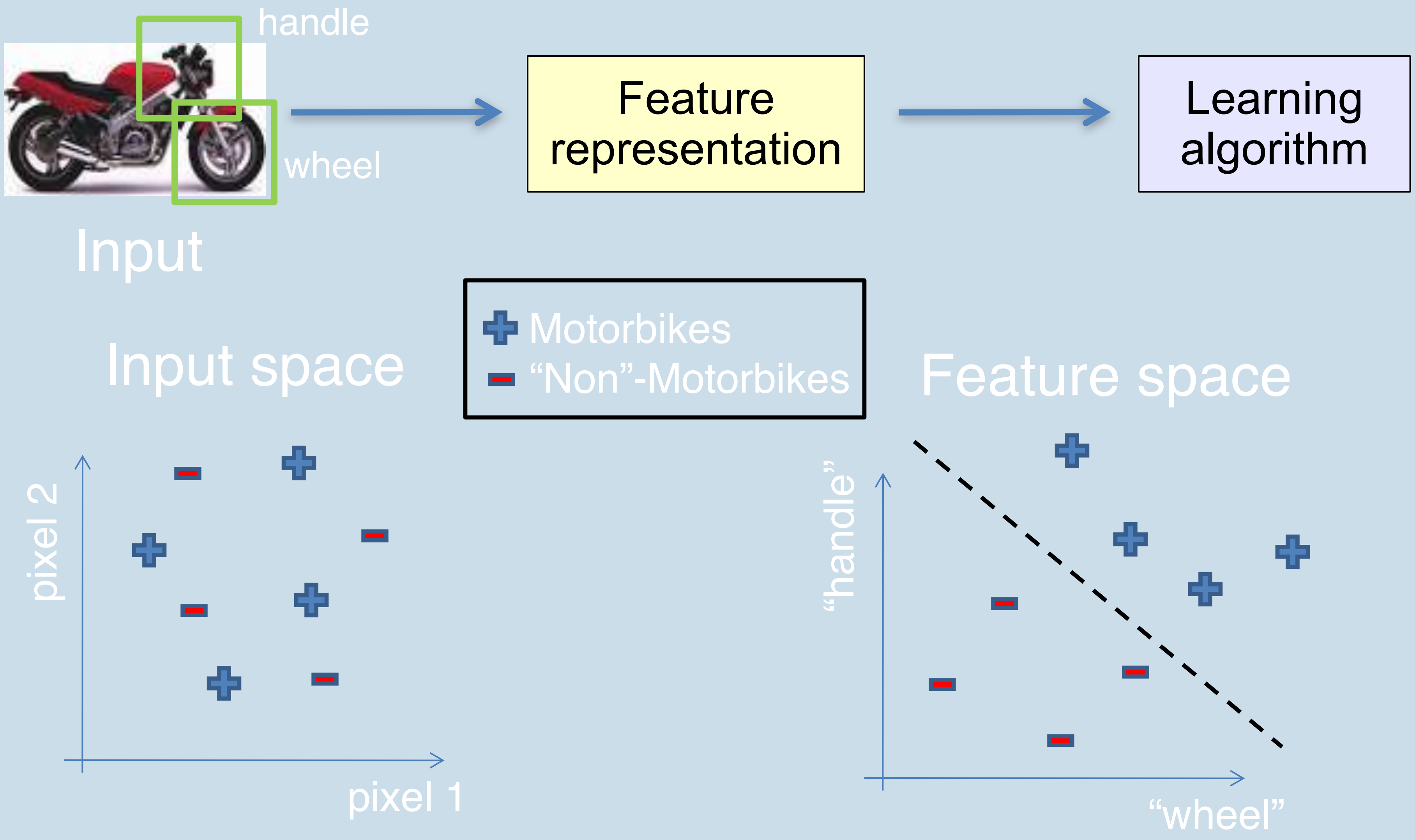
Problem: Even if you have a trillion rows of data to learn from, you still need to learn $2^{100} - 2^{12}$ rows!

"Luckily, the functions we want to learn in the real world are *not* drawn uniformly from the set of all mathematically possible functions!" (Pedro Domingos)

Large numbers of features is critical. The specific learning algorithm is important, but ones that can scale to many features also have a big advantage.

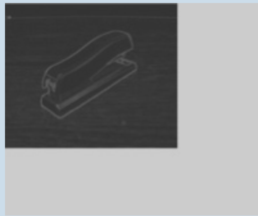


Source: Adam Coates, Andrew Ng


handle
wheel
Input
Feature representation
Learning algorithm
Motorbikes
“Non”-Motorbikes
Input space
pixel 2
pixel 1
Feature space
“handle”
“wheel”

Object detection

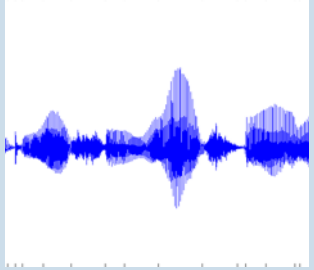
Image



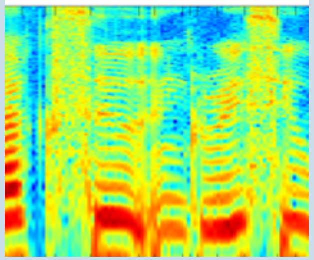
Low-level vision features

Recognition

Audio classification



Audio

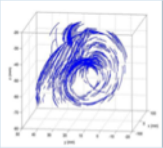


Low-level audio features

Speaker identification

Helicopter control

Helicopter



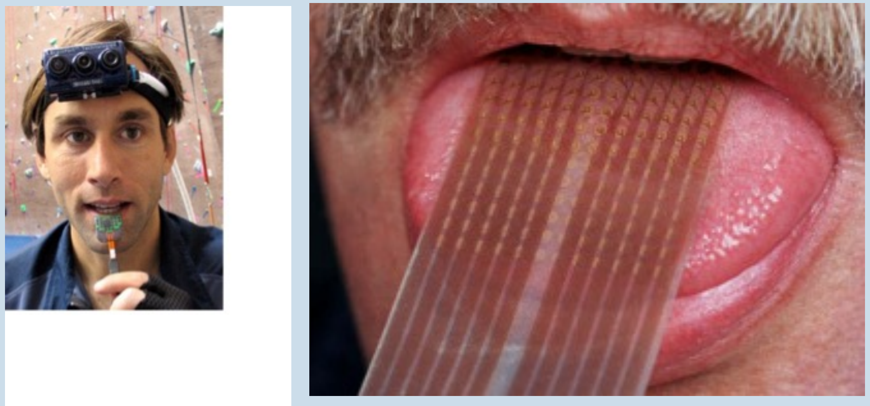
Low-level state features

Action

Problems of hand-tuned features

1. Need expert knowledge
2. Time-consuming and expensive
3. Does not generalize to other domains

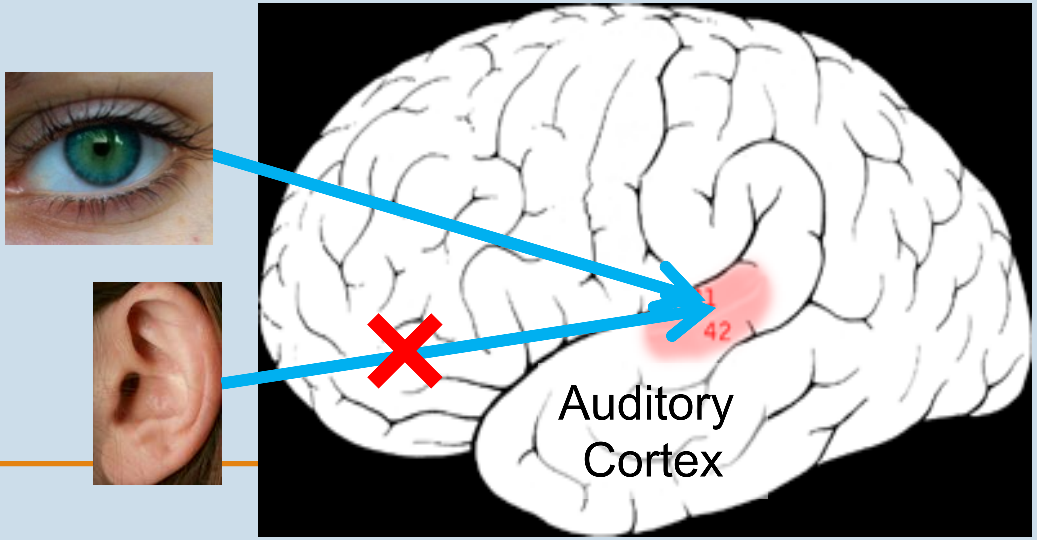
The One Learning Algorithm Hypothesis



Seeing with your tongue

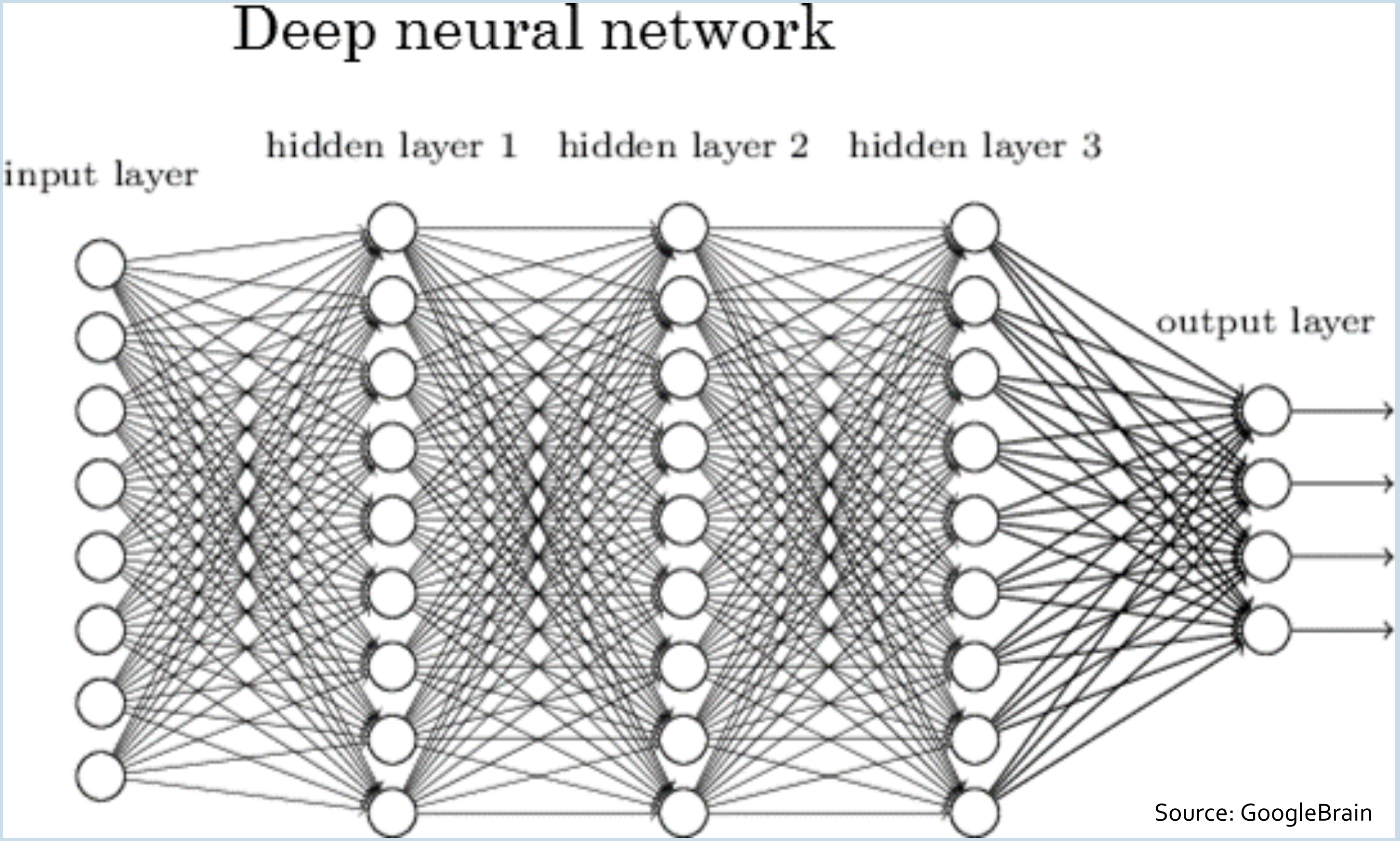
Human echolocation (sonar)

Sensory cortex learns to see.



Auditory cortex learns to see.

Source: Andrew Ng; BrainPort; Martinez et al; Roe et al.

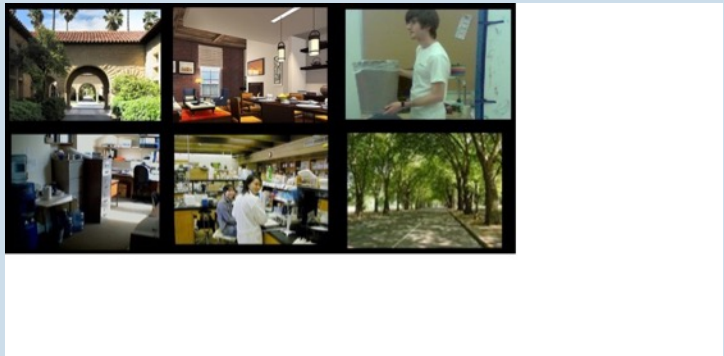


Universality Theorem: A three-layer neural network can calculate *any* function (to a close approximation).

CAUTION: That a function can be *calculated* doesn't mean it can be *learned*.

Check out Michael Nielsen's online book for more information: http://neuralnetworksanddeeplearning.com/chap4.html

The goal of unsupervised feature extraction



Unlabeled images

Learning algorithm

Feature representation

Source: Andrew Ng; Mahendran and Veldadi 2015

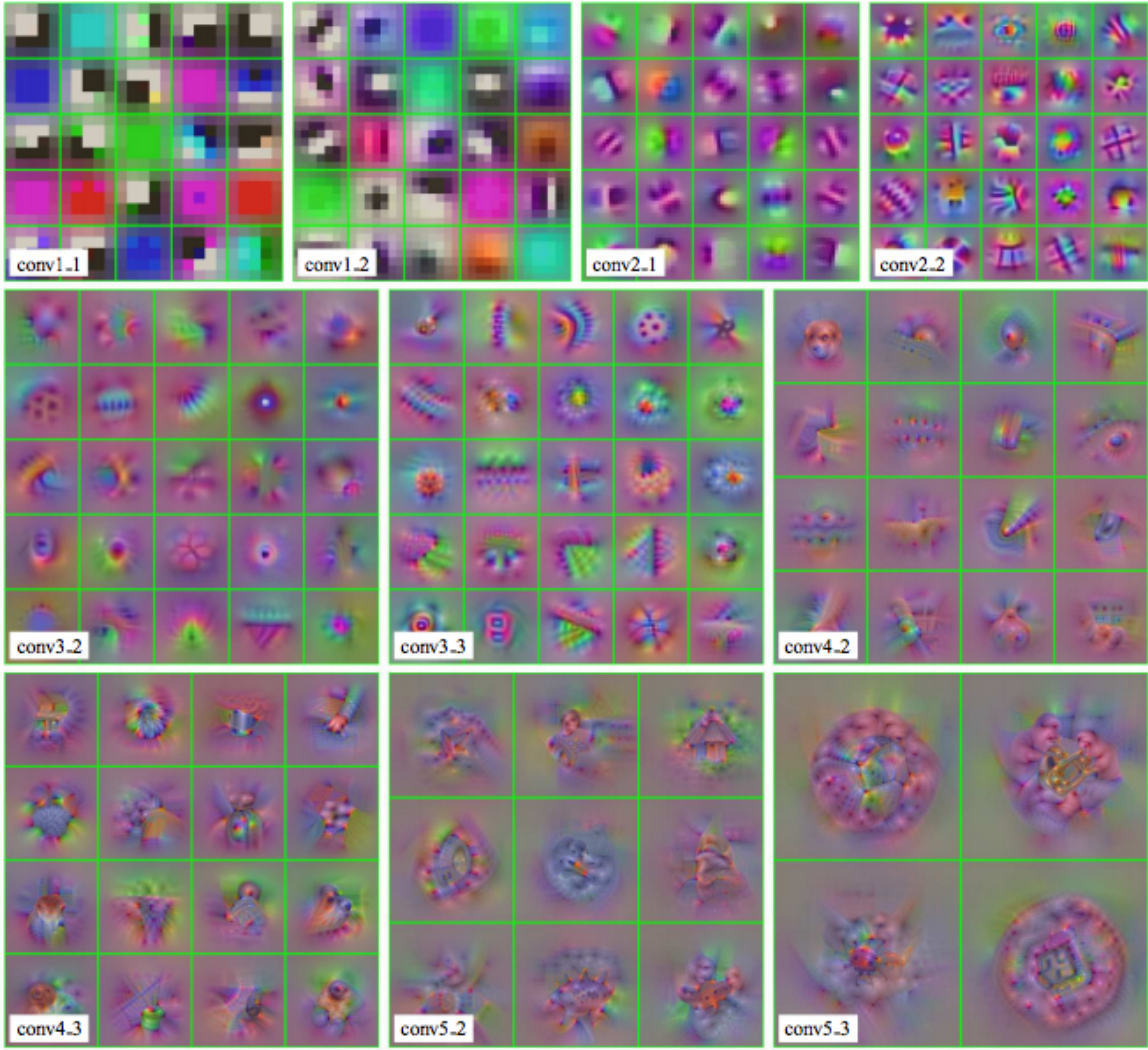
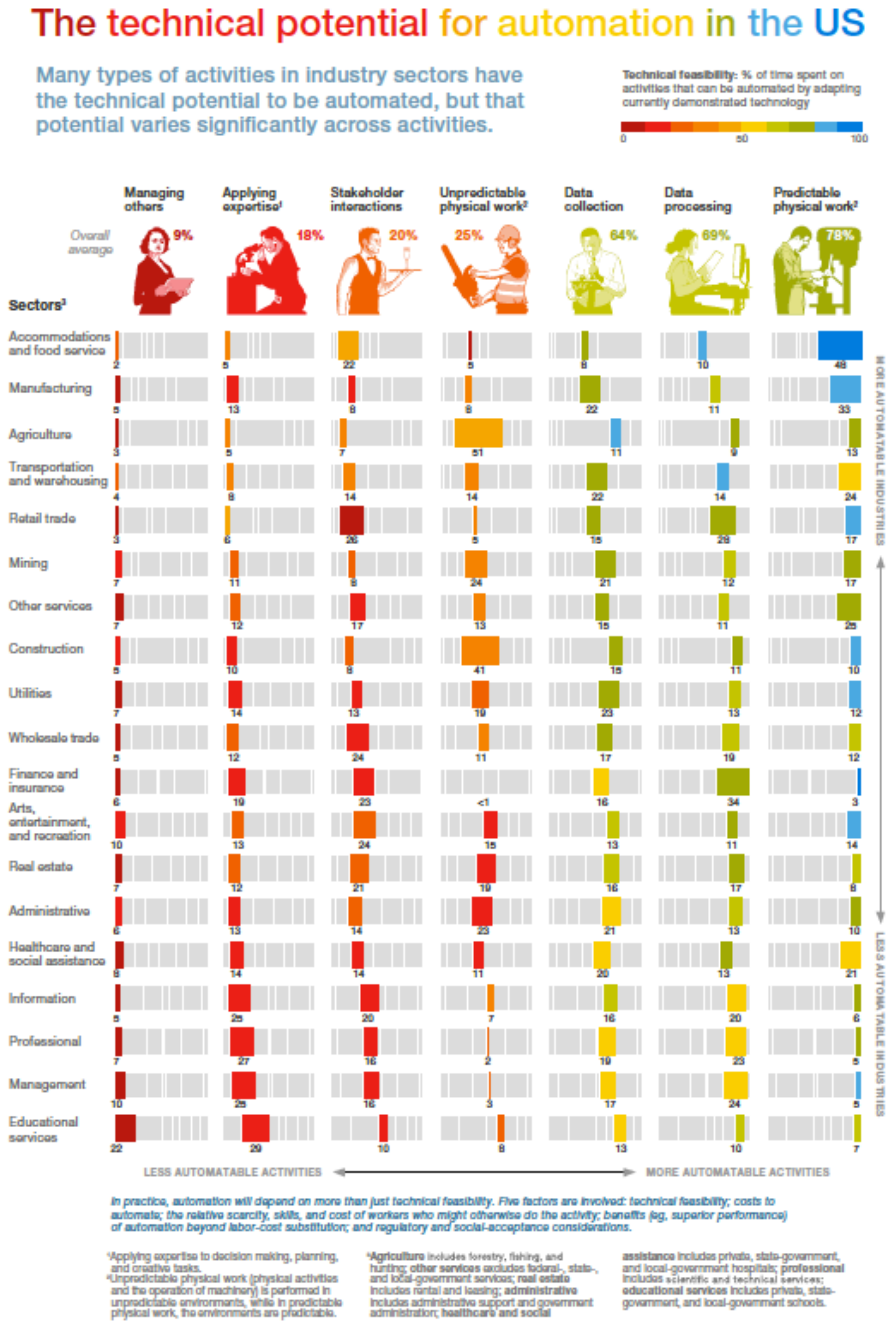


Fig. 17 Activation maximization of the first filters for each convolutional layer in VGG-VD-16 (Color figure online)

"People sometimes ask how quickly I think we will get there, and my honest answer is I don't know. We could get there in 3 years or in 30 years. **But I do believe that it will happen in this century.**"

Marek Rosa

CEO, GoodAI



Source: McKinsey Quarterly, July 2016

“I must study politics and war that my sons may have liberty to study mathematics and philosophy. My sons ought to study mathematics and philosophy, geography, natural history, naval architecture, navigation, commerce, and agriculture, in order to give their children a right to study painting, poetry, music, architecture, statuary, tapestry, and porcelain.”

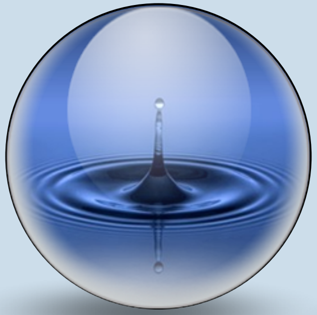
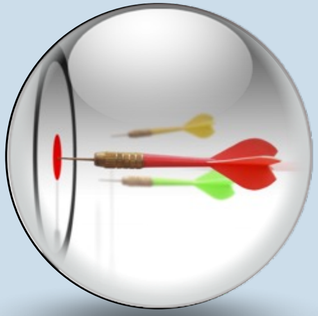
-- John Adams, 1780

"Despite these astonishing advances, we are a long way from machines that are as intelligent as humans—or even rats. So far, we've seen only 5% of what AI can do."

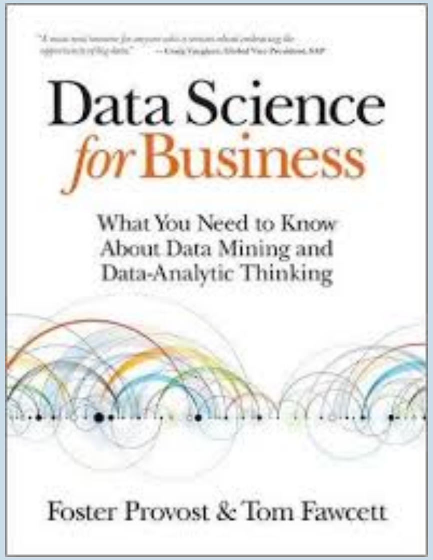
Yann LeCun

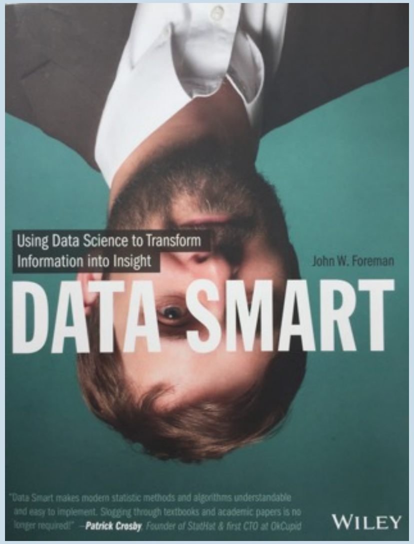
Director of research, Facebook

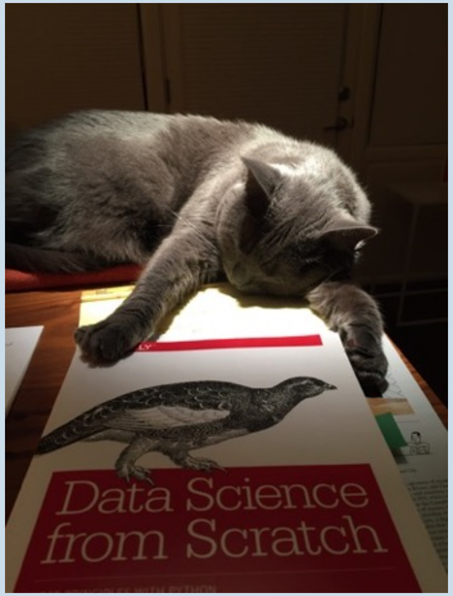
Source: www.cbinsights.com

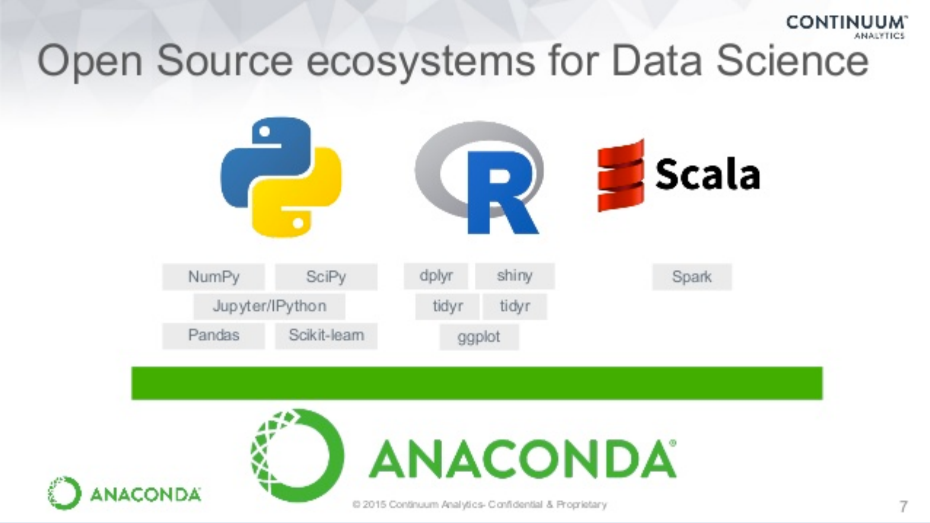



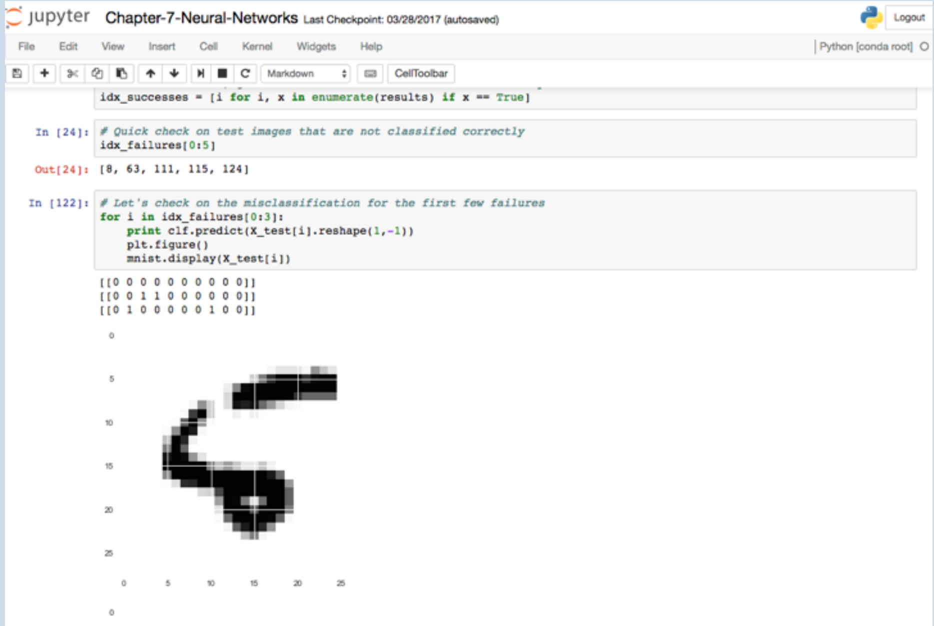
	Perception	Learning	Reasoning	Abstraction
Rules–Based Systems	No	No	Yes	No
Machine Learning	Yes	Yes	No	No
Deep Learning	Yes	Yes	No	Yes
Contextual Adaptation	Yes	Yes	Yes	Yes











Jenn Wortman Vaughn
Daphne Koller
Andrew Ng
Adam Geitgey
Jason Brownlee
Sonya Sawtelle
Michael Nielsen
Chris Olah
Andrej Karpathy
Hilary Mason

Datasets

- Kaggle
- UCI ML Repository
- Open Data for Deep Learning

"Organizations that make the most of machine learning are those that have in place an infrastructure that makes experimenting with many different learners, data sources and learning problems easy and efficient, and where there is a close collaboration between machine learning experts and application domain ones."

Pedro Domingos, *A Few Useful Things to Know About Machine Learning*.

"More broadly, companies must have two types of people to unleash the potential of machine learning. 'Quants' are schooled in its language and methods. 'Translators' can bridge the disciplines of data, machine learning, and decision making by reframing the quants' complex results as actionable insights that generalist managers can execute."

An executive's guide to machine learning,
McKinsey Quarterly, June 2015