



State of the Art in Machine Learning

Poul Petersen
BigML



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



UNIVERSITAT
DE VALÈNCIA

What is ML?

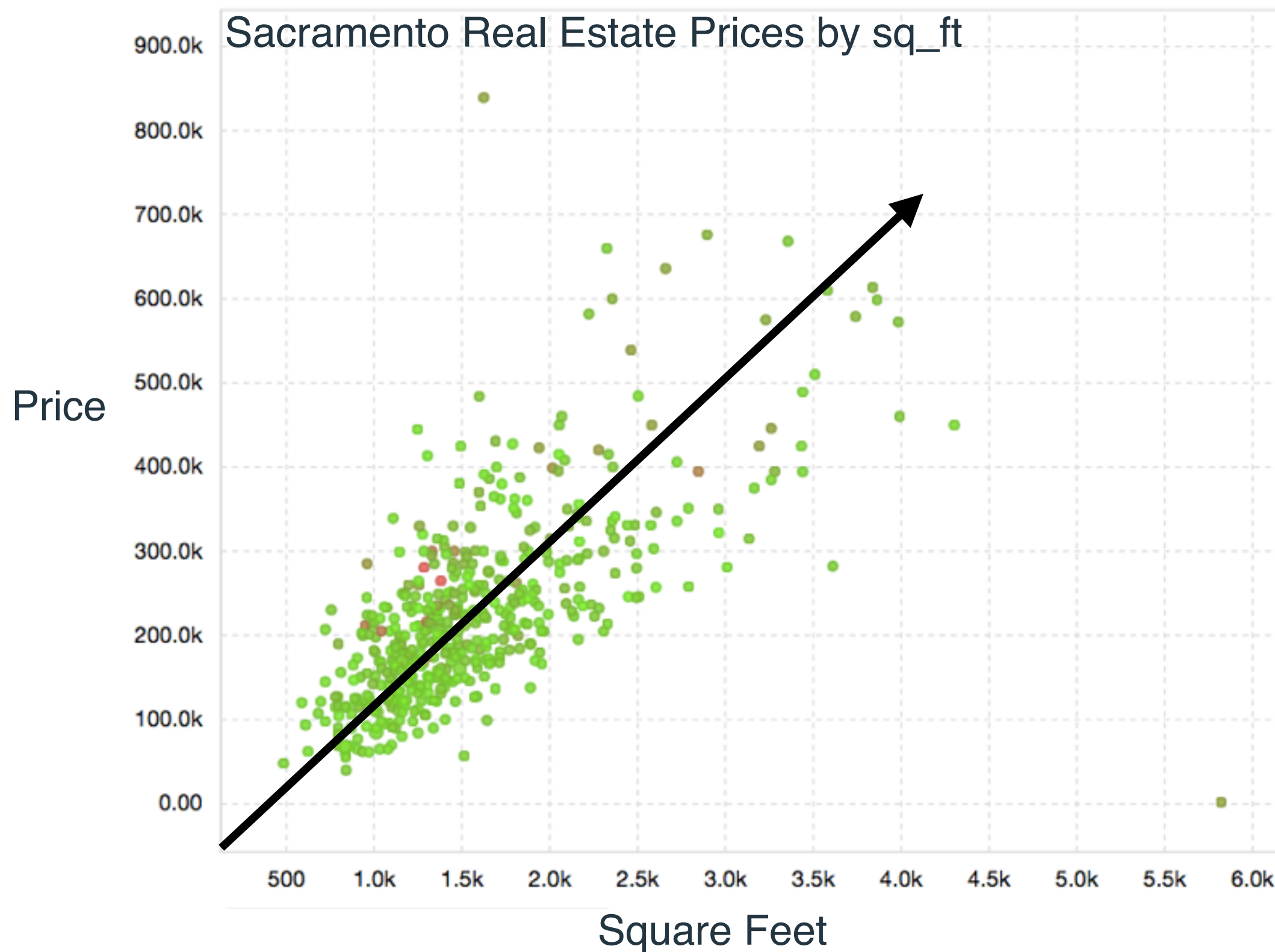


“a field of study that gives computers the ability to learn without being explicitly programmed”

Professor Arthur Samuel, 1959

- What “ability to learn” do computers have?
- What does “explicitly programmed” mean?

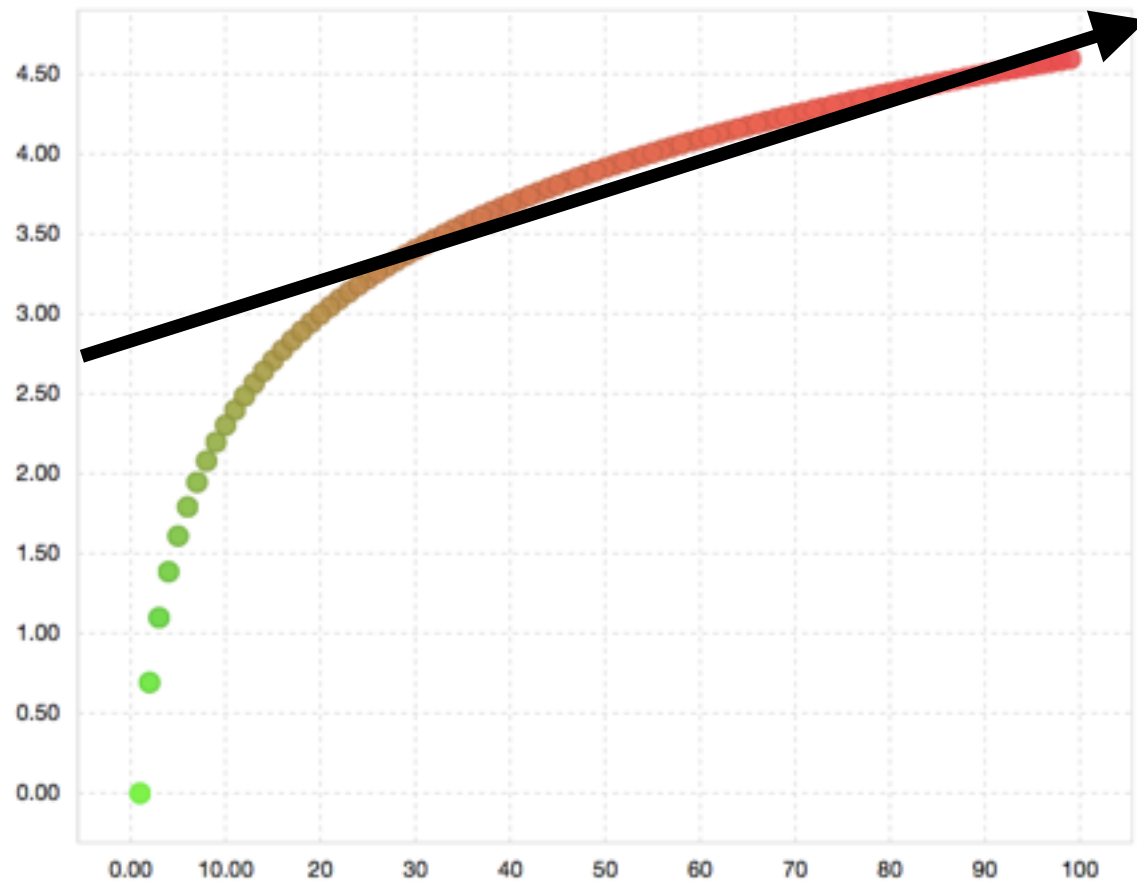
Ability to Learn



Ability to Learn

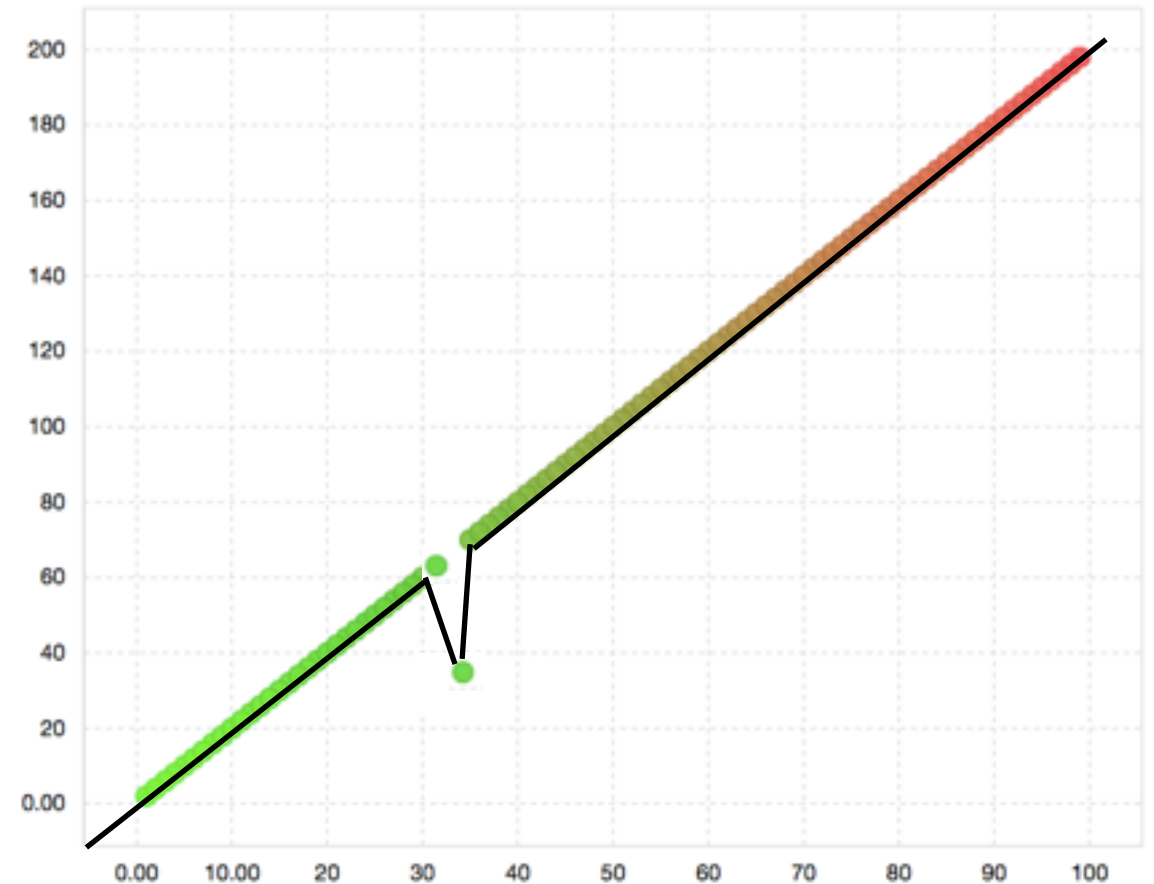
- Provide computer with **examples** of the relationship between square footage X and price Y .
- The computer **learns** the equation of a line $F()$ which fits the examples.
- The computer can now **predict** the price of any home knowing only the square footage: $F(x) = y$
- This model is known as *Linear Regression*. There are other types of models.
- You may have noticed there was some points that did not fit. This is important!

Learning Problems (fit)



Under-fitting

- Model does not fit well enough
- Does not capture the underlying trend of the data
- Change algorithm or features



Over-fitting

- Model fits too well does not “generalize”
- Captures the noise or outliers of the data
- Change algorithm or filter outliers

Learning Problems (missing)

- Missing values at training/prediction time
- Some algorithms can handle missing values, some no
- Missing data is sometimes important
- Replace missing values
- Predict missing values

Learning Problems (missing)

Missing@	Decision Trees	KNN	Logistic Regression	Naive Bayes	Neural Networks
Training	Yes	No	No	Yes	Yes*
Prediction	Yes	No	No	Yes	No

Not Explicitly Programmed



Control System



“Customers go on vacation... We need a program that predicts if the light will come on when the control system flips the switch.”

Not Explicitly Programmed

Switch Light?	
on	TRUE
off	FALSE



@iLoveRuby

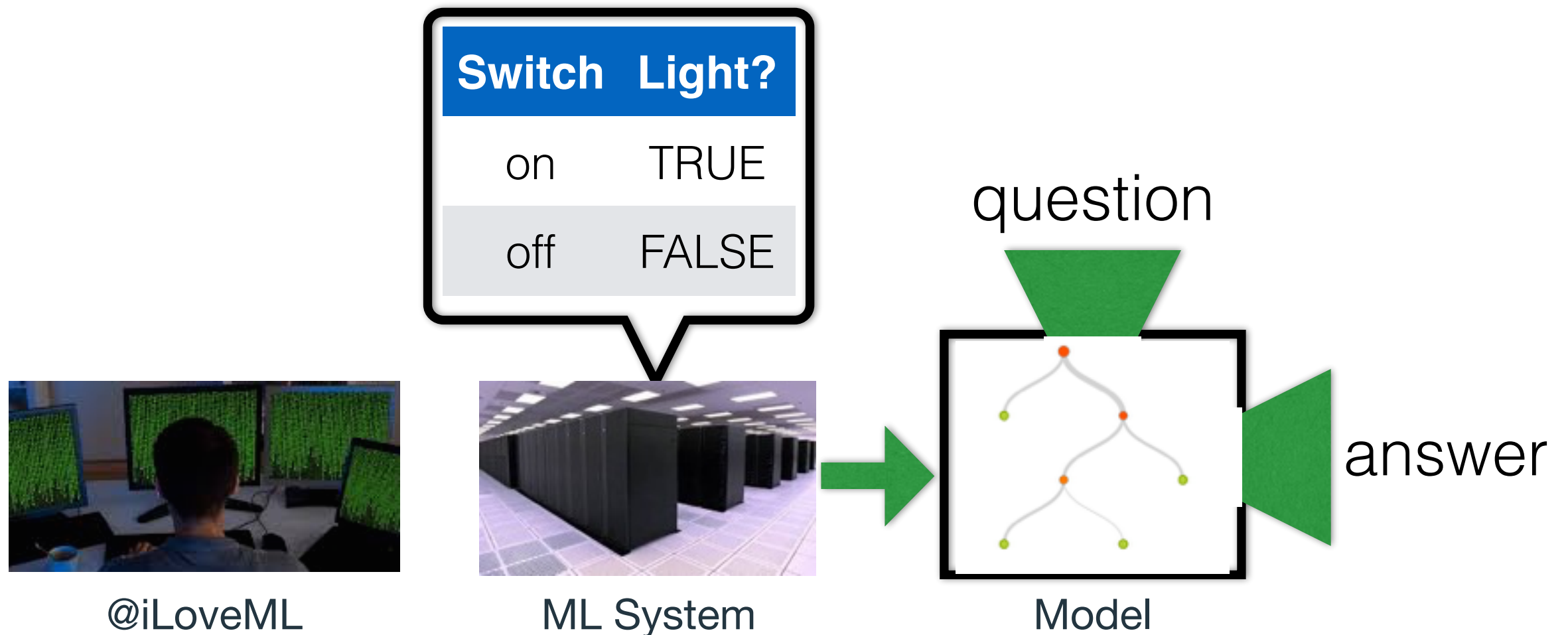
```
# ruby is da best
age => 27,
occupation => 'Rubyist',
hobby => 'Yak Shaving'

def isLightOn(switch)
  if switch == "on"
    puts true
  else
    puts false
  end
end

on => 'Sysadmin and Rubyist'
hobby => 'Promoting OS/2'
```

- @iLoveRuby reasoned the rules by experience
- Then programmed the rules **explicitly**

Not Explicitly Programmed



- The ML System reasoned the rules from the data and created a Model
- Functionally the Model is the same as @iLoveRuby's explicit model

Not Explicitly Programmed



Goal is to *predict* if the bulb will come on but the switch is not the important *variable*:

- how long has the bulb has been in service
- reliability of brand
- rated power: higher power shorter life
- duty cycle
- room conditions: temperature, humidity

Even worse: **None** of these conditions is absolute.

Not Explicitly Programmed

brand	power	age	duty	temp	humidity	FAIL?
koala	45	338	1	16	0.03	FALSE
otter	15	140	1	27	0.27	FALSE
koala	15	315	1	19	0.37	TRUE
otter	45	338	1	29	0.27	TRUE
koala	45	211	1	23	0.85	TRUE
otter	15	328	1	17	0.56	FALSE
koala	15	318	2	22	0.45	TRUE
koala	15	273	1	27	0.18	FALSE
koala	45	102	1	21	0.48	FALSE
koala	15	110	2	15	0.99	TRUE
otter	45	355	2	15	0.01	FALSE
otter	15	69	1	24	0.70	FALSE
koala	15	69	1	24	0.70	FALSE
koala	15	337	2	27	0.83	TRUE

Not Explicitly Programmed

brand	power	age	duty	temp	humidity	FAIL?
koala	45	338	1	16	0.03	FALSE
otter	15	140	1	27	0.27	FALSE
koala	15	315	1	19	0.37	TRUE
otter	45	338	1	29	0.27	TRUE
koala	45	211	1	23	0.85	TRUE
otter	15	328	1	17	0.56	FALSE
koala	15	318	2	22	0.45	TRUE
koala	15	273	1	27	0.18	FALSE
koala	45	102	1	21	0.48	FALSE
koala	15	110	2	15	0.99	TRUE
otter	45	355	2	15	0.01	FALSE
otter	15	69	1	24	0.70	FALSE
koala	15	69	1	24	0.70	FALSE
koala	15	337	2	27	0.83	TRUE



@iLoveRuby



JOHN DOE
Full Address • City, State, ZIP • Phone Number • E-mail

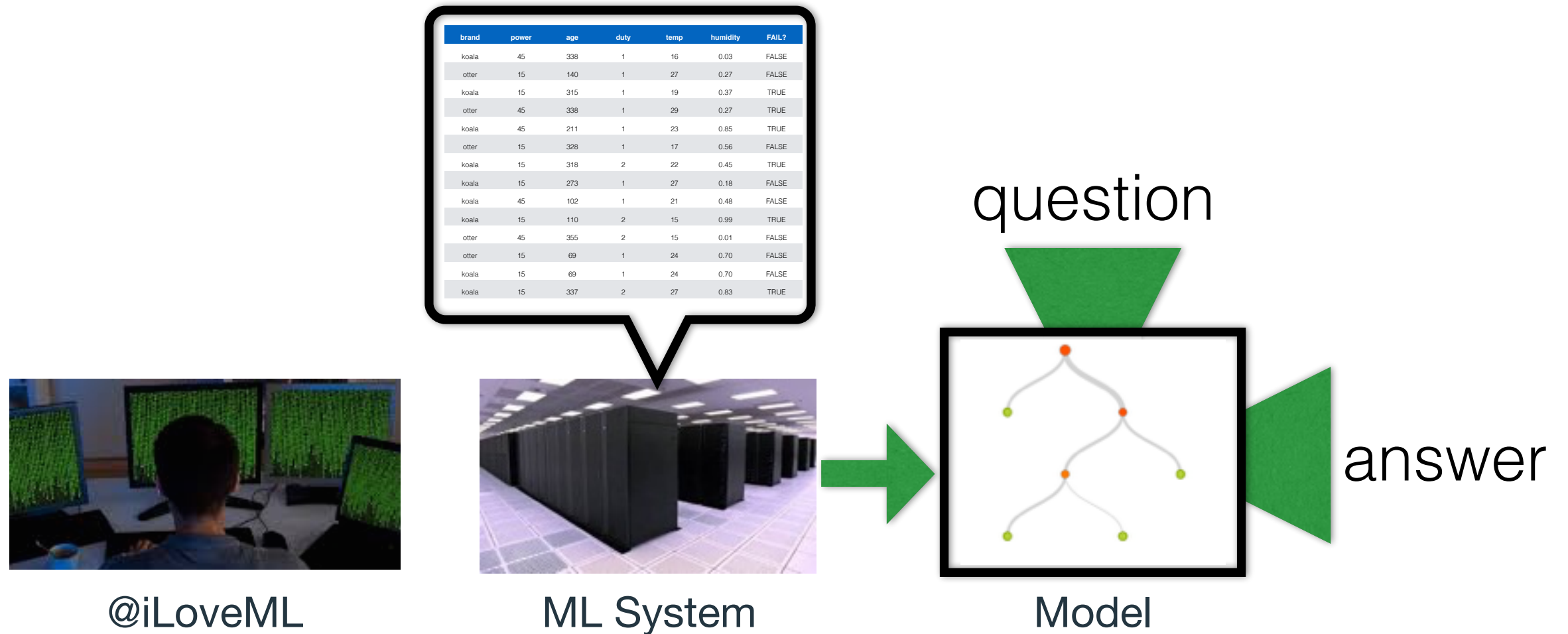
OBJECTIVE: Design apparel plans for an innovative retail company

EDUCATION:
UNIVERSITY OF MINNESOTA
College of Design
• Bachelor of Science in Graphic Design
• Cumulative GPA 3.87, Dean's List
• Twinkleton Iron Range Scholarship
City, State
May 2011

WORK EXPERIENCE:
AMERICAN EAGLE
Sales Associate
• Collaborated with the store merchandiser creating displays to attract clientele
• Use my visual awareness to assist customers in their shopping experience
• Thoroughly scan every piece of merchandise for inventory control
• Process shipment to increase my product knowledge
City, State
July 2009 - present
PLANET BEACH
Spa Consultant
• Sell retail and memberships to meet company sales goals
• Build organizational skills by single-handedly running all operating procedures
• Communicate with clients to fulfill their wants and needs
• Attend promotional events to market our services
• Handle cash and deposits during opening and closing
• Received employee of the month award twice
City, State
Aug. 2008 - present
HEARTBREAKER
Sales Associate
• Stocked sales floor with fast fashion inventory
• Marked down items allowing me to see successful merchandise in a retail market
• Offered advice and assistance to each guest
City, State
May 2008 - Aug. 2008
VICTORIA'S SECRET
Fashion Representative
• Applied my leadership skills by assisting in the training of coworkers
• Set up mannequins and displays in order to entice future customers
• Provided superior customer service by helping with consumer decisions
• Took seasonal inventory
City, State
Jan. 2006 - Feb. 2009
VOLUNTEER EXPERIENCE:
TARGET CORPORATION
Brand Ambassador
• Represented Periscope Marketing and Target Inc. at a college event
• Engaged University of Minnesota students in the Target brand experience
City, State
August 2009

- Multi-dimensional data is much harder to find rules
- Explicit program requires modification

Not Explicitly Programmed



- ML System easily re-trains on new data

Terminology

datasource

brand	power	age	duty	temp	humidity	FAIL?
koala	45	338	1	16	0.03	FALSE
otter	15	140	1	27	0.27	FALSE
koala	15	315	1	19	0.37	TRUE
otter	45	338	1	29	0.27	TRUE
koala	45	211	1	23	0.85	TRUE
otter	15	328	1	17	0.56	FALSE
koala	15	318	2	22	0.45	TRUE
koala	15	273	1	27	0.18	FALSE
koala	45	102	1	21	0.48	FALSE
koala	15	110	2	15	0.99	TRUE
otter	45	355	2	15	0.01	FALSE
otter	15	69	1	24	0.70	FALSE
koala	15	69	1	24	0.70	FALSE
koala	15	337	2	27	0.83	TRUE

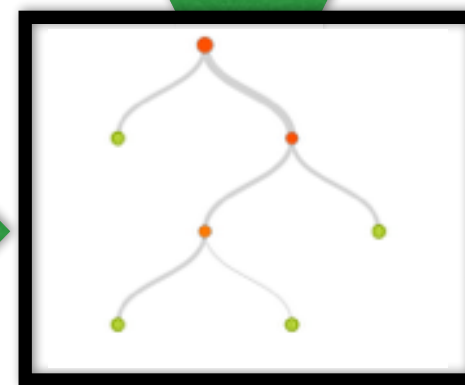
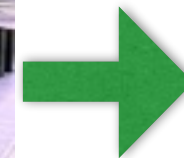
→ **training**



@iLoveML



ML System



input

Model

prediction

↓
“putting into production”

Supervised Learning

Labeled Data

features →

instances ↓

brand	power	age	duty	temp	humidity	FAIL?
koala	45	338	1	16	0.03	FALSE
otter	15	140	1	27	0.27	FALSE
koala	15	315	1	19	0.37	TRUE
otter	45	338	1	29	0.27	TRUE
koala	45	211	1	23	0.85	TRUE
otter	15	328	1	17	0.56	FALSE
koala	15	318	2	22	0.45	TRUE

Supervised Learning

Classification

animal	state	...	proximity	label
tiger	hungry	...	close	run
elephant	happy	...	far	take picture

Regression

animal	state	...	proximity	min_kmh
tiger	hungry	...	close	70
hippo	angry	...	far	10

Multi-Label Classification

animal	state	...	proximity	action1	action2
tiger	hungry	...	close	run	look untasty
elephant	happy	...	far	take picture	call friends

Unsupervised Learning

Unlabeled Data

features →

instances ↓

date	customer	account	auth	class	zip	amount
Mon	Bob	3421	pin	clothes	46140	135
Tue	Bob	3421	sign	food	46140	401
Tue	Alice	2456	pin	food	12222	234
Wed	Sally	6788	pin	gas	26339	94
Wed	Bob	3421	pin	tech	21350	2459
Wed	Bob	3421	pin	gas	46140	83
The	Sally	6788	sign	food	26339	51

Unsupervised Learning

Clustering

date	customer	account	auth	class	zip	amount
Mon	Bob	3421	pin	clothes	46140	135
Tue	Bob	3421	sign	food	46140	401
Tue	Alice	2456	pin	food	12222	234
Wed	Sally	6788	pin	gas	26339	94
Wed	Bob	3421	pin	tech	21350	2459
Wed	Bob	3421	pin	gas	46140	83
The	Sally	6788	sign	food	26339	51

similar

Anomaly Detection

date	customer	account	auth	class	zip	amount
Mon	Bob	3421	pin	clothes	46140	135
Tue	Bob	3421	sign	food	46140	401
Tue	Alice	2456	pin	food	12222	234
Wed	Sally	6788	pin	gas	26339	94
Wed	Bob	3421	pin	tech	21350	2459
Wed	Bob	3421	pin	gas	46140	83
The	Sally	6788	sign	food	26339	51

unusual

Semi-Supervised Learning

Labeled and Unlabeled Data

						label
brand	power	age	duty	temp	humidity	FAIL?
koala	45	338	1	16	0.03	FALSE
otter	15	140	1	27	0.27	
koala	15	315	1	19	0.37	
otter	45	338	1	29	0.27	TRUE
koala	45	211	1	23	0.85	TRUE
otter	15	328	1	17	0.56	
koala	15	318	2	22	0.45	TRUE

infer

Data Types

1 2 3

1, 2.0, 3, -5.4

numeric

A B C

true, yes, red, mammal

categorical

DATE-TIME

2013-09-25 10:02

DATE-TIME

text

Be not afraid of greatness:
some are born great, some
achieve greatness, and
some have greatness
thrust upon 'em.

text

YYYY-MM-DD

YEAR

2013

YYYY-MM-DD

MONTH

September

YYYY-MM-DD

DAY-OF-MONTH

25

M-T-W-T-F-S-D

DAY-OF-WEEK

Wednesday

HH:MM:SS

HOUR

10

HH:MM:SS

MINUTE

02

great
being afraid some

“great”

appears 2 times

“afraid”

appears 1 time

“born”

appears 1 time

“some”

appears 2 times

Text Analysis

Be not afraid of greatness:
some are born great, some
achieve greatness, and
some have greatness
thrust upon 'em.

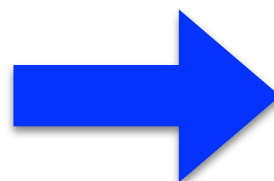
Text Analysis

Be not afraid of greatness:
some are born great, some
achieve greatness, and
some have greatness
thrust upon 'em.

great: appears 4 times

Text Analysis

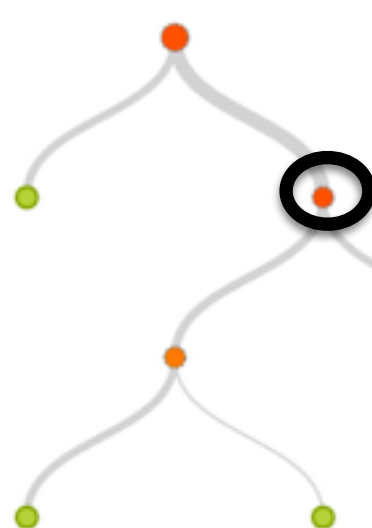
Be not afraid of greatness:
some are born great, some achieve
greatness, and some have greatness
thrust upon 'em.



great	afraid	born	achieve
-------	--------	------	---------

4	1	1	1
---	---	---	---

...
-----	-----	-----	-----



The token “great”
does not occur

The token “afraid”
occurs more than once

Model

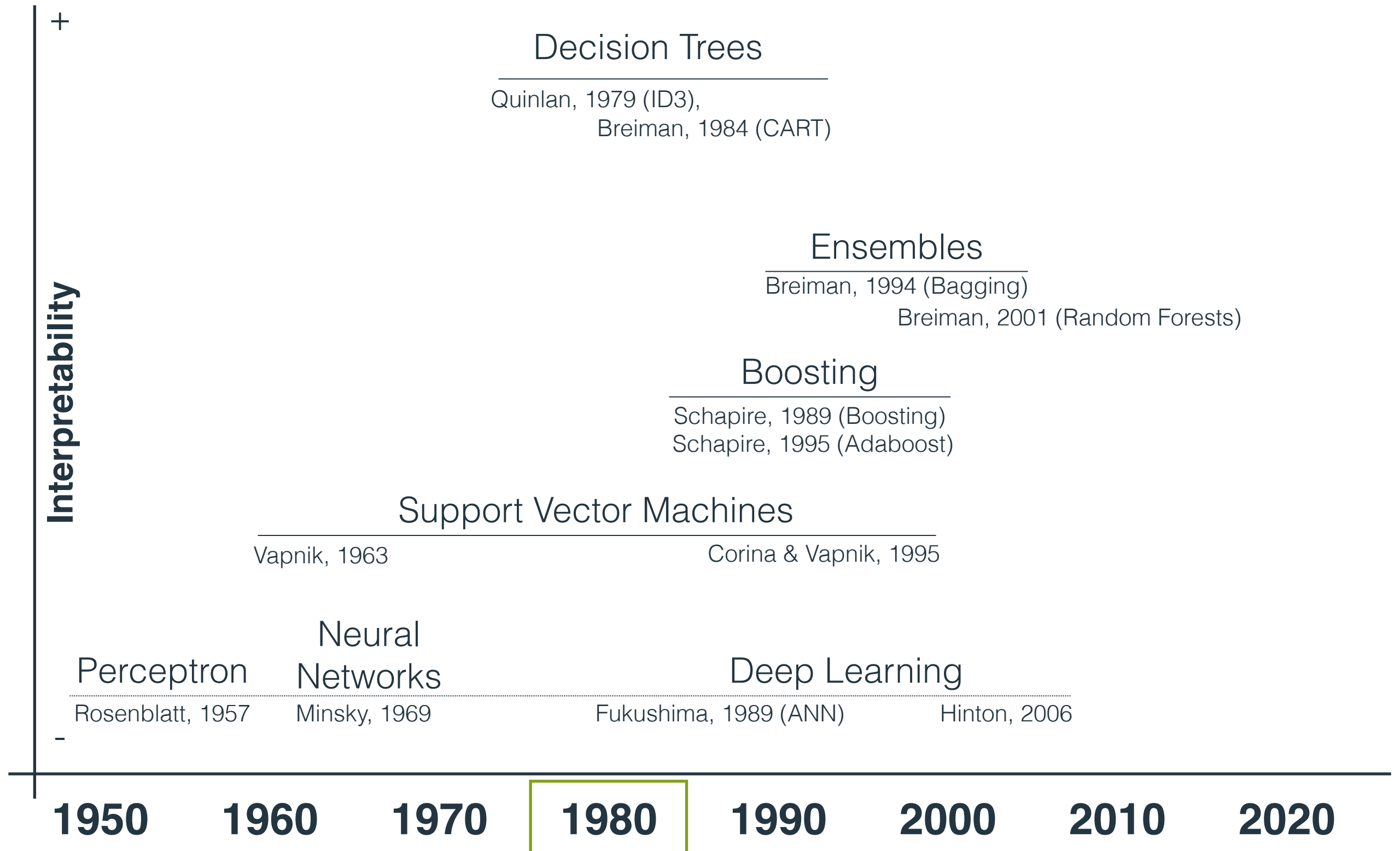
Topic Modeling

Four score and seven years ago **our fathers** brought forth on this continent a **new nation, conceived in liberty, and dedicated to the proposition that all men are created equal.**

Now **we are engaged in a great civil war**, testing whether that **nation**, or any **nation so conceived and so dedicated**, can long endure. **We are met on a great battlefield of that war. We have come to dedicate a portion of that field, as a final resting place for those who here gave their lives that that nation might live. It is altogether fitting and proper that we should do this.**

But, in a larger sense, **we can not dedicate, we can not consecrate, we can not hallow this ground. The brave men, living and dead, who struggled here, have consecrated it, far above our poor power to add or detract. The world will little note, nor long remember what we say here, but it can never forget what they did here. It is for us the living, rather, to be dedicated here to the unfinished work which they who fought here have thus far so nobly advanced. It is rather for us to be here dedicated to the great task remaining before us—that from these honored dead we take increased devotion to that cause for which they gave the last full measure of devotion—that we here highly resolve that these dead shall not have died in vain—that this nation, under God, shall have a new birth of freedom—and that government of the people, by the people, for the people, shall not perish from the earth.**

Brief History of ML



Why ML Now?

- Decreasing cost of data
- Abundant computing power, especially cloud
- Machine Learning APIs
- Abundance of APIs + internet to combine easily

Composability

Enhancing your Cloud Applications with Artificial Intelligence

Gluecon 2014

Putting it all together: Sample A.I. App Workflow

Example app: "Voice + visual assistant for french hikers"

Use AT&T Speech API to accept speech input

"Est-ce que toxique?"



Use Google Translate to convert from French to English

"Is this poisonous?"



Take a picture of the item the question is about



Use AlchemyAPI to interpret the photograph

"fly agaric"



Use IBM Watson to answer the question

Is this poisonous + fly agaric



The Stages of a ML App

