

Introduction to Machine Learning

Spring 2018

Today's Topics

- Class logistics
- Why machine learning?
- What does a machine learn?
- Designing a “supervised” machine learning algorithm
- Lab

Today's Topics

- Class logistics
- Why machine learning?
- What does a machine learn?
- Designing a “supervised” machine learning algorithm
- Lab

Introductions

Instructor: Danna Gurari

Danna: pronounced like “Donna”

Gurari: rhymes with Ferrari



Interdisciplinary class: introduce yourself & share about your

- career aspirations
- experience with machine learning and programming

Class Overview and Logistics

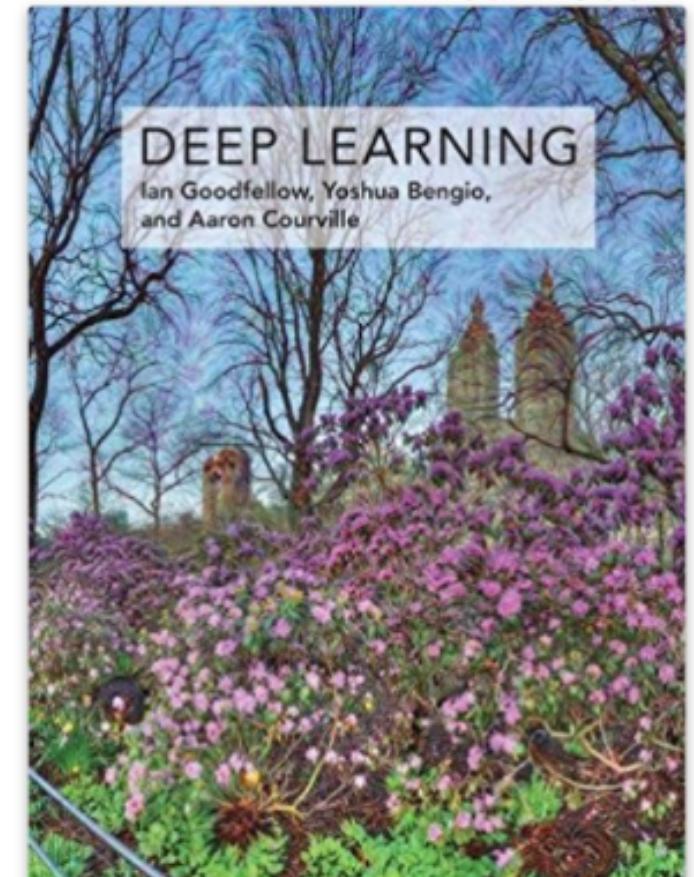
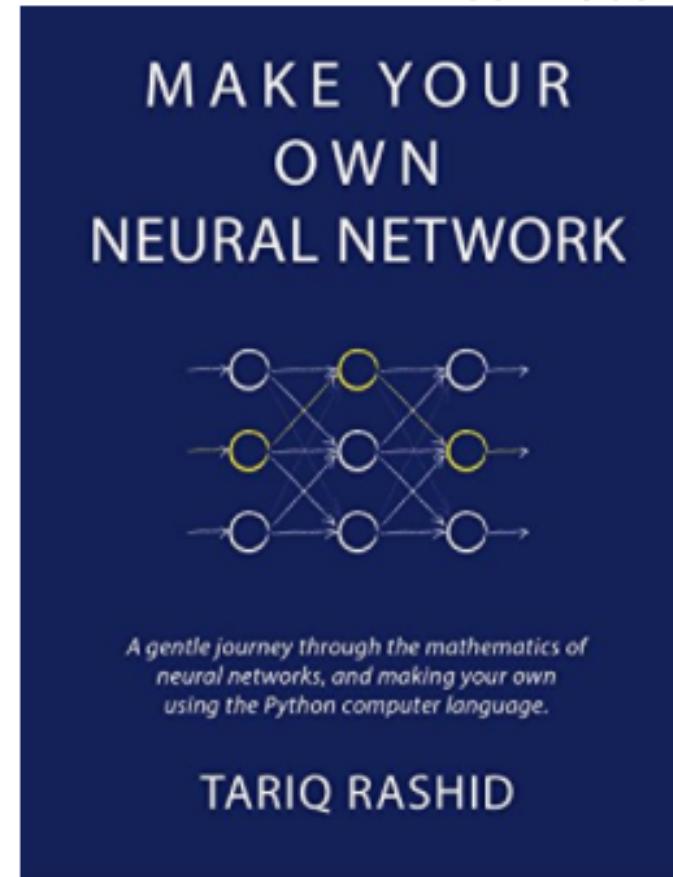
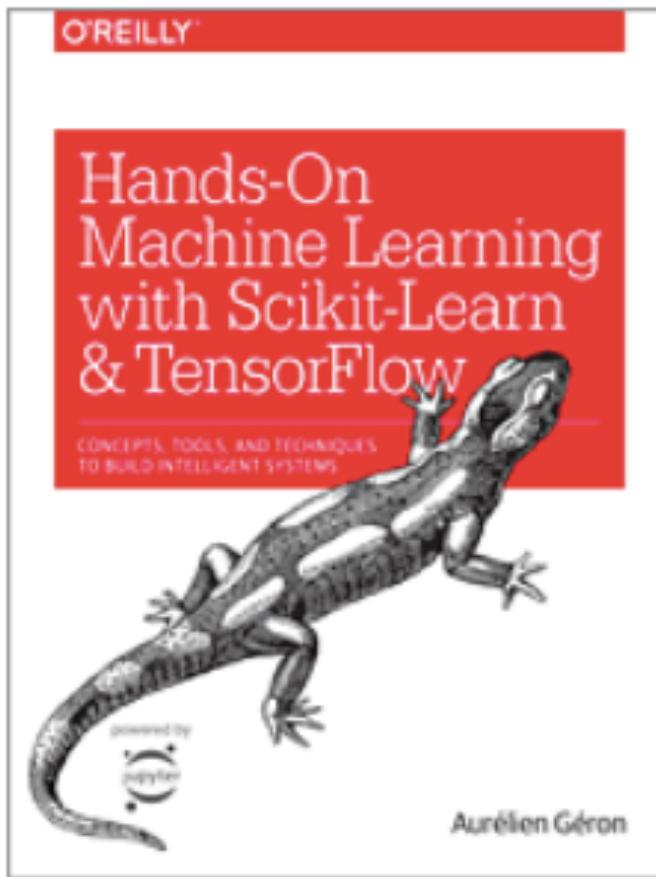
- Typical class format:
 - First half = lecture + discussion
 - Break = snacks provided
 - Second half = hands-on lab session
- Class website
 - Assignments
 - Schedule
 - Final Project
- Syllabus

Q&A: “Do I have the appropriate pre-requisites/background?”

- Yes. There are no pre-requisites.
- You will be expected to further develop skills we cover in class on your own; e.g.,
 - Programming; e.g., Python
 - Linear algebra; e.g., vector/matrix manipulations
 - Calculus; e.g. partial derivatives
 - Probability; Bayes rule
 - Statistics; mean/median/maximum likelihood

Q&A: ““What are required textbooks?”

- None. The following are strongly recommended:



Congratulations!

- By taking this class, you receive a gift of:



- Thanks to: Microsoft Azure

Today's Topics

- Class logistics
- Why machine learning?
- What does a machine learn?
- Designing a “supervised” machine learning algorithm
- Lab

Origin of Machine Learning: Human “Computer”

- 1613-Present: https://en.wikipedia.org/wiki/Human_computer
 - i.e., a person performs calculations towards solving complex problems
 - e.g., astronomy, fluid dynamics, nuclear fission
 - e.g.,



• •



Dorothy Vaughn



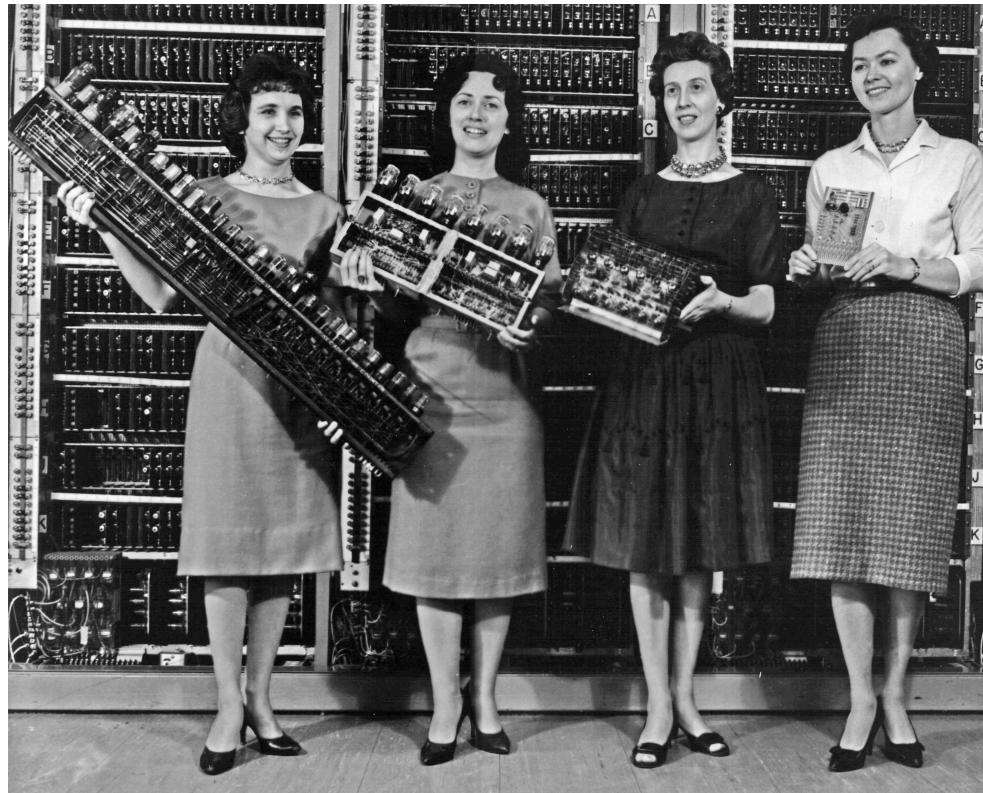
Mary Jackson



Miriam Mann

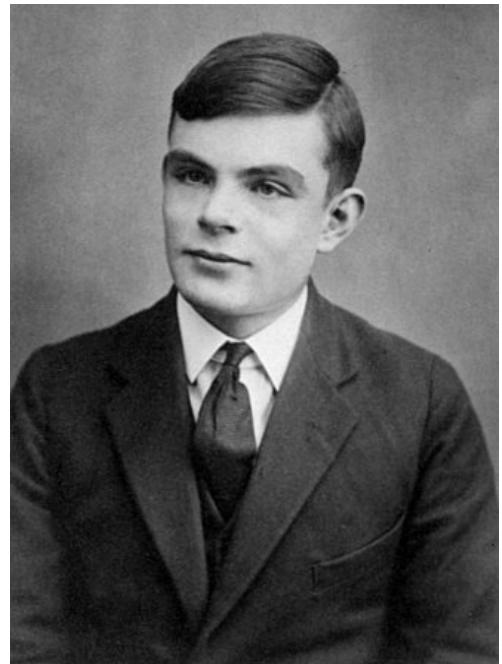
Origin of Machine Learning

- Timeline
 - Early 1940s: first computing machines created; e.g., ENIAC in World War II



Origin of Machine Learning

- Timeline
 - Early 1940s: first computing machines created; e.g., ENIAC in World War II
 - 1950: Alan Turing devises a test to decide if a machine's behavior is indistinguishable from that of a human



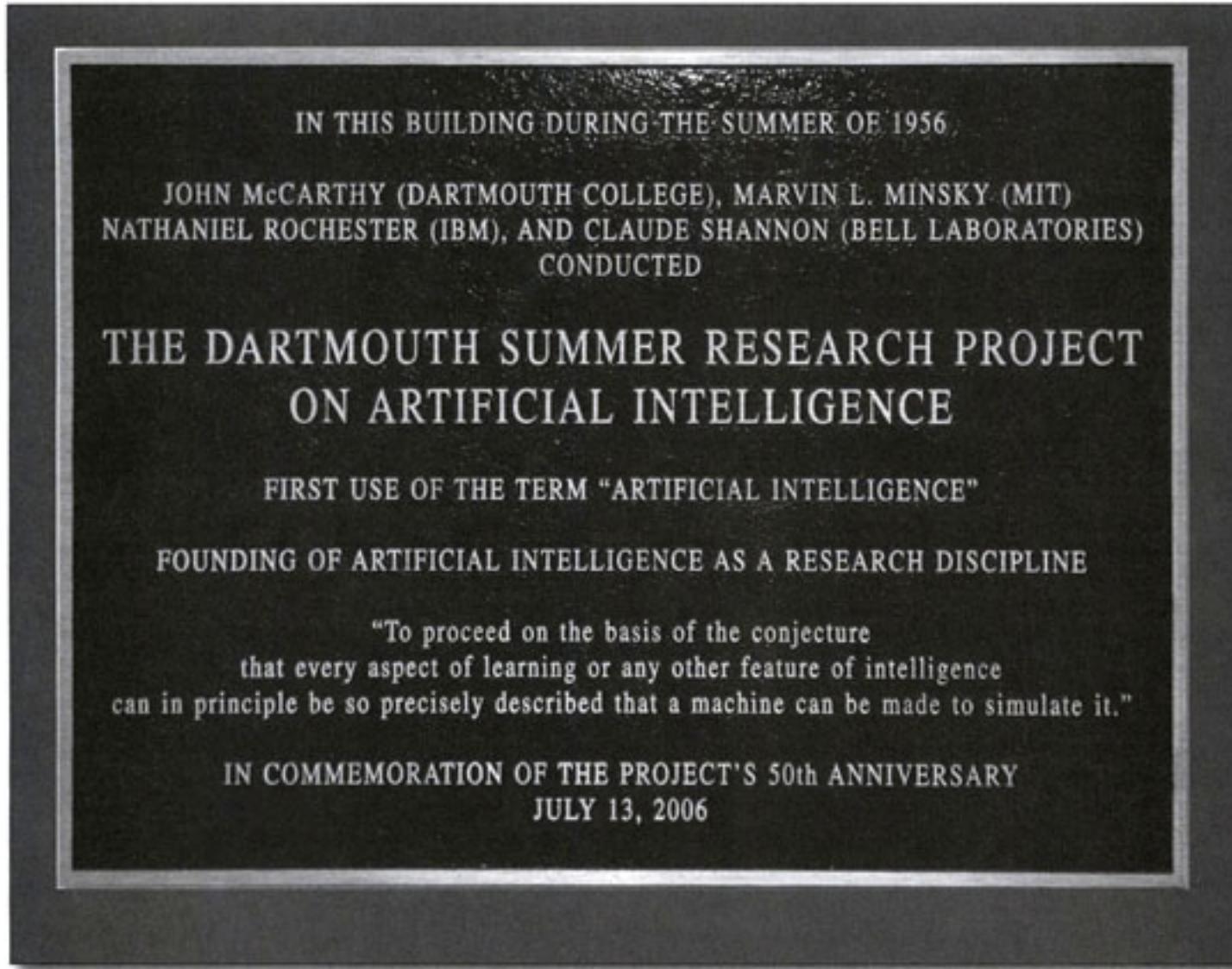
Turing Test:



Origin of Machine Learning

- Timeline
 - Early 1940s
 - 1950: Alan Turing's paper on indistinguishability of machines from humans
 - 1956: birth of the field at the Dartmouth Conference

Proposal Executive Summary
carried out during the summer of 1956.
The study is to determine whether or not other features can be made to simulate learning or any other feature of intelligence. The machine can be made to simulate some form of abstract thought and improve themself by solving these problems.



d War II
s
}

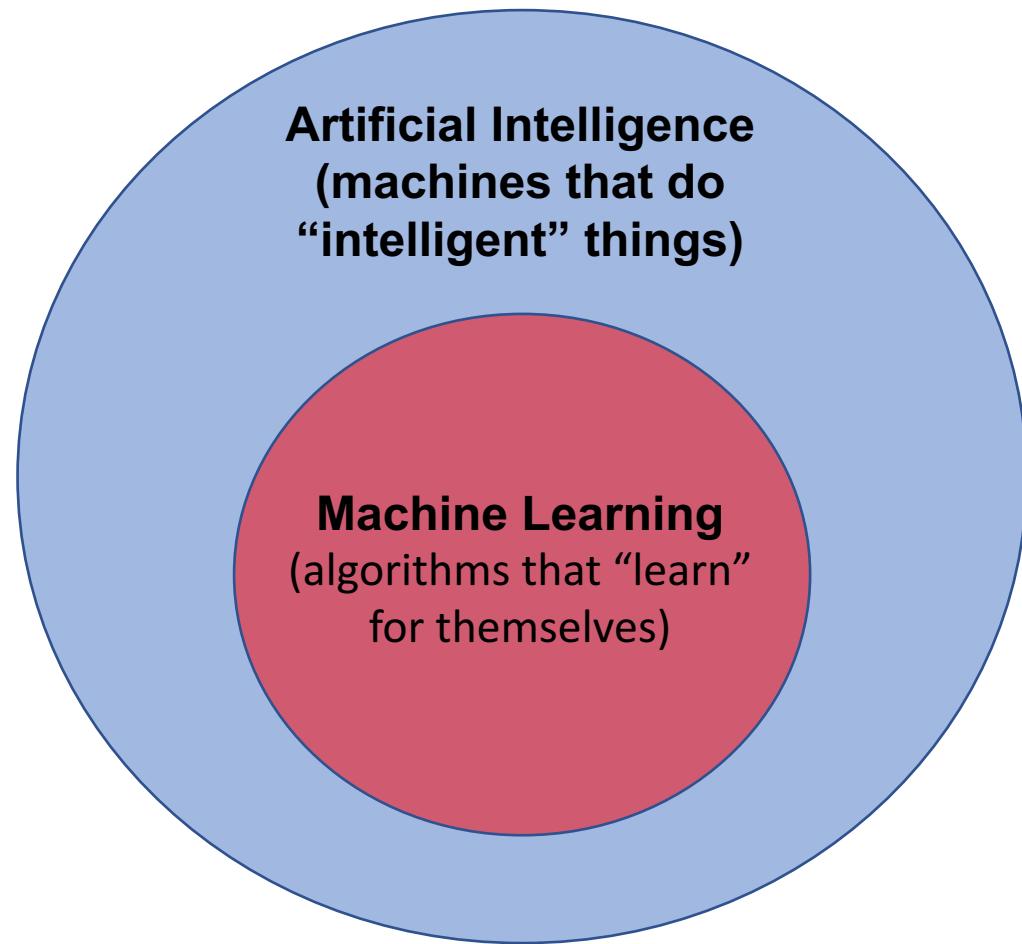
elligence be
Hampshire.
learning or any
machine can be
language,
nans, and
or more of
r a summer."

Origin of Machine Learning

- Timeline
 - Early 1940s: first computing machines created; e.g., ENIAC in World War II
 - 1950: Alan Turing devises a test to decide if a machine's behavior is indistinguishable from that of a human
 - 1956: birth of “artificial intelligence” (AI) as a field at a workshop called “Dartmouth Summer Research Project on Artificial Intelligence”
 - 1959: AI researcher coins the term “machine learning”

“Field of study that gives computers **the ability to learn without being explicitly programmed.**”

What is Machine Learning?

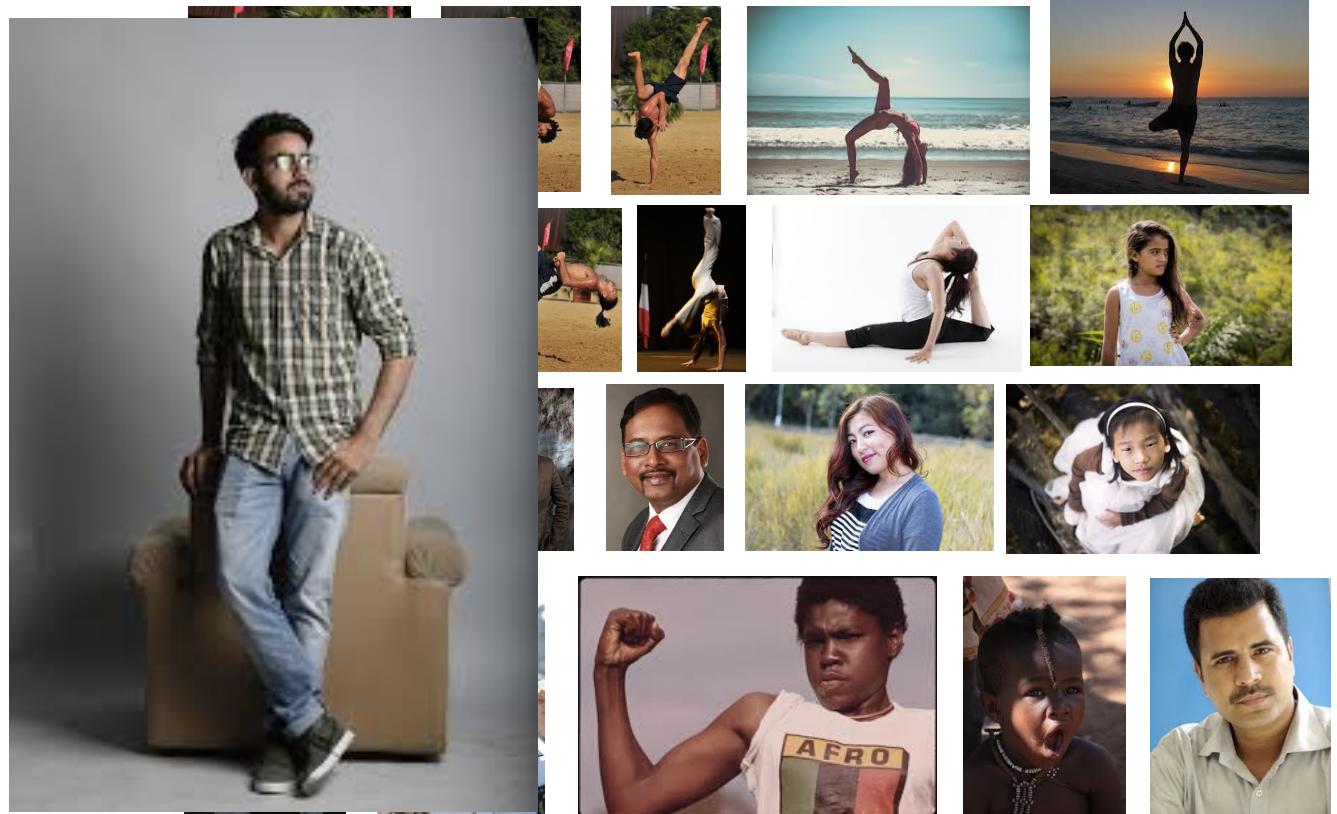


Why Have Machines “Learn”?

- It is hard to hand-craft a complete set of rules

Why Have Machines “Learn”?

e.g., What rules would you use to answer:
“Is there a person in the image?”

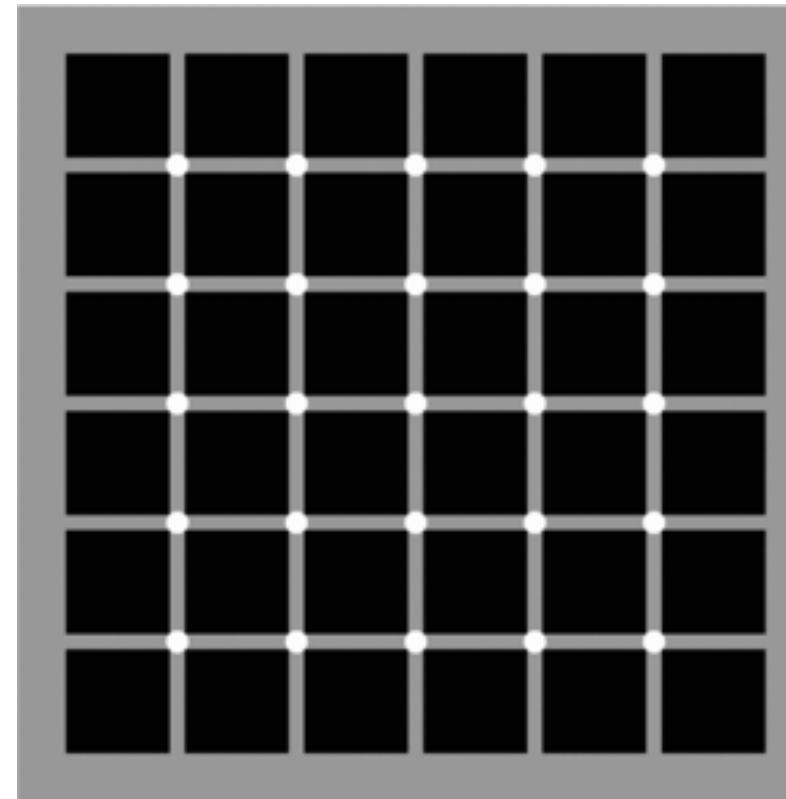


Why Have Machines “Learn”?

- It is hard to hand-craft a complete set of rules
- We, as humans, may not devise the best rules for a machine since our brains (unconsciously) pre-process the data we sense

Why Have Machines “Learn”?

e.g., how many black dots do you see in the image?



Origin of Machine Learning

- Recall Timeline
 - Early 1940s: First computing machines developed
 - 1950: Alan Turing devises a test to decide if a machine's behavior is indistinguishable from that of a human
 - 1956: birth of Artificial Intelligence. At a conference workshop called "Dartmouth Summer Research Project on Artificial Intelligence"
 - 1959: AI researcher coins the term “machine learning”

“Field of study that gives computers **the ability to learn without being explicitly programmed.**”

AI Problems Solved by Machine Learning Today

Recommendation Systems

The screenshot shows the Netflix homepage with a sidebar for "Movies". It features a "BUYERS' GUIDE" section with a photo of a couple. Below it, under "Critically-acclaimed", is another photo of a couple. The sidebar includes a "Books" section with categories like "Computers & Technology", "AI & Machine Learning", and "Data Processing". A "Departments" dropdown menu is open, showing "Books" and "Advanced Search". The main content area displays a "Recommen" banner.

The screenshot shows the "Recommendations" section of The New York Times website. It features a grid of book covers from Packt Publishing, including titles like "Python Machine Learning", "Computer Vision with Python 3", "Python Machine Learning By Example", "Deep Learning with TensorFlow", and "Python Machine Learning Projects". The Packt logo is prominently displayed. Below the grid, a green banner reads: "Get personalized recommendations, based on what you enjoy reading on NYTimes.com". At the bottom, there are "Log In", "Register Now", and "Log In With Facebook" buttons.

AI Problems Solved by Machine Learning Today

Information Retrieval

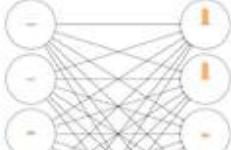
Baidu News machine learning 百度一下

网页 新闻 贴吧 知道 音乐 图片 视频 地图 文库 更多»

找到相关新闻92篇 新闻全文 新闻标题 | 按焦点排序 ▾

[...GWAS summary statistics for data mining and machine learning](#)
中国矿业大学 2017年12月26日 16:58
报告题目:Using GWAS summary statistics for data mining and machine learning 时间:12月29日上午9:00 地点:文昌校区逸夫楼邵206 主办单位:中国矿业大学信息与 ... 百度快照

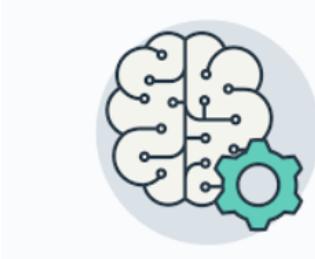
[\[Machine Learning\] 深度学习中消失的梯度](#)
深圳热线 2017年12月07日 09:46
原标题:[Machine Learning] 深度学习中消失的梯度 好久没有更新blog了,最近抽时间看了Nielsen的《Neural Net 百度快照



Google machine learning

All News Videos Images Books More

artificial intelligence big data iot distributed robotic cyber s



Machine learning & artificial intelligence
ARTIFICIAL INTELLIGENCE
Design an intelligent agent that perceives its environment and makes decisions to maximize chances of achieving its goal. Subfields: vision, robotics, machine learning, natural language processing, planning, ...
MACHINE LEARNING
Gloss "computes the ability to learn without being explicitly programmed" (Arthur Samuel, 1959)
SUPERVISED LEARNING Classification, regression
UNSUPERVISED LEARNING Clustering, dimensionality reduction, recommendation
REINFORCEMENT LEARNING Reward maximization
Machine Learning for Nature

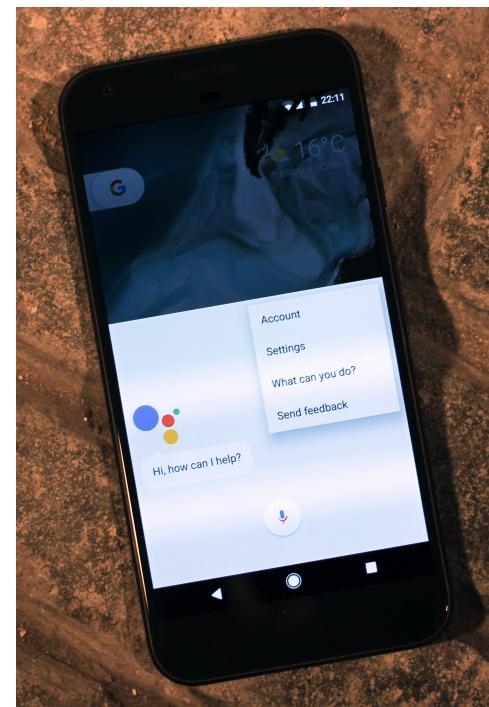


AI Problems Solved by Machine Learning Today

Recognition



(Face)



(Speech)



(Fraud)

AI Problems Solved by Machine Learning Today

Robotics



(Self-driving Vehicles)



(Medical Surgery)



(Manufacturing)

Today's Topics

- Class logistics
- Why machine learning?
- **What does a machine learn?**
- Designing a “supervised” machine learning algorithm
- Lab

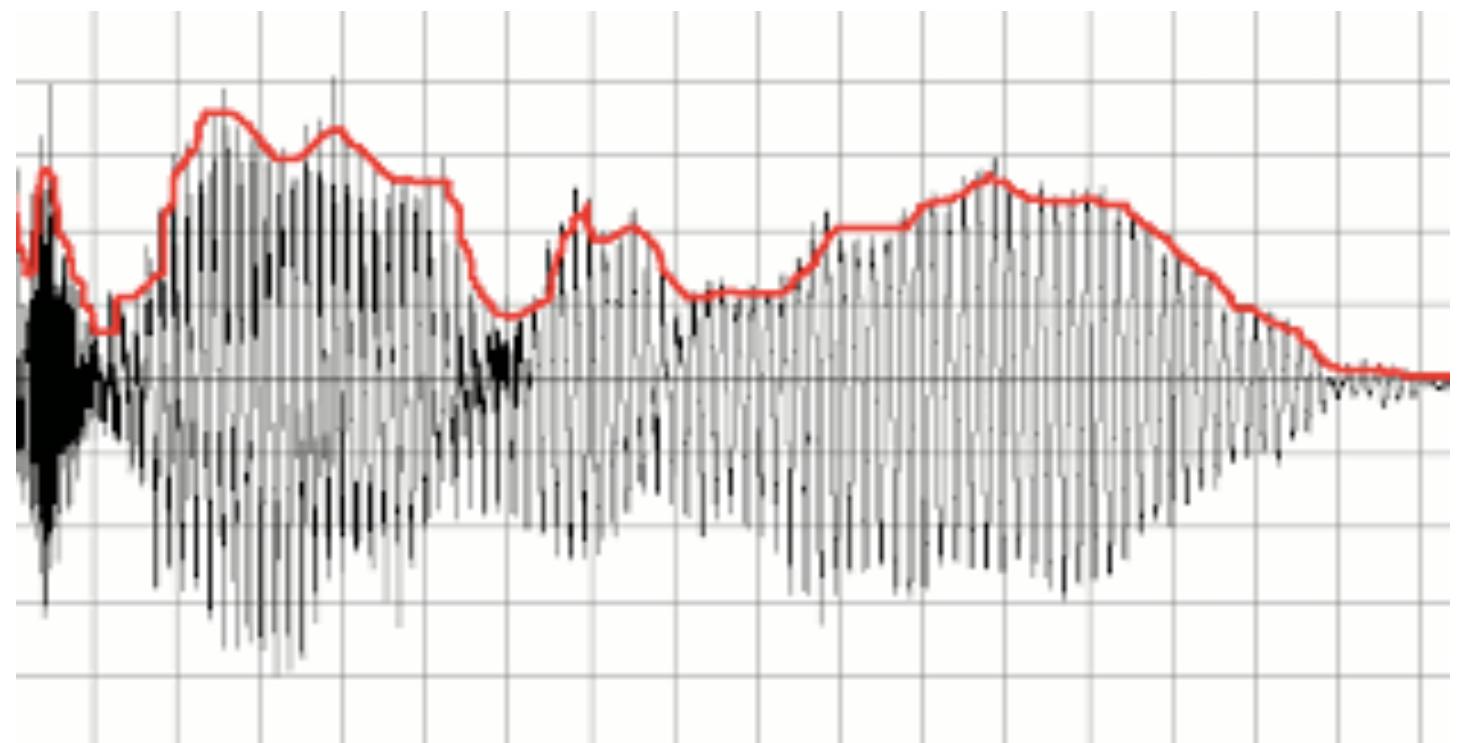
First: What Does a Machine Get?



- Audio
 - Input?



e.g.,

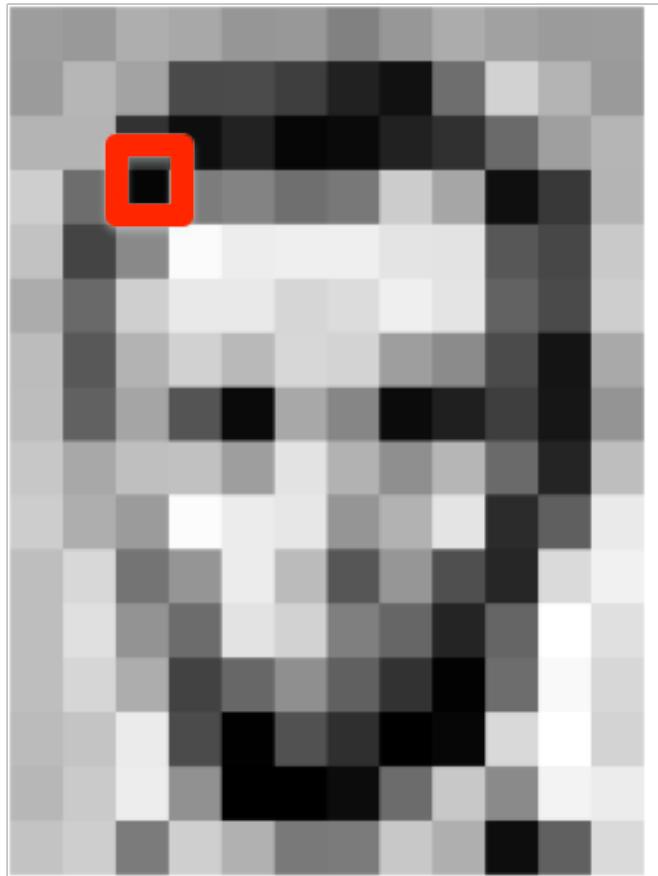


First: What Does a Machine Get?



- Audio
 - Input?
- Images
 - Input?

157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	40	14	34	6	10	33	48	106	159	181
206	105	54	131	111	120	204	166	15	56	180	156
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	239	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	156	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	109	143	96	50	2	109	249	216
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	176	13	96	218



255

First: What Does a Machine Get?



- Audio
 - Input?
- Images
 - Input?
- Video
 - Input?

157	153	174	168	160	152	129	151	172	161	155	156
155	182	163	74	75	62	39	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	227	87	71
188	88	179	209	185	215	211	158	139	75	20	169
189	97	166	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	195	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

Time 1

167	153	174	168	150	152	129	161	172	161	155	156
155	182	163	74	75	62	39	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	166	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	195	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

1 hour

Analogous to:



First: What Does a Machine Get?



- Audio
 - Input?
- Images
 - Input?
- Video
 - Input?
- Text
 - Input?

e.g.,

Confidential letter sh



David-Khoza@mmoscacsv.com

to ▾

.

2 Attachments



I would like to share this confidential and opportunity with you, with the hope that we can both work together for its success and mutual benefit. Our government service is currently, profiling and processing payments for their contractors with accumulated results to the current year. My proposition to you is to present you to our Ministry for profiling as a contractor and to showcase a beneficiary of contractor success to the value of vital addition.

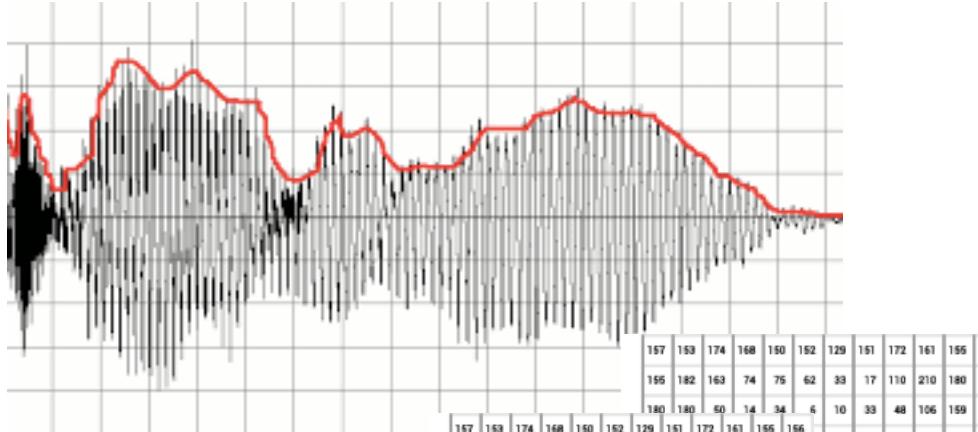
It is within my authority as the Director General in the ministry to do so in your profile by profiling as one of our contractors and beneficiaries of payment, regular and agreed that payment that will be transparent, accounted and approved by the ministry and its appointed agency bank. This will officially place you as the correct payment beneficiary. Once the contract payment is approved on your name, via ministry's appointed bank will pay you the annual UNQUOTE. This measure will be beneficial to both of us and I have received over all, motivation to achieve a successful completion of the ministry's task.

Note that you do not require to possess any qualification or profession to proceed. This transaction is free and there are no strings or obligation for ongoing periodic profiling. I guarantee that it will be measured under a legitimate arrangement without any breach of the law either here in South Africa or elsewhere. Once your contract payment is approved and the transfer is concluded to your account, with that in mind, you and I will withdraw the funds to the ratio of 50% for me, and 50% for you as your benefits. I urge for your full cooperation in this deal, and offer to maintain an amicable level of communication and confidentiality, respecting your position in the government ministry. The transaction will be done officially in the most secure manner while retaining an official and transparent banking procedures.

First: What Does a Machine Get?

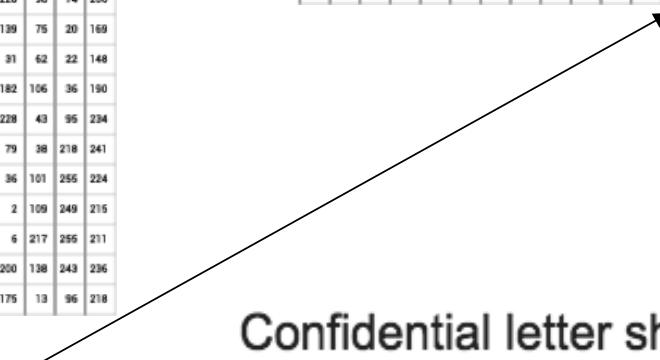


- Audio
 - Input?
- Images
 - Input?
- Video
 - Input?
- Text
 - Input?
- Multi-modal
 - Input? - combination of the above



157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	38	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	228	227	87	71	201	
172	105	207	233	233	214	220	239	228	98	74	148
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	193	158	139	75	20	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

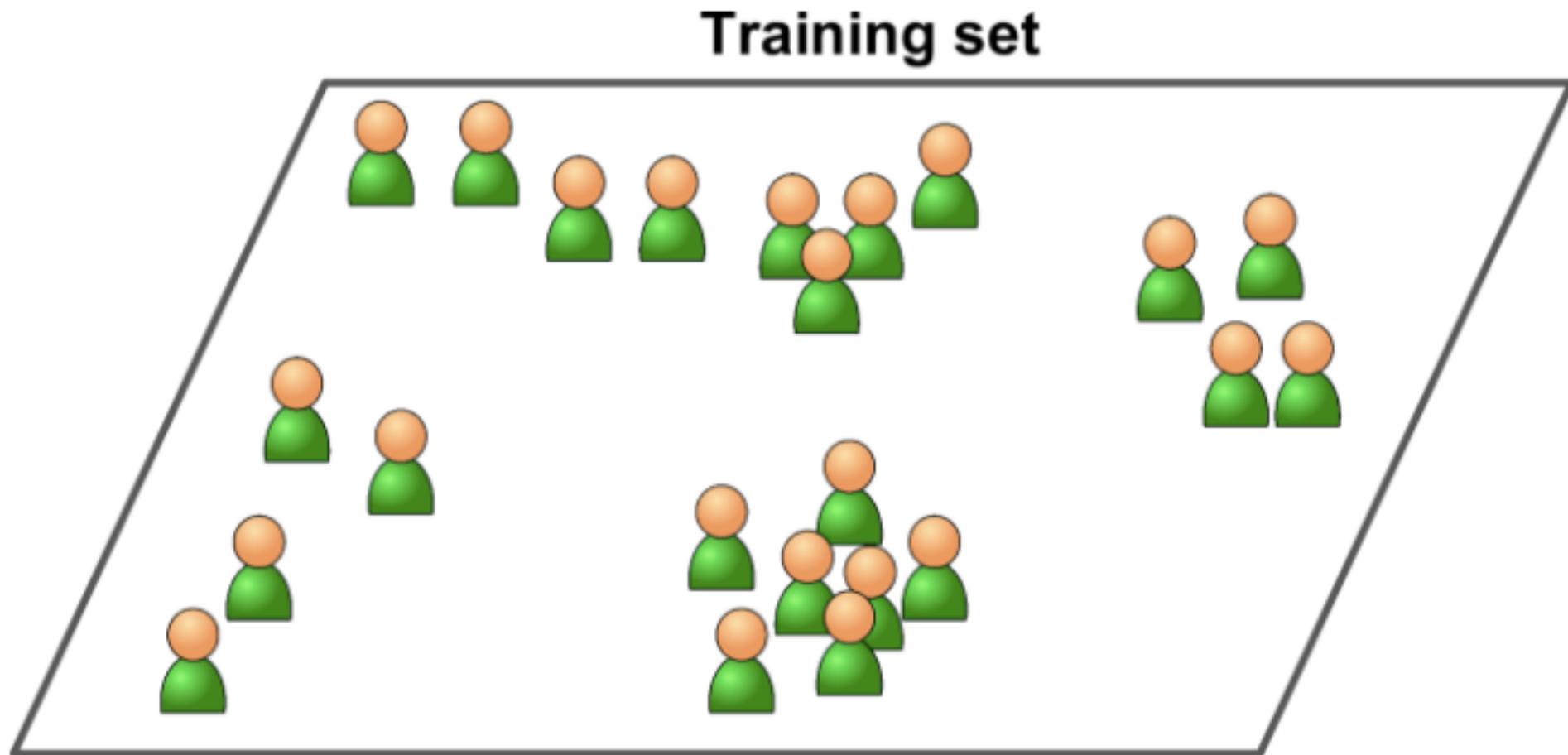
157	153	174	168	160	152	129	151	172	161	155	156
155	182	163	74	75	62	38	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	228	227	87	71	201	
172	105	207	233	233	214	220	239	228	98	74	148
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218



Confidential letter sh

? David-Khoza@mmoscacsv.com
to

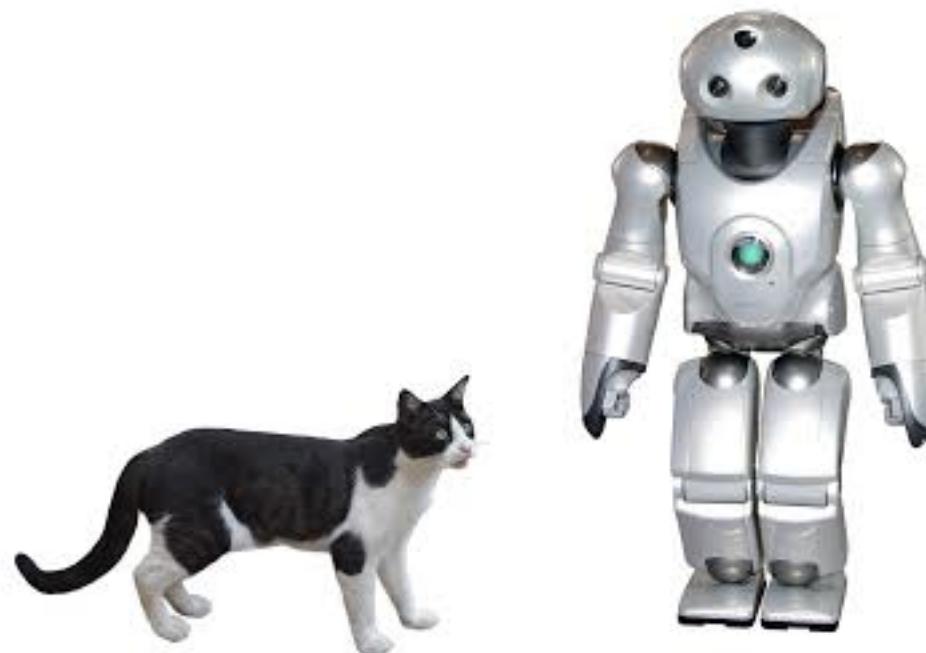
First: What Does a Machine Get?



Types of Learning Tasks

- Unsupervised
 - No label given for training data
- Supervised
 - Label given for training data: e.g., “cat”

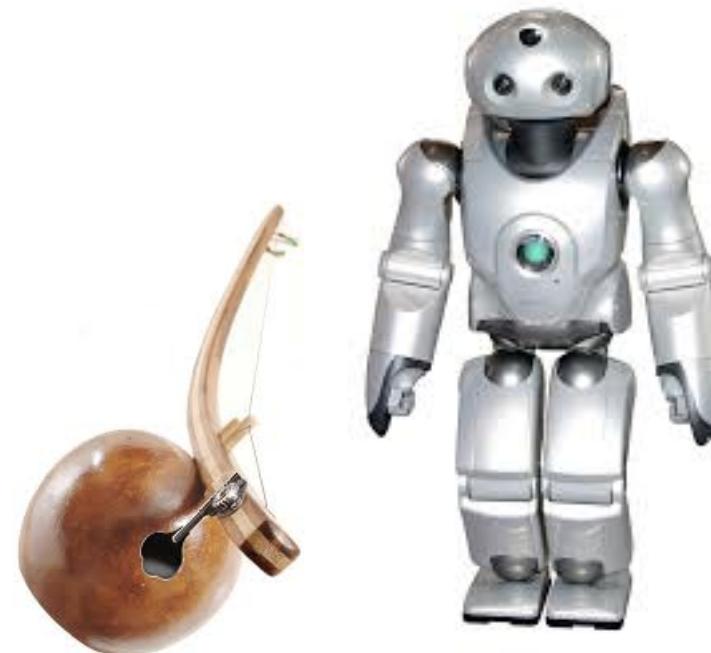
What is this?



Types of Learning Tasks

- Unsupervised
 - No label given for training data
- Supervised
 - Label given for training data: e.g., “berimbau”

What is this?



Types of Learning Tasks

- Unsupervised
 - No label given for training data

Is this email spam?

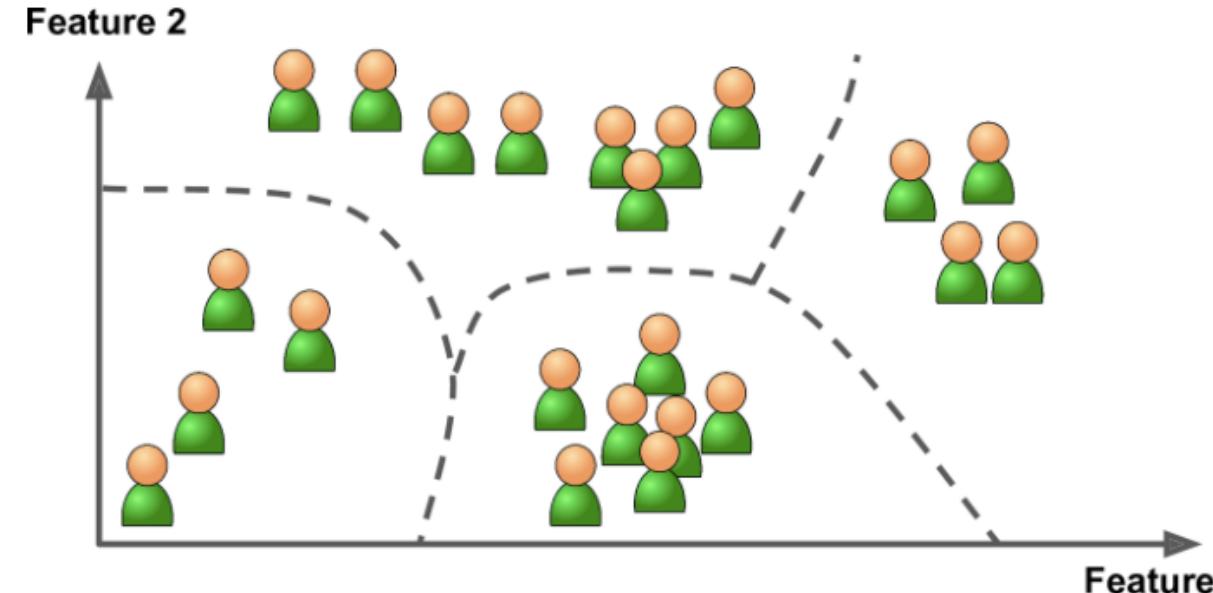
Confidential letter sh
? David-Khoza@mmoscacsv.cc
to



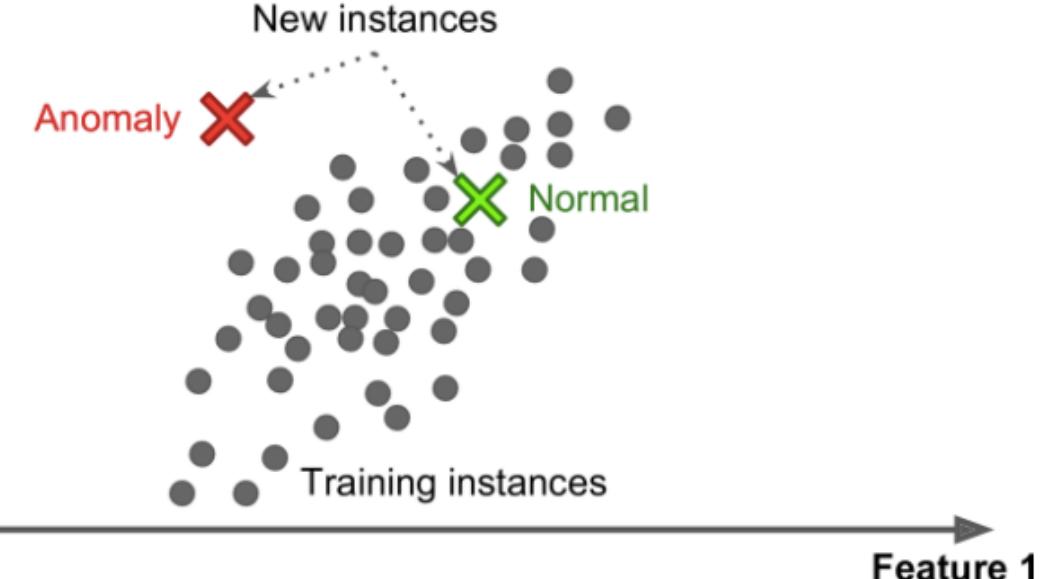
- Supervised
 - Label given for training data: e.g., “yes”

Types of “Unsupervised” Learning

Clustering



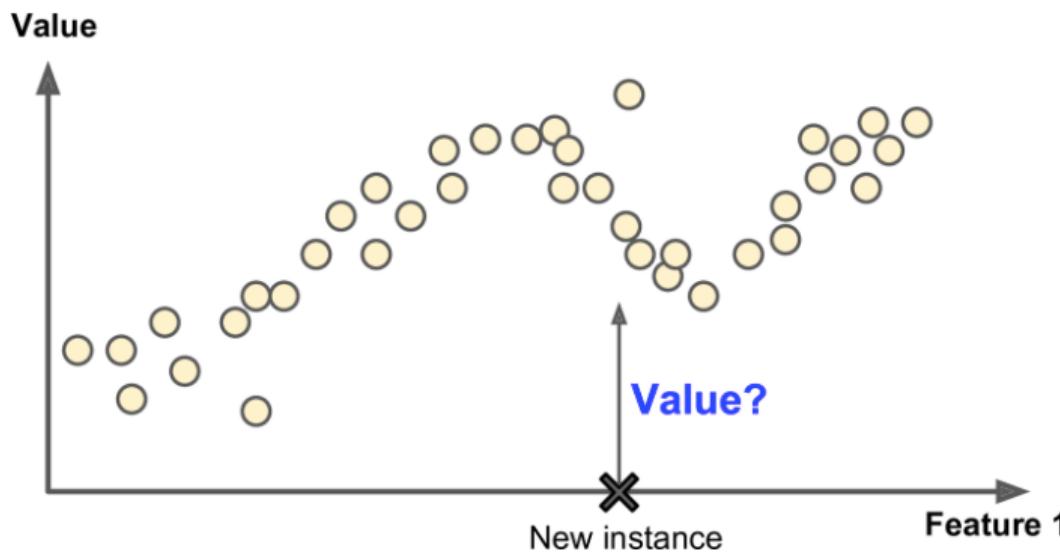
Anomaly Detection



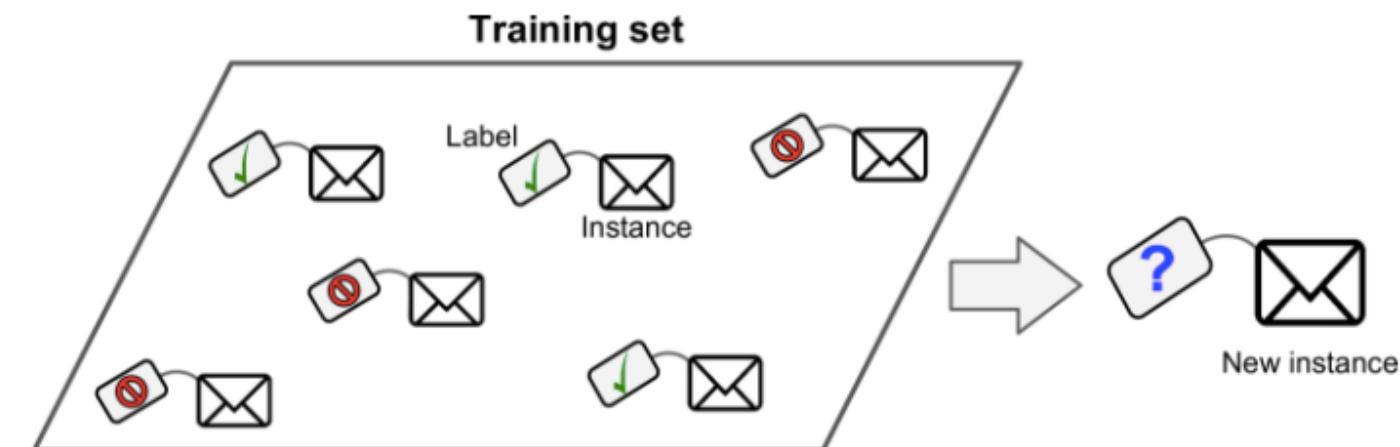
What are real world applications for these types?

Types of “Supervised” Learning

Regression
(predict **continuous** value)



Classification
(predict **discrete** value)



What are real world applications for these types?

Learning Challenge: Sufficient Training



Little Training



Highly Trained

Learning Challenge: Sufficient Training

e.g., images

images on basic hard drive:
(500 GB/2 MB = 250,000)

10^5



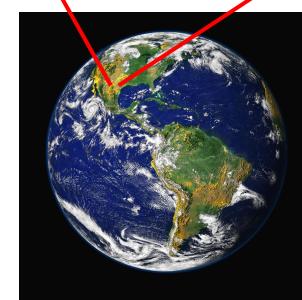
images seen during my first 10 years:
(24 images/sec * 60 sec * 60 min * 16 hr * 365 days * 10 yrs = 5,045,760,000)

10^9



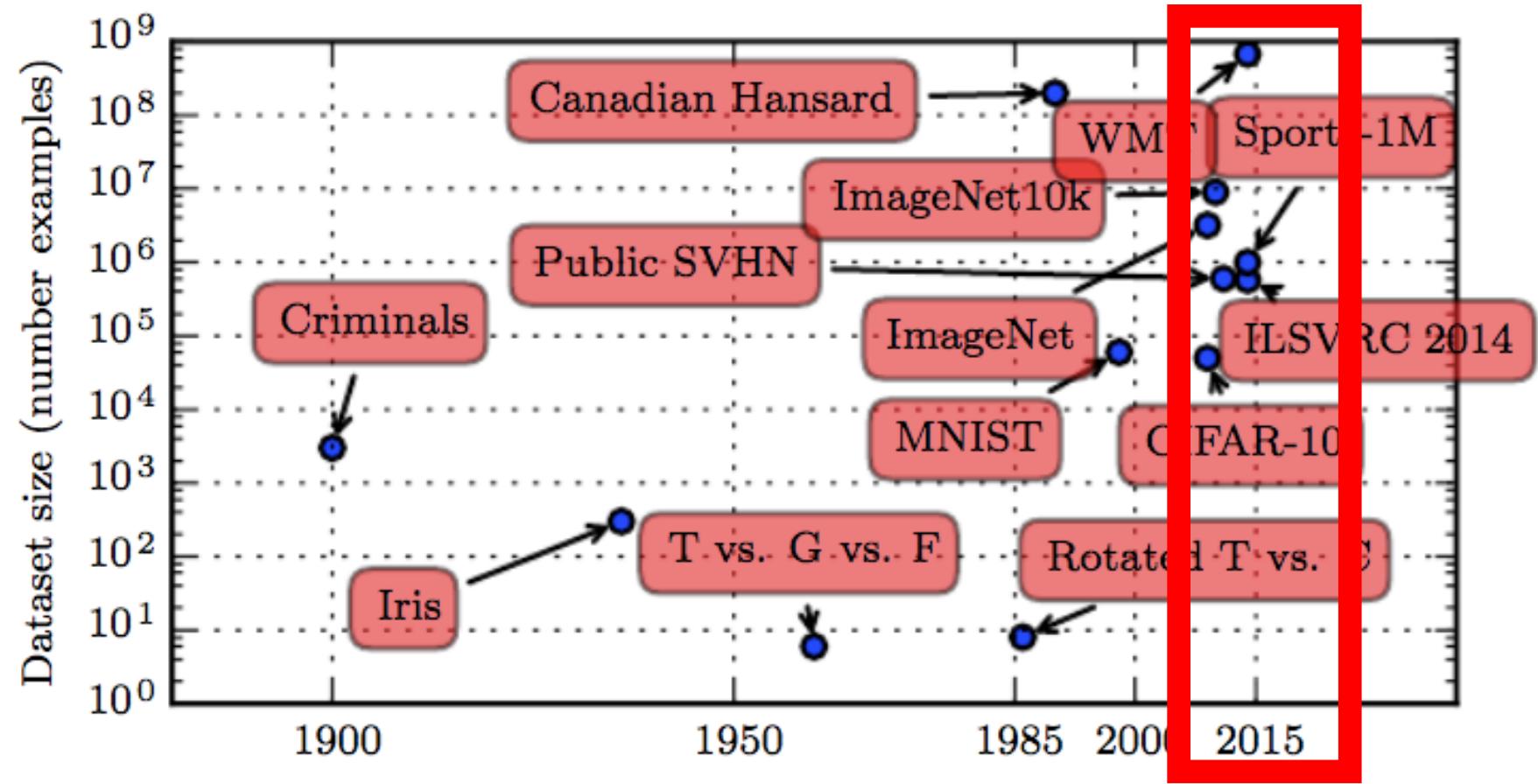
images seen by all humanity:
(7.5 billion humans¹ * 24 images/sec * 60 * 60 * 16 * 365 * 60 yrs = 2.23×10^{20})

¹ <http://www.worldometers.info/world-population/>



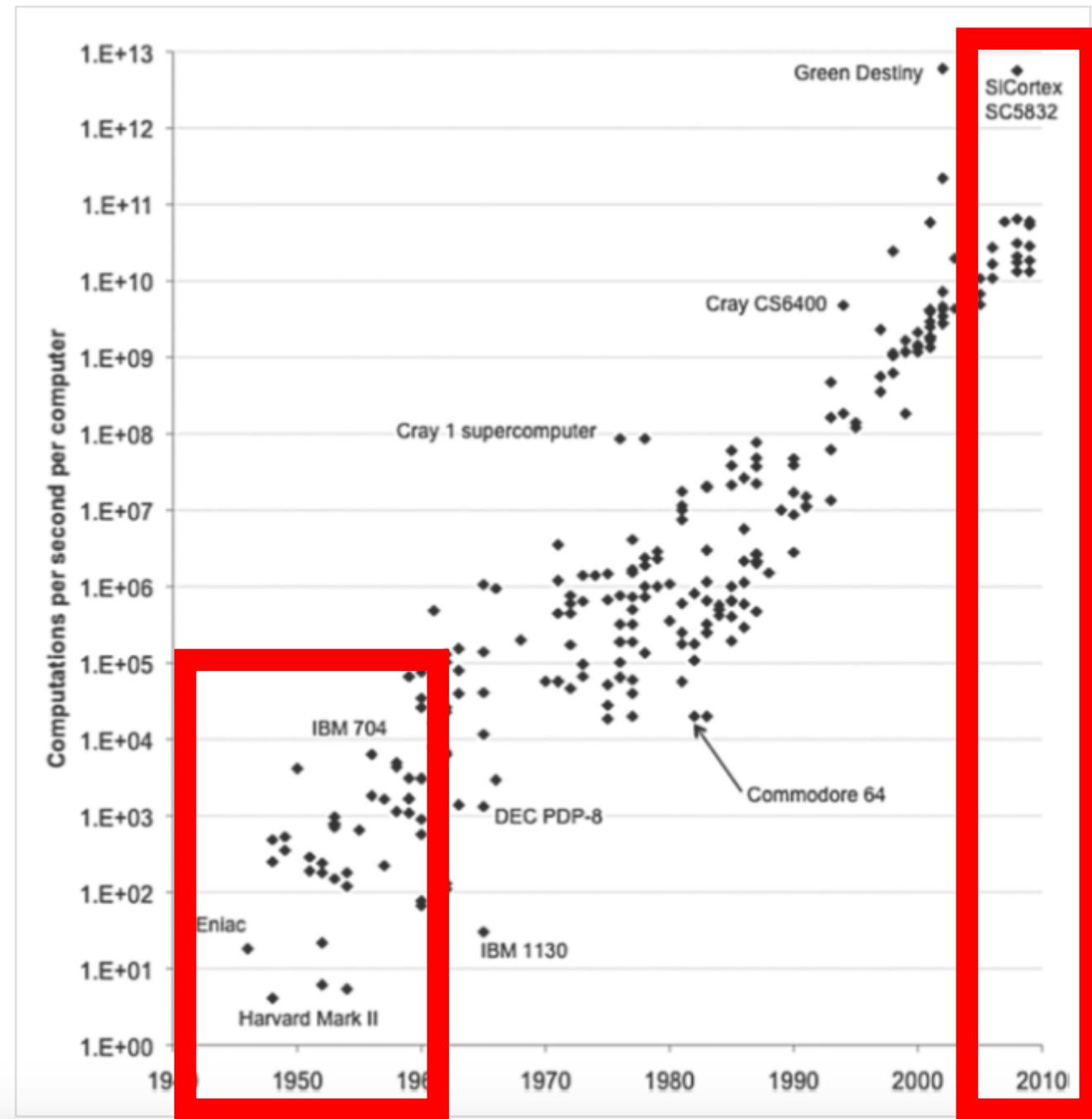
Learning: Where Machine Learning is Today

- “Big”-ger Data
 - e.g., internet



Learning: Where Mac

- “Big”-ger Data
 - e.g., internet
- Better Hardware
 - e.g., faster processing -- GPUs



Learning: Where Machine Learning is Today

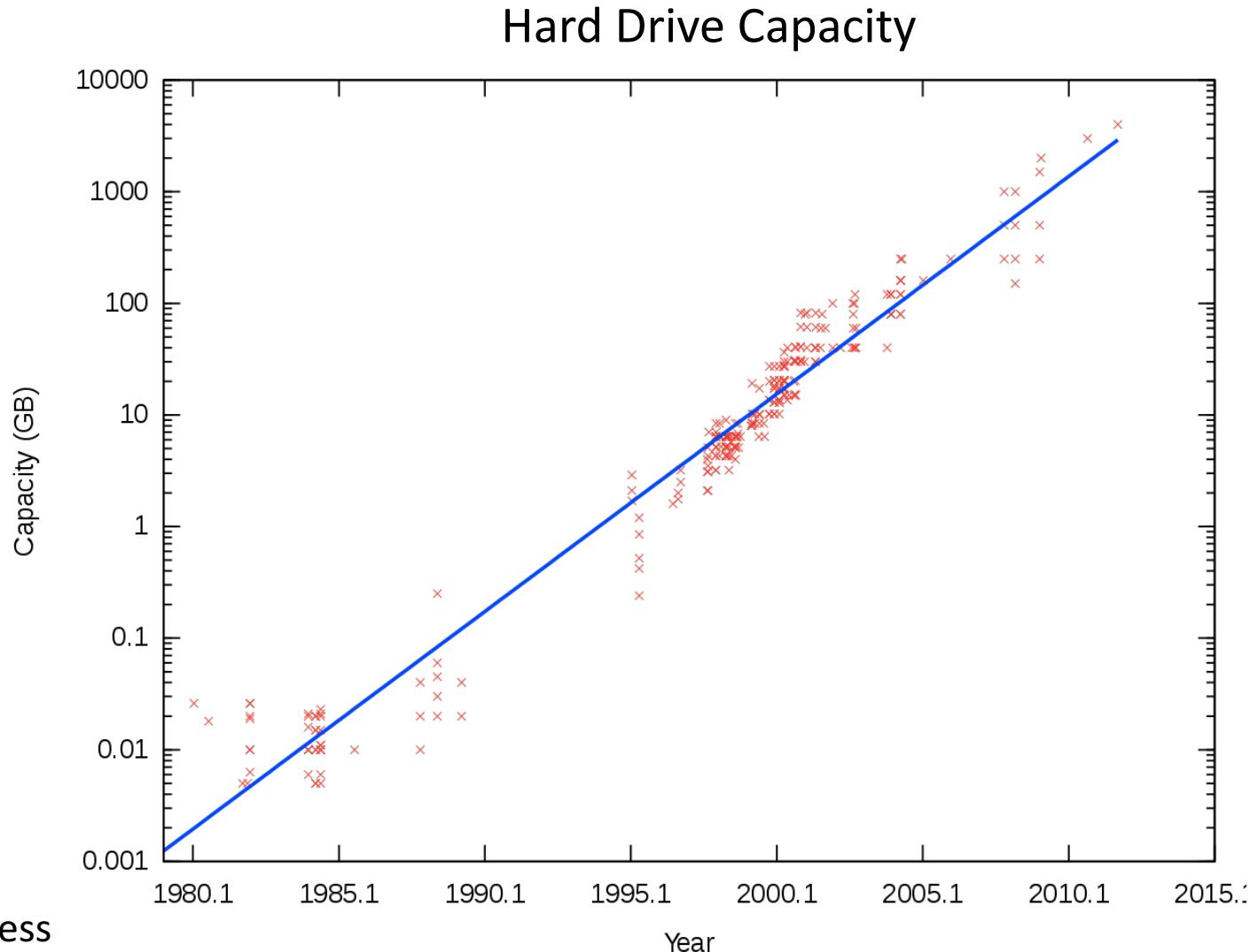
- “Big”-ger Data
 - e.g., internet
- Better Hardware
 - e.g., faster processing -- GPUs
 - e.g., **more memory**

The IBM Model 350 disk file with a storage space of 5MB from [1956](#) and a Micro SD Card



Learning: Where Machine Learning is Today

- “Big”-ger Data
 - e.g., internet
- Better Hardware
 - e.g., faster processing -- GPUs
 - e.g., **more memory**



Learning: Where Machine Learning is Today

- “Big”-ger Data

In the past 59 years, machines were:

- Better Hardware
 - e.g., faster processing -- GPUs
 - e.g., more memory

- (1) given more training data,
- (2) built to think faster, and
- (3) built to remember more!

What does your money in this class buy you?

- State-of-Art Hardware
 - e.g., faster processing -- GPUs
 - e.g., more memory



- Thanks to: Microsoft Azure

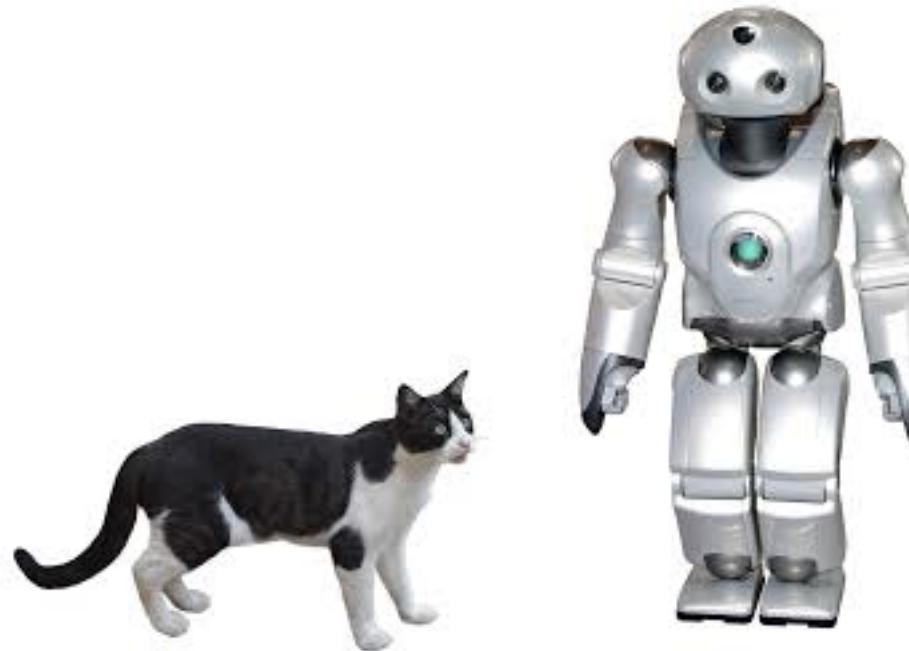
Today's Topics

- Class logistics
- Why machine learning?
- What does a machine learn?
- Designing a “supervised” machine learning algorithm
- Lab

How to Evaluate a Machine’s “Learning”?

- Training Data:

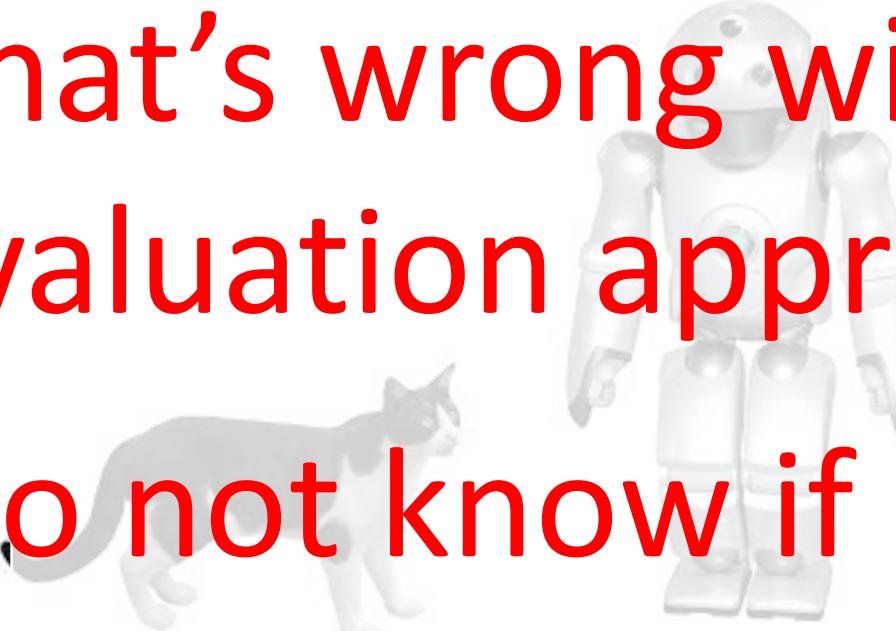
What is this?



How to Evaluate a Machine’s “Learning”?

- Testing Data:

What is this?



What's wrong with this
evaluation approach?

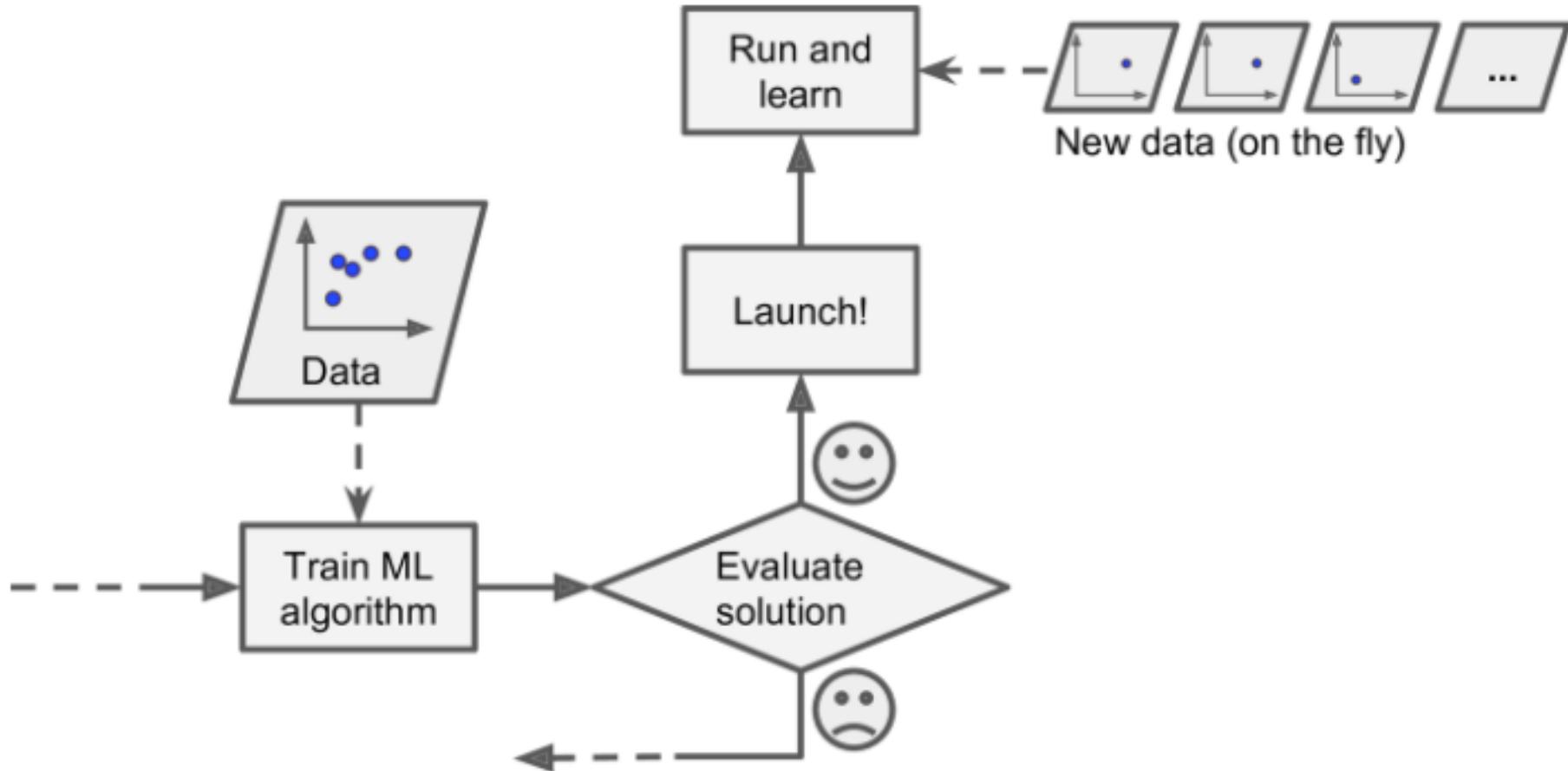
We do not know if a machine's
learned knowledge generalizes.

How to Evaluate a Machine’s “Learning”?

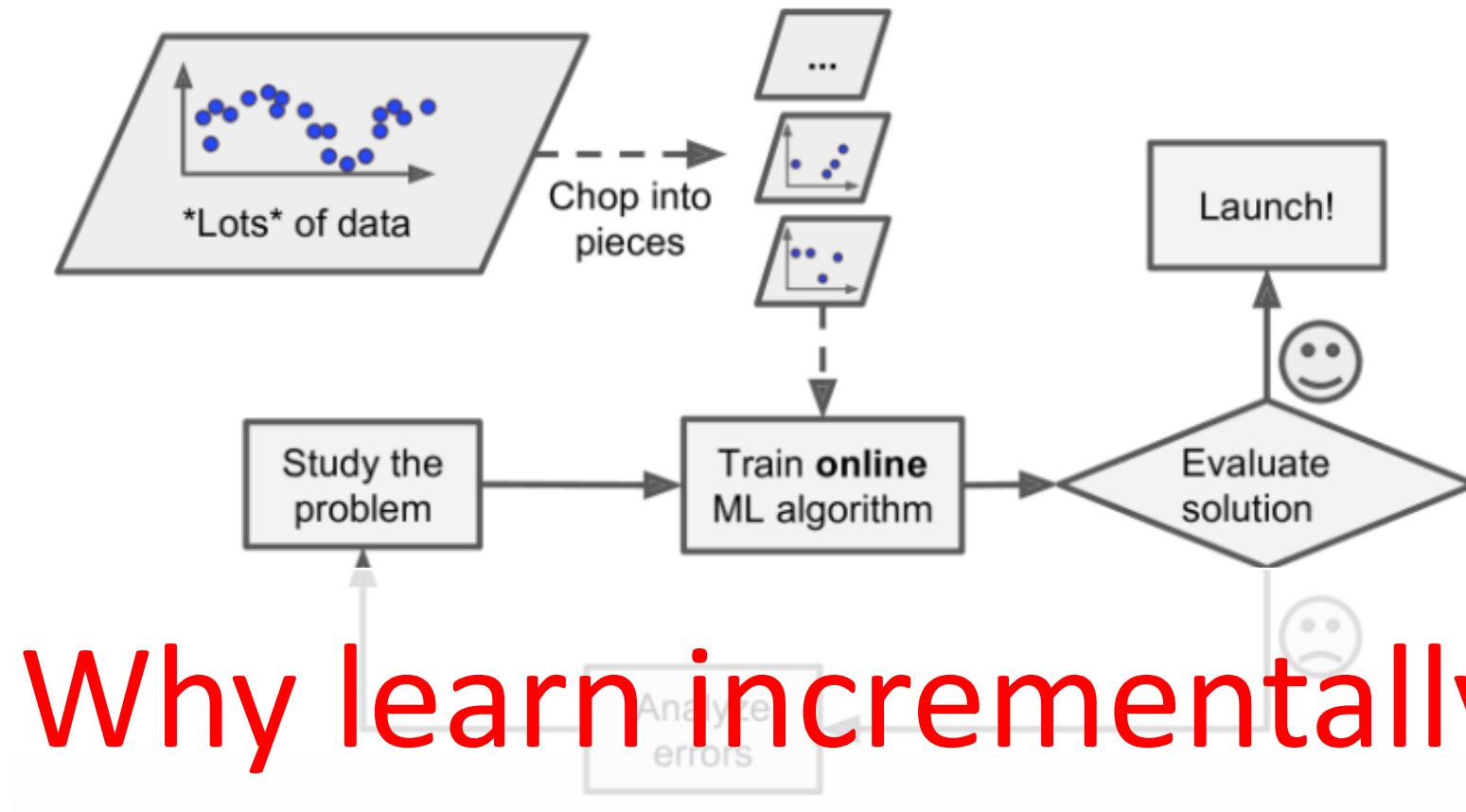
- Split data into a “**training set**” and “**testing set**”

	Feature 1	Feature 2	...	Feature M	Label
Sample 1:	0.7	100	● ● ●	0.81	Yes ● ●
Sample N:	0.5	121	● ● ●	0.3	No

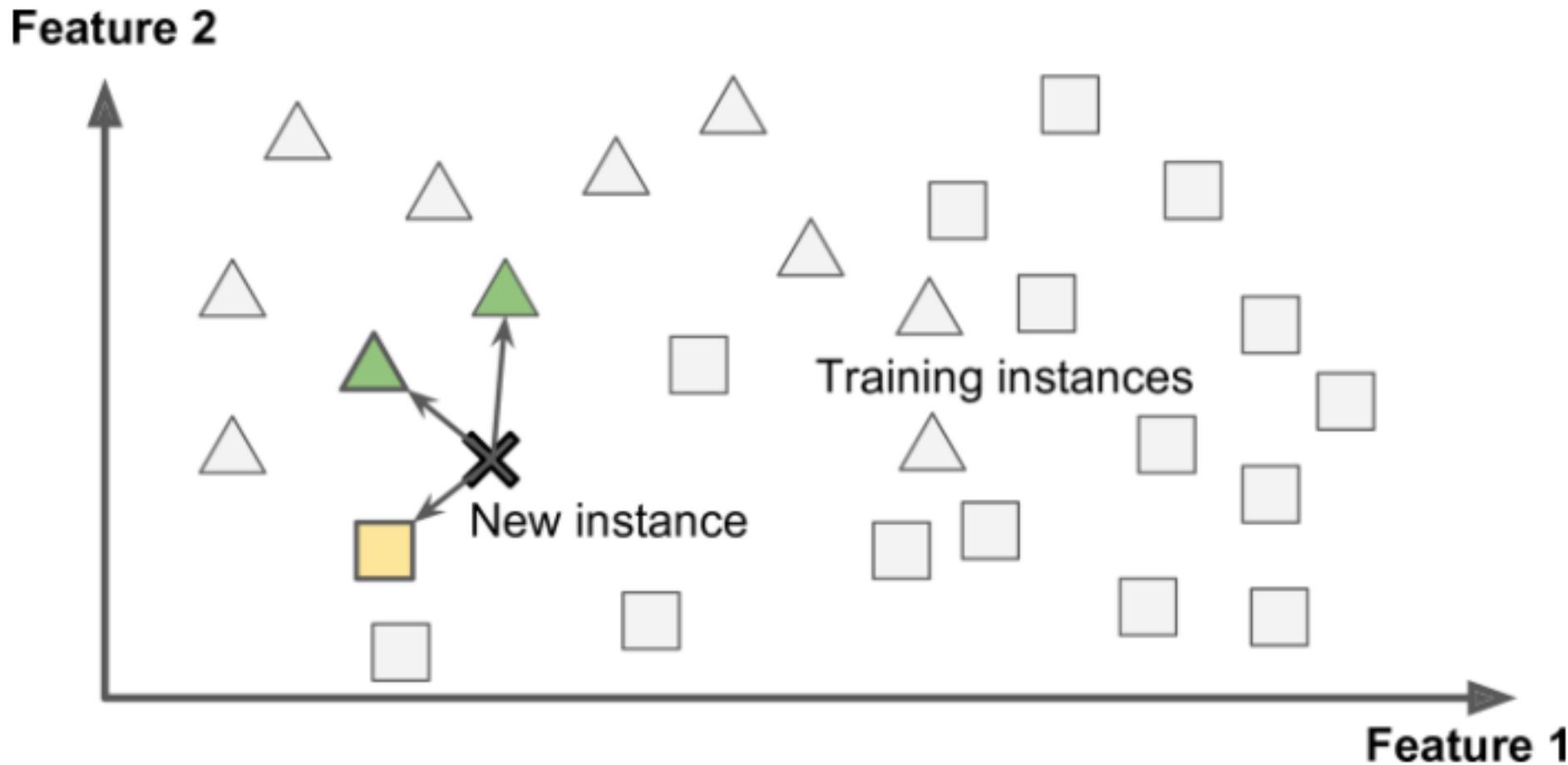
How to Teach a Machine? Online Learning vs Offline Learning



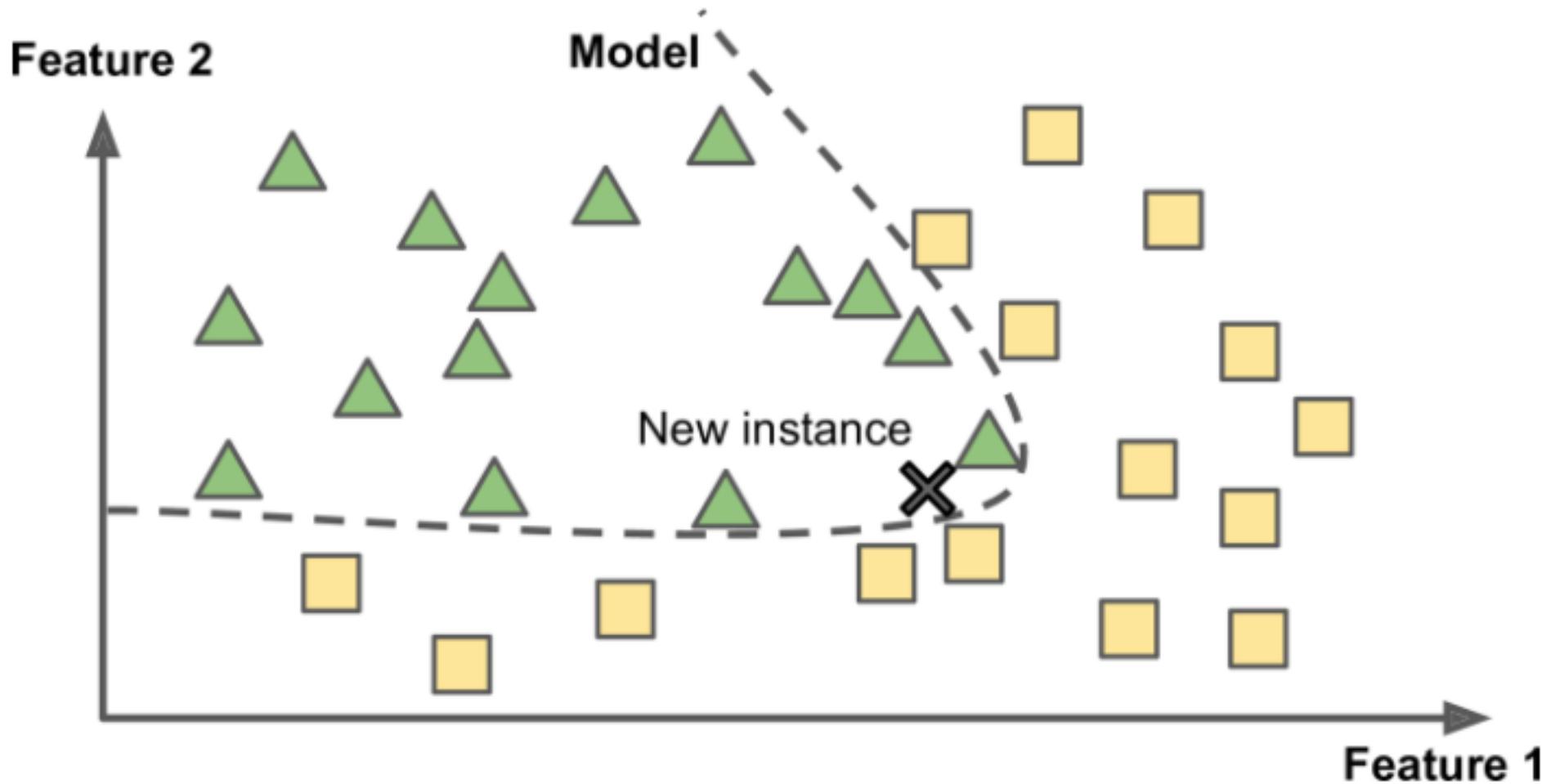
How to Teach a Machine? Online Learning vs Offline Learning



How to Teach a Machine? Instance-Based Learning



How to Teach a Machine? Model-Based Learning



Machine Learning Model

- Goal: learn data distribution in the “real world”
- Modeling:
 - Can **increase** model’s representational capacity
 - e.g., linear versus quadratic?
 - Can **decrease** model’s representational capacity
 - e.g., linear versus quadratic?

Machine Learning Model

- Underfitting
 - Model is too simple
- Overfitting
 - Model is too complex
- Regularization
 - Simplifies a complex model

Today's Topics

- Class logistics
- Why machine learning?
- What does a machine learn?
- Designing a “supervised” machine learning algorithm
- Lab

Account Set-up

- Log into a machine: EID and UT password
- Request iSchool account (if you do not yet have an account)
 - https://www.ischool.utexas.edu/accounts/request_account.php
- Reset iSchool password (if you do not yet have an account)
 - Instructions included in email approving the account