Applied Machine Learning

Introduction

Reihaneh Rabbany

McGill
School of Computer Science

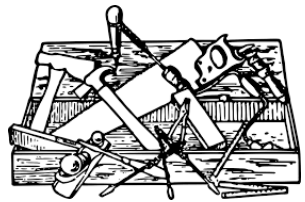
COMP 551 (winter 2020)

Objectives

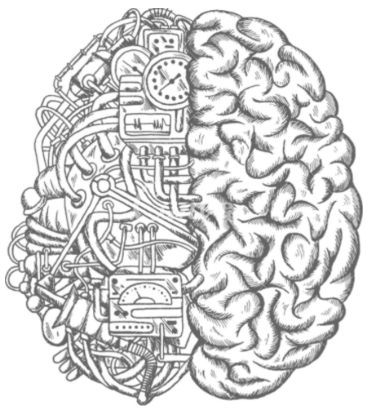
- understanding the scope of machine learning
  - relation to other areas
- understanding types of machine learning

What is Machine Learning?

ML is the set of "algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions"



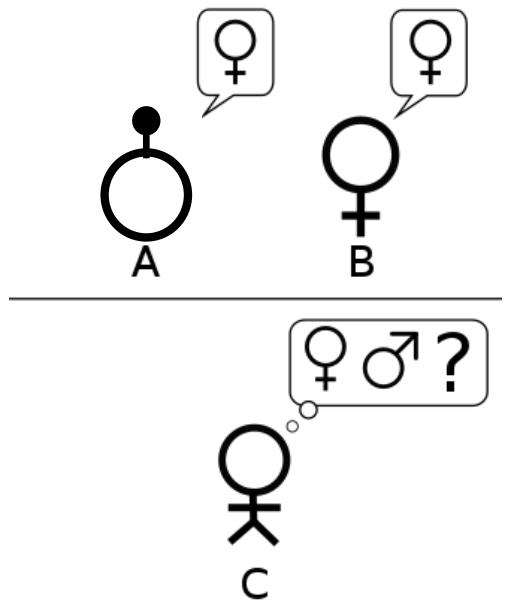
an inadequate history of ML



- 1950: Turing test
- 1956: checker player that learned as it played (Arthur Samuel)
  - coined the term Machine Learning
- 1958: first artificial neural networks called Perceptron (Frank Rosenblatt),
  - 1959: ADELINE (Widrow and Hoff)
- 1963: support vector machines (Vapnick & Ya)
- 1969: Minskey and Pappert show the limitations of single-layer neural networks
- 1970-80s rule-based and symbolic AI dominates (two AI winters)
- 1980's Bayesian networks (Judea Pearl)
- 1986 Backpropagation rediscovered (Rumelhart, Hinton & Williams)
- 1991 Kernel trick for SVM
- 2012 AlexNet wins Imagenet by a large margin
- 2012 - now deep learning explosion...
- next? AI winter? AGI?

Imitation game

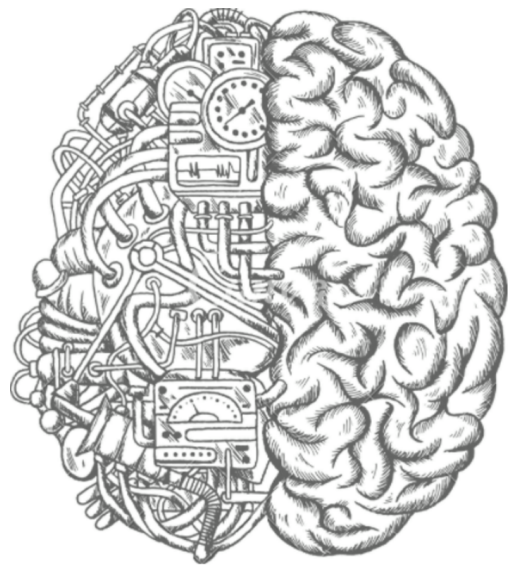
What will happen when a machine takes the part of A in this game? Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, **"Can machines think?"**



Computing Machinery and Intelligence by Alan Turing (**1950**)

Thinking Machines, the search for Artificial Intelligence (AI)

John McCarthy coined the term Artificial Intelligence and organized the first AI conference in 1955



Checkers

A computer can learn to play a better game of checkers than its programmer

list of parameters that have something to do with the game but whose values are inknown and unspecified

Some Studies in **Machine Learning** Using the Game of Checkers by Arthur Samuel (1959)

*“ I learned of Strachey’s work from a paper he presented in Toronto in September 1952. Because his program was, at that time, already in published form, I must concede defeat. It was not until **1954**, with the advent of the IBM 704, that my program was able to play an interesting game. My contribution was to add “learning” to the program, and I believe that I can claim a first for this. Strachey didn’t pursue further development of his program.*

Some Studies in **Machine Learning** Using the Game of Checkers by Arthur Samuel (1959)

The Perceptron —

A Perceiving and Recognizing Automaton

by Frank Rosenblatt (1957)

The first device to think as the human brain

Learns by doing

Learned to differentiate between right and left after 50 attempts, registering a Q for the left squares and an O for the right squares

Image Recognition

How does human think?

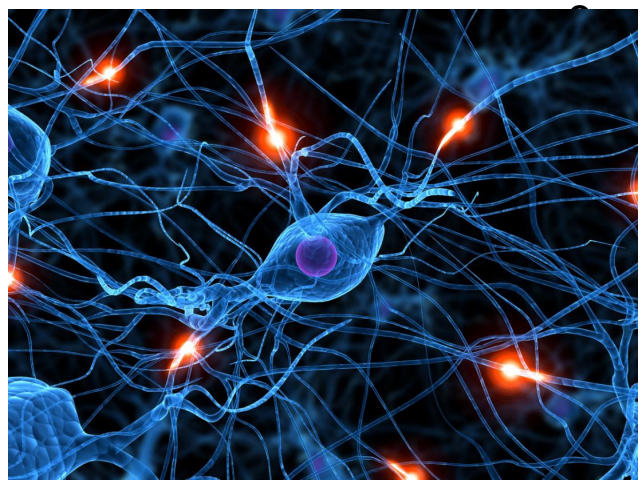
Earliest works in the 1930s and 1940 by Donald Hebb, a psychologist at McGill

Studying behaviour in terms of brain function and connections between neuron assemblies

1943: First mathematical model by Warren S. McCulloch, a neuroscientist, and Walter Pitts, a mathematician

> *" The brain of man, like that of the animals, is made up of many cells of a certain type called neurons. These cells... react on an all-or-none basis ('fire'; ...) and transmit a pulse to other neurons through synaptic connections. Each neuron is connected to many others, and a number of input signals are, in general, required before a neuron will 'fire'. ...Learning seems to consist of alterations in the strength and even perhaps in the number of these synaptic interconnections. Now it is possible to devise a variety of mechanical, chemical, and electrical devices which simulate the behavior of individual neurons in a crude sort of way, and we can interconnect these devices in some random fashion to simulate the synaptic interconnections that exist within the brain, and, finally, we can arrange for the automatic strengthening or weakening of these interconnections using a training routine.*

Samuel, A. L. (1962). Artificial intelligence: a frontier of automation. The Annals of the American Academy of Political and Social Science, 340(1), 10-20.



Multilayer Perceptrons

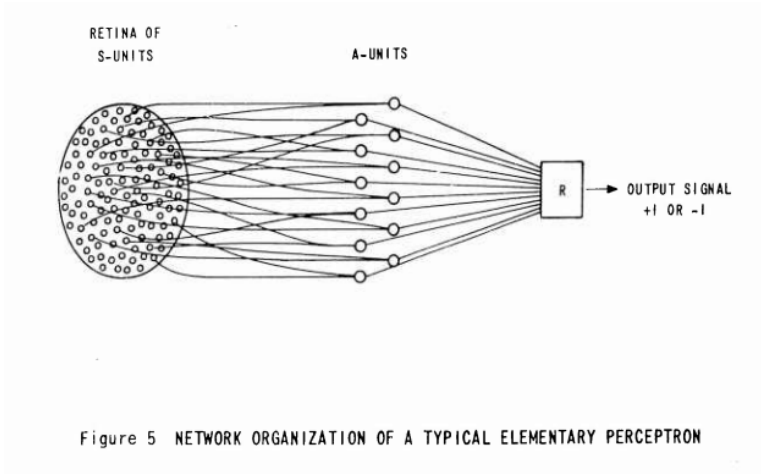
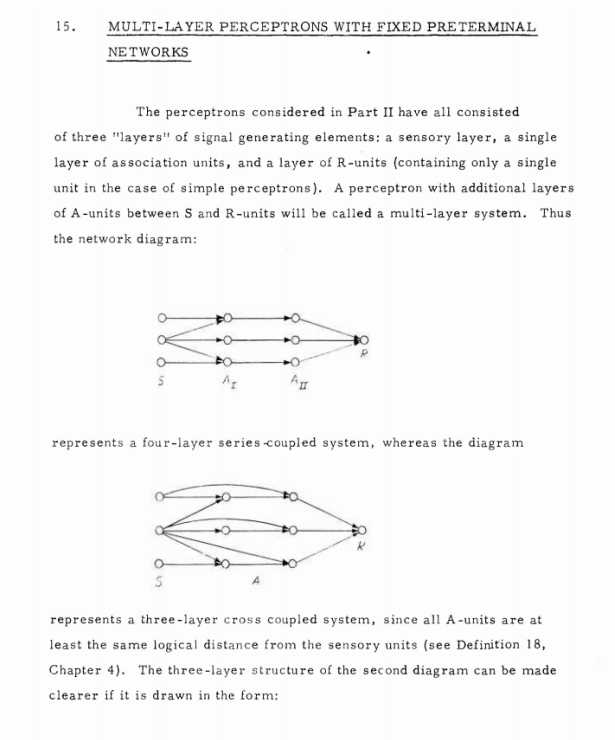


Figure 5 NETWORK ORGANIZATION OF A TYPICAL ELEMENTARY PERCEPTRON

Principles of neurodynamics. Perceptrons and the theory of brain mechanisms by Rosenblatt (1961)



15. MULTI-LAYER PERCEPTRONS WITH FIXED PRETERMINAL NETWORKS

The perceptrons considered in Part II have all consisted of three "layers" of signal generating elements: a sensory layer, a single layer of association units, and a layer of R-units (containing only a single unit in the case of simple perceptrons). A perceptron with additional layers of A-units between S and R-units will be called a multi-layer system. Thus the network diagram:

represents a four-layer series-coupled system, whereas the diagram

represents a three-layer cross coupled system, since all A-units are at least the same logical distance from the sensory units (see Definition 18, Chapter 4). The three-layer structure of the second diagram can be made clearer if it is drawn in the form:

ADALINE and MADALINE and LMS algorithm

Adaptive Linear Neuron and Many ADALINEs

by Bernard Widrow et al., 1958 and 1960

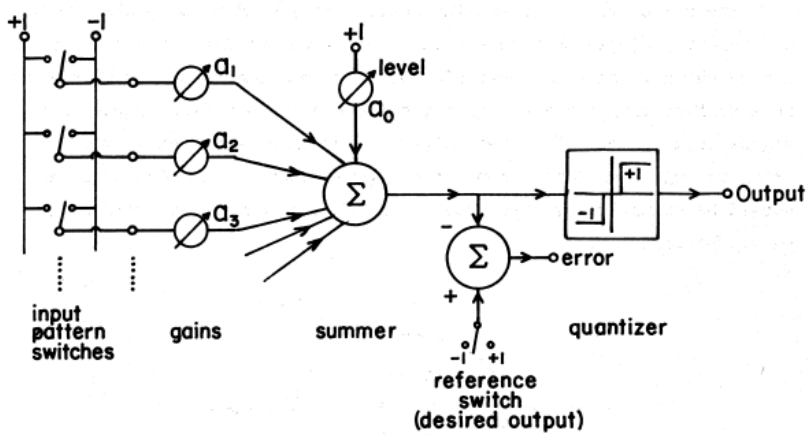
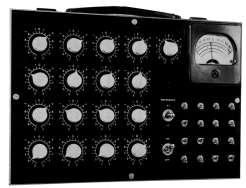
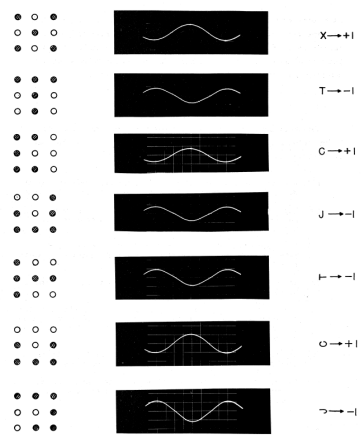


FIG. 3.--SCHEMATIC OF ADALINE.

Letter Recognition





60 ~ sine waves; vertical scale is 0.1 volts/cm

FIG. 12. WAVE-FORMS OF A MEMISTOR NEURON AFTER A TRAINING EXPERIMENT.

Real World Applications?

Speech Recognition (1963)

29. L.R. Talbert, G.F. Groner, J.S. Koford, R.J. Brown, P.R. Low, and C.H. Mays, *A Real-Time Adaptive Speech-Recognition System*, Technical Report, Stanford University, 1963.

Adaptive Speech-Recognition System. L. R. ember), G. F. GRONER (nonmember), AND nmember), *Department of Electrical Engineer-iversity, Stanford, California.*—This paper de-scribes a real-time speech-recognition system employing adaptive threshold logic elements called "Adalines." Time-normalized digital patterns representing the time-frequency spectrum are obtained from amplitude-normalized outputs of 8 bandpass filters. Adaline networks that perform the speech-tion are simulated in an IBM-1620 computer. iment, the Adalines were trained on 8 samples of 16 phonetically balanced words. After cor-of these samples, different samples, spoken by the same speaker, were correctly identified 112 times without error. When tested on new voices of the same sex, the machine achieved an average recognition rate of 90%; however, substantial improvement is realized by including the the training group. The use of adaptive net-n classifiers has achieved enormous system complete redesign of the classification system can be accomplished by a training process. The system has successfully carried out many speech-recognition tasks, among them recognition of the 10 digits spoken in 4 different languages and identification of different speakers saying the same spoken

Weather Forecasting (1964)

30. M.J.C. Hu, *Application of the Adaline System to Weather Forecasting*, E.E. Degree Thesis, Technical Report 6775-1, Stanford Electron. Labs., Stanford, CA, June 1964.

Modems (1965)

36. R.W. Lucky, "Automatic equalization for digital communication," *Bell Syst. Tech. J.*, **44**, 547–588 (1965).

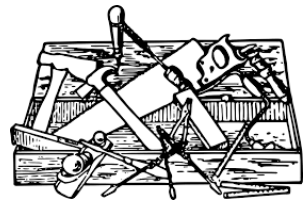
Telephones (1965)

38. M.M. Sondhi, "An adaptive echo canceller," *Bell Syst. Tech. J.*, **46**, 497–511 (1967).

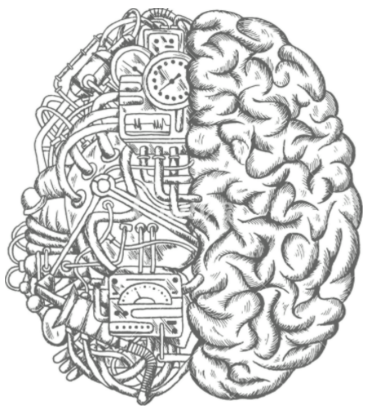
Adaptive neural networks and their applications. International Journal of Intelligent Systems (1993) by Widrow B, Lehr MA.

What is Machine Learning?

ML is the set of "algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions"



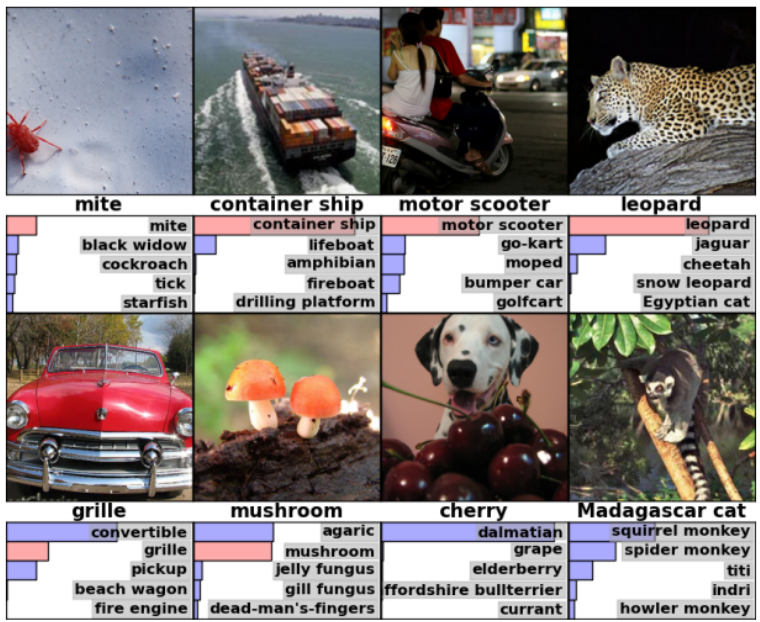
an inadequate history of ML

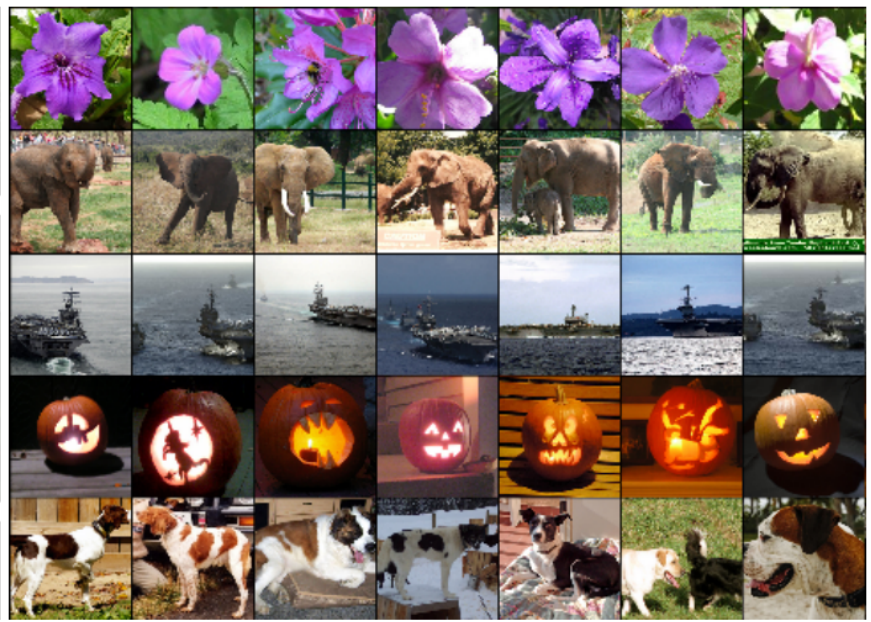


- 1950: Turing test
- 1959: checker player that learned as it played (Arthur Samuel)
  - coined the term Machine Learning
- 1958: first artificial neural networks called Perceptron (Frank Rosenblatt),
  - 1959: ADELINE (Widrow and Hoff)
- 1963: support vector machines (Vapnick & Ya)
- 1969: Minskey and Pappert show the limitations of single-layer neural networks
- 1970-80s rule-based and symbolic AI dominates (two AI winters)
- 1980's Bayesian networks (Judea Pearl)
- 1986 Backpropagation rediscovered (Rumelhart, Hinton & Williams)
- 1991 Kernel trick for SVM
- 2012 AlexNet wins Imagenet by a large margin
- 2012 - now deep learning explosion...
- next? AI winter? AGI?

Deep Blue vs Kasparov 1997

AlexNet 2012





AlphaGo vs Ke Jie 2017

computers could be developed to perform better than humans in highly complex tasks

Google's AlphaGo Defeats Chinese Go Master in Win for A.I.

Ke Jie, the world's top Go player, reacting during his match on Tuesday against AlphaGo, artificial intelligence software developed by a Google affiliate. China Stringer Network, via Reuters

Placing ML: overlapping fields

Artificial Intelligence: broader includes search, planning, multiagent systems, robotics, etc.
Statistics: ML is more focused on algorithmic, practical and powerful models (e.g., neural networks) and is built around AI
Pattern Detection: automatic discovery of regularities in data, basis of ML
Vision & Natural Language Processing: the origin of many advances
Data mining: more emphasis on making algorithms that work in the real world, scalability and performance comes before having strong theoretical foundations, more space for using heuristics, exploratory analysis and unsupervised algorithms
Data science: an umbrella term for the above mostly used in industry when the output is knowledge/information to be used for decision making

Placing ML: main venues

top computer science conferences

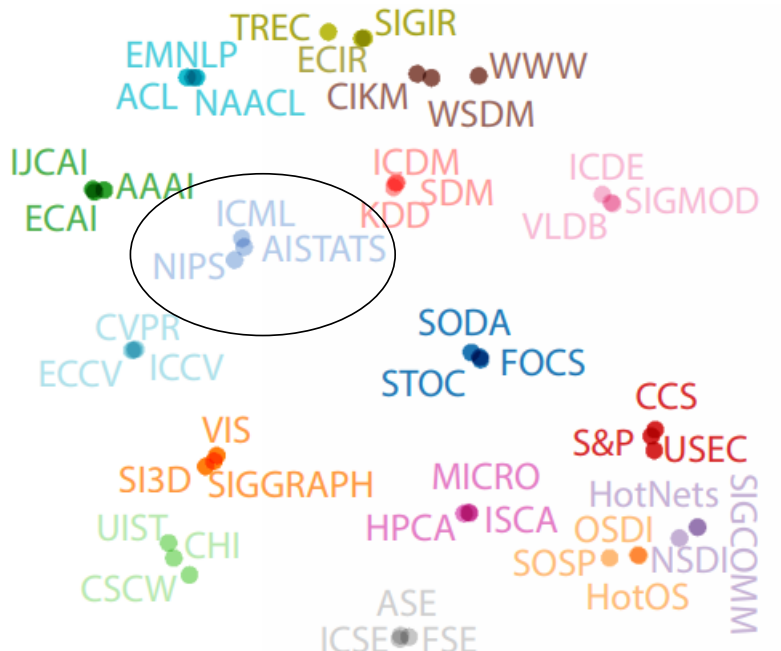
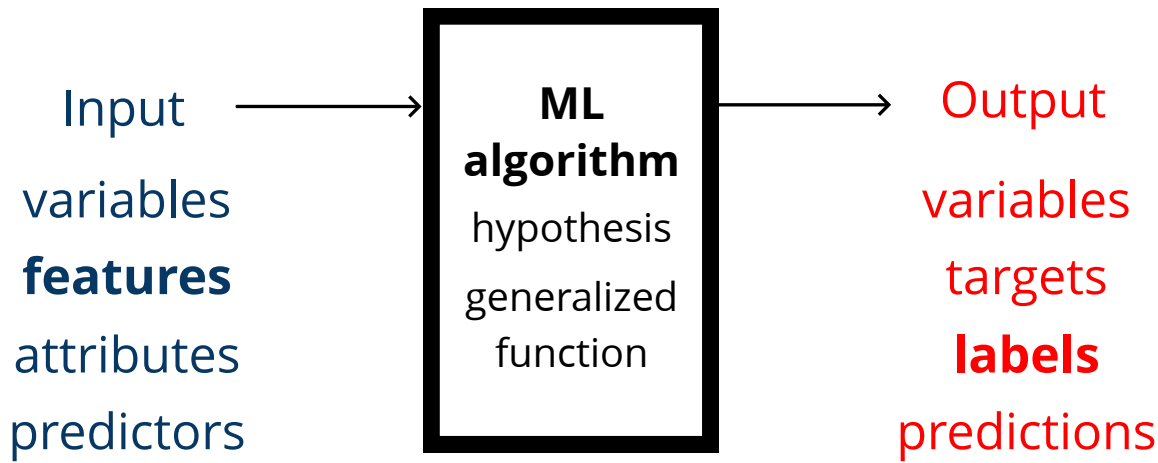


Figure from Dong et al.

Common Terminology



instance: <input, output>

<tumorsize, texture, perimeter> = <18.2, 27.6, 117.5> => cancer = No

Terminology

(Labelled) **Dataset**: many training examples or instances

<tumorsize,	texture,	perimeter>	,	<cancer,	size change>
<18.2,	27.6,	117.5>	,	< No ,	+2 >
<17.9,	10.3,	122.8>	,	< No ,	-4 >
<20.2,	14.3,	111.2>	,	< Yes ,	+3 >
<15.5,	15.2,	135.5>	,	< No ,	0 >

.

.

.

Terminology

We split the dataset into

Train dataset

<tumorsize,	texture,	perimeter>	,	<cancer,	size change>
<18.2,	27.6,	117.5>	,	< No ,	+2 >
<17.9,	10.3,	122.8>	,	< No ,	-4 >
<20.2,	14.3,	111.2>	,	< Yes ,	+3 >
<15.5,	15.2,	135.5>	,	< No ,	0 >

.
.
.

used to build the model

Test dataset

<tumorsize,	texture,	perimeter>	,	<cancer,	size change>
<12.4,	15.7,	120.1>	,	< No ,	+5 >
<15.2,	17.2,	113.3>	,	< Yes ,	+1 >
<19.3,	15.9,	125.4>	,	< No ,	+2 >
<17.5,	11.9,	122.7>	,	< No ,	-3 >

.
.
.

used to evaluate the model

algorithm shouldn't have access to this set when being trained

Terminology

Evaluation of performance

Test dataset

Ground-Truth

Input features	True labels	Output labels
<tumorsize, texture, perimeter>	<cancer, size change>	<cancer, size change>
<12.4, 15.7, 120.1>	< No , +5 >	< Yes , +4 >
<15.2, 17.2, 113.3>	< Yes , +1 >	< Yes , +1 >
<19.3, 15.9, 125.4>	< No , +2 >	< No , +1 >
<17.5, 11.9, 122.7>	< No , -3 >	< Yes , -2 >
.		.
.		.
.		.

algorithm shouldn't see the true labels when being evaluated, these are only used to compare against the results to measure performance

"Types" of Machine Learning

Supervised learning

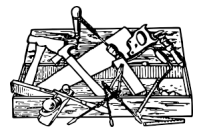
Unsupervised learning

Semi-supervised learning

Reinforcement learning

...

"Types" of Machine Learning

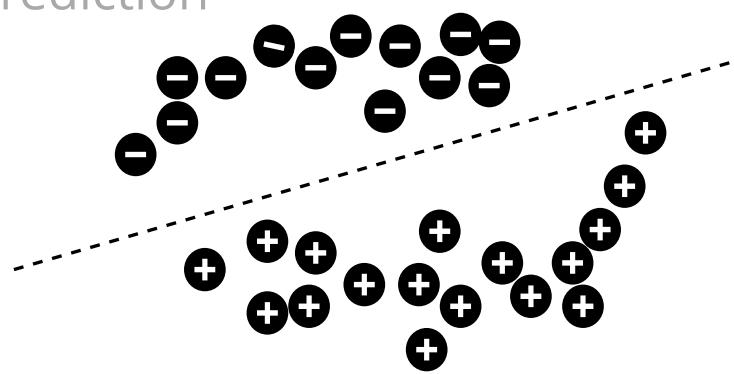


Supervised Learning: we have labeled data

- classification
- regression
- structured prediction

most of this course!

give desired output to the algorithm to learn from it



Supervised Learning

<tumorsize, texture, perimeter> , <cancer, size change>

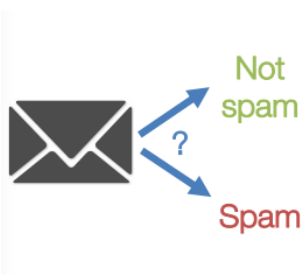
<18.2, 27.6, 117.5> , < No , +2 > ⊖

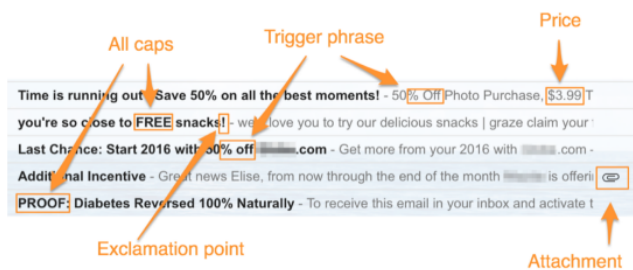
<17.9, 10.3, 122.8> , < No , -4 > ⊖

<20.2, 14.3, 111.2> , < Yes , +3 > ⊕

<15.5, 15.2, 135.5> , < No , 0 > ⊖

.





Example image from Hamilton 551 slides

Classification vs Regression

<tumorsize, texture, perimeter>	,	<cancer>
<18.2, 27.6, 117.5>	,	< No >
<17.9, 10.3, 122.8>	,	< No >
<20.2, 14.3, 111.2>	,	< Yes >
<15.5, 15.2, 135.5>	,	< No >

Classification

categorical output

discrete

<tumorsize, texture, perimeter>	,	<size change>
<18.2, 27.6, 117.5>	,	< +2 >
<17.9, 10.3, 122.8>	,	< -4 >
<20.2, 14.3, 111.2>	,	< +3 >
<15.5, 15.2, 135.5>	,	< 0 >

Regression

numerical output

continuous

Supervised Learning: Example

MIT Technology Review

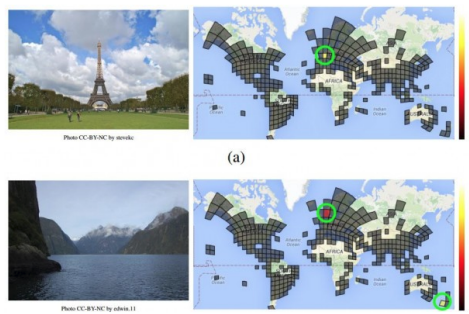
Topics+ The Download

Intelligent Machines

Google Unveils Neural Network with "Superhuman" Ability to Determine the Location of Almost Any Image

Guessing the location of a randomly chosen Street View image is hard, even for well-traveled humans. But Google's latest artificial-intelligence machine manages it with relative ease.

by Emerging Technology from the arXiv February 24, 2016



trained on a database of geolocated images from the Web

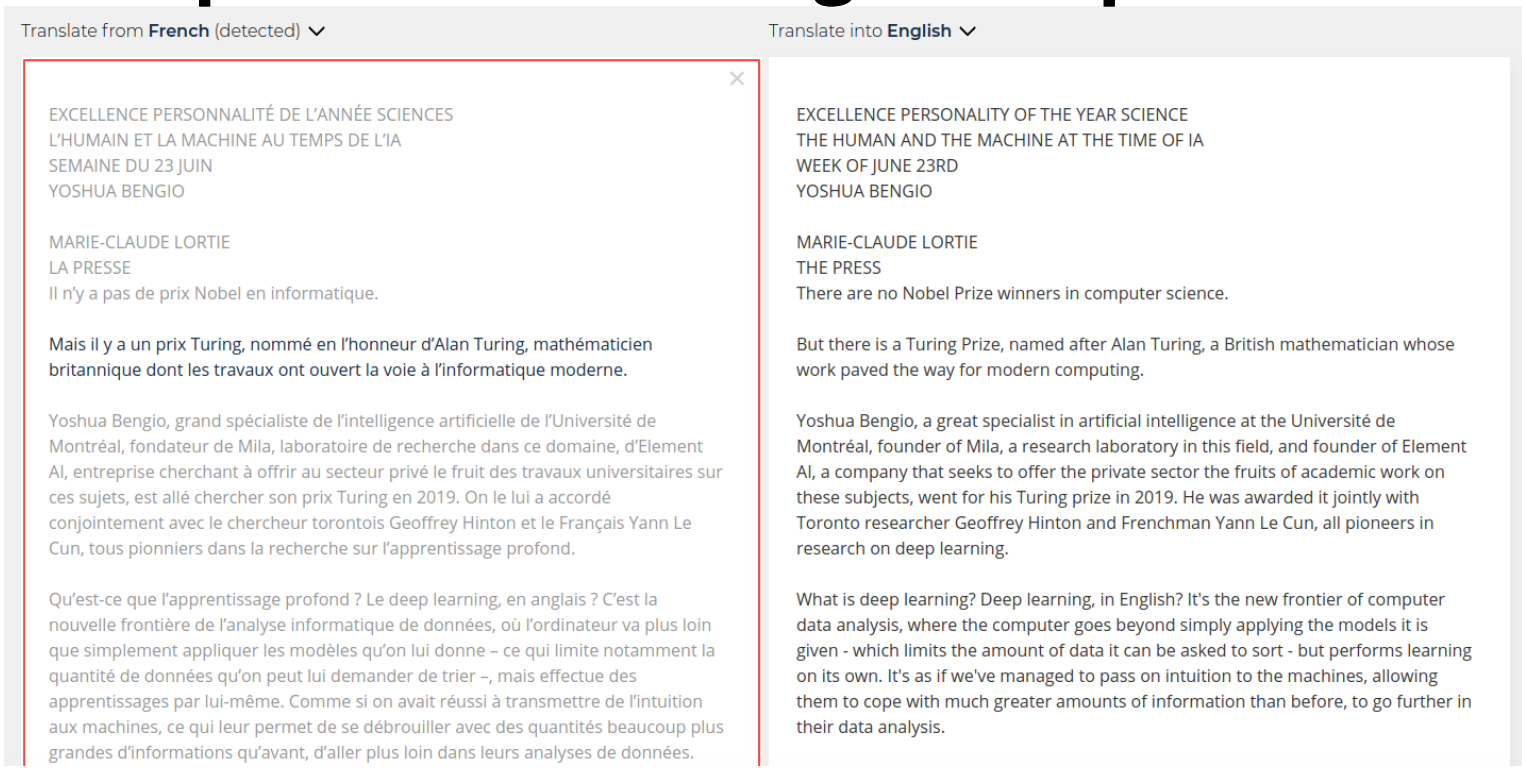
Supervised Learning: Example

Machine Translation

trained using

bilingual sentences collected by Linguee's web crawler on the Internet

Supervised Learning: Example



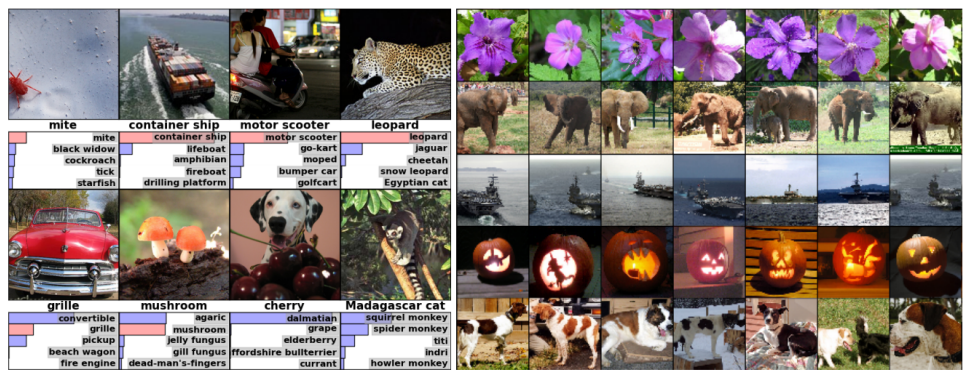
Large Labelled Datasets

Supervised methods are powered by large datasets often crawled from the web or curated with crowdsourcing, e.g.

ImageNet => AlexNet

bilingual sentences collected by Linguee's web **crawler** on the Internet

a database of geolocated images from the **Web**



Supervised Learning: Example Dataset

What is COCO?

COCO is a large-scale object detection, segmentation, and captioning dataset. COCO has several features:

- Object segmentation
- Recognition in context
- Superpixel stuff segmentation
- 330K images (>200K labeled)
- 1.5 million object instances
- 80 object categories
- 91 stuff categories
- 5 captions per image
- 250,000 people with keypoints

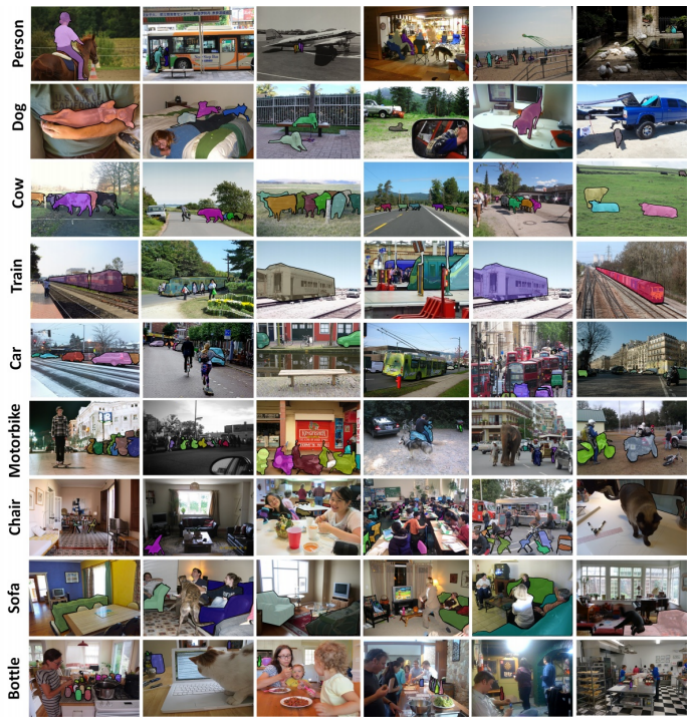


Fig. 6: Samples of annotated images in the MS COCO dataset.

Object Recognition

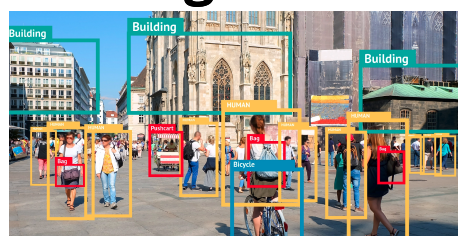


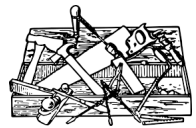
image: https://bitmovin.com/object-detection/

Supervised Learning: Example

Image Captioning

from: https://arxiv.org/pdf/1609.06647.pdf

"Types" of Machine Learning

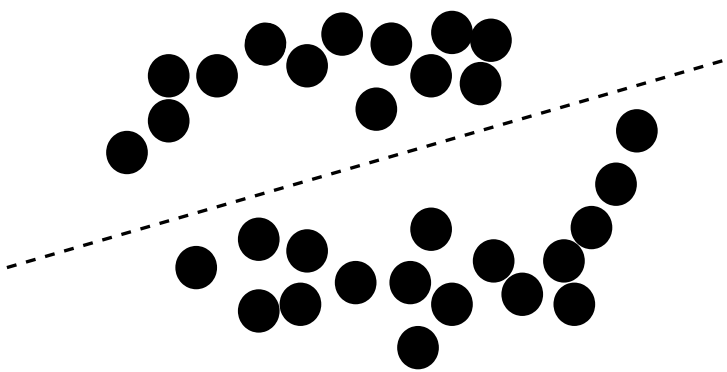


Unsupervised Learning: only unlabeled data

- clustering
- dimensionality reduction
- density estimation / generative modeling
- anomaly detection
- discovering latent factors and structures

- helps explore and understand the data
- closer to data mining
- we have much more unlabeled data
- more open challeges

The algorithm doesn't see the desired outputs, mines the patterns in the input data



Unsupervised Learning

<tumorsize, texture, perimeter>

<18.2, 27.6, 117.5>

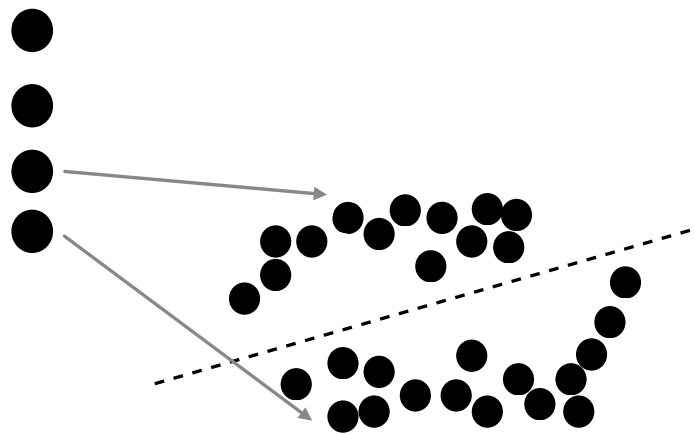
<17.9, 10.3, 122.8>

<20.2, 14.3, 111.2>

<15.5, 15.2, 135.5>

.

.

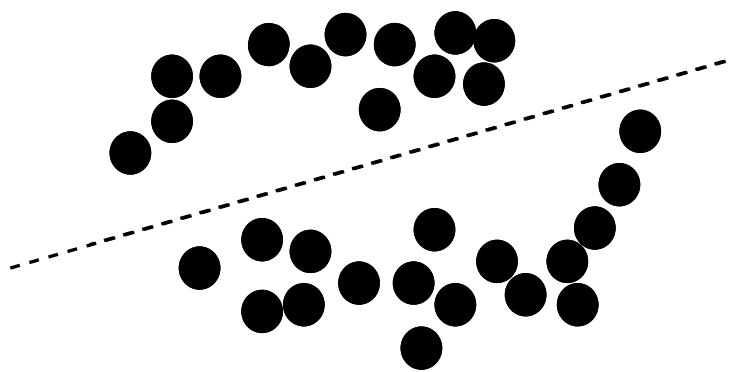


there seems to be two types of tumors, largers and rough vs smaller and smooth

Clustering

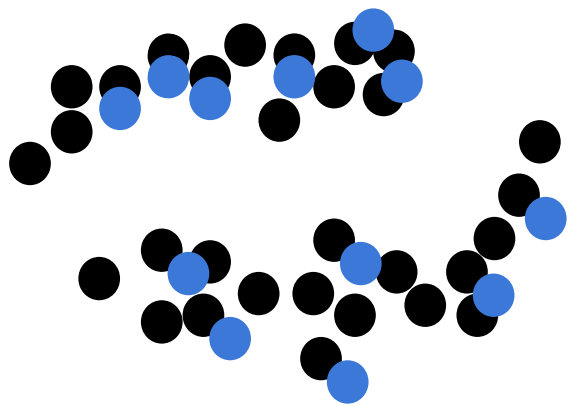
Similar to classification but labels/classes should be inferred and are not given to the algorithm

<tumorsize,	texture,	perimeter>
<18.2,	27.6,	117.5>
<17.9,	10.3,	122.8>
<20.2,	14.3,	111.2>
<15.5,	15.2,	135.5>



Generative Modelling

Models the distribution of the data and learns to generate the data instead of directly categorizing/discriminating the instances into different classes

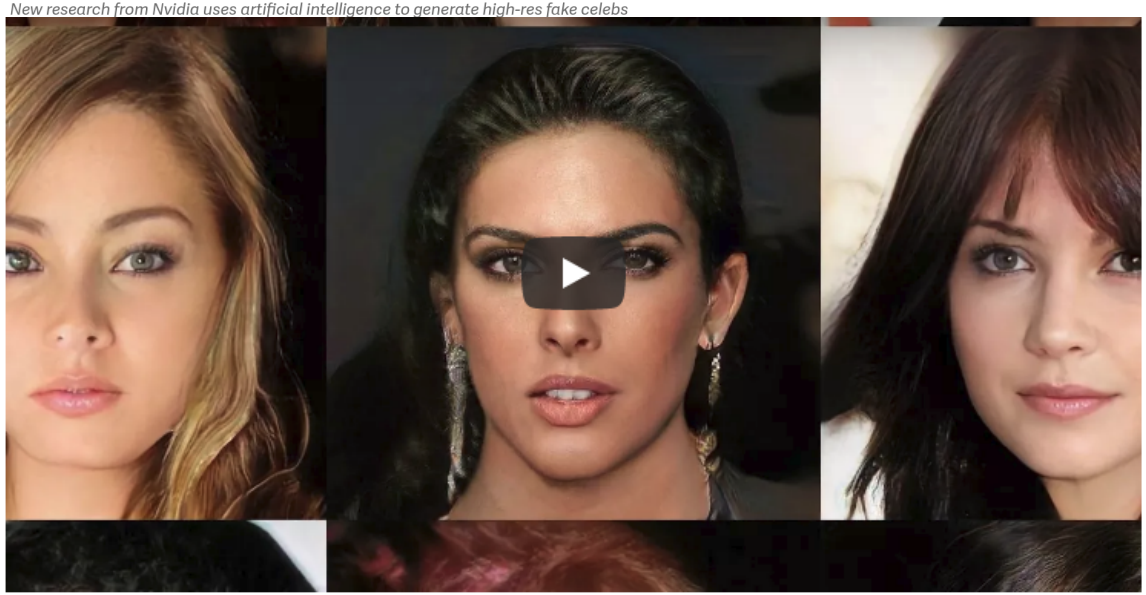


Unsupervised Learning: Example

TECH ARTIFICIAL INTELLIGENCE

All of these faces are fake celebrities spawned by AI

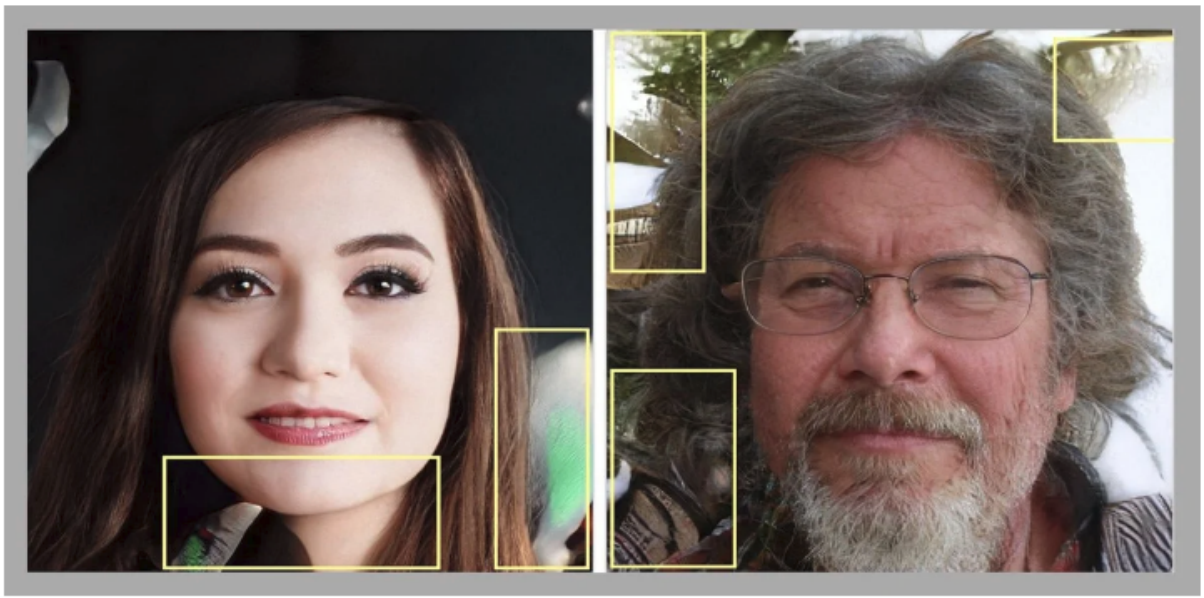
New research from Nvidia uses artificial intelligence to generate high-res fake celebs



PARIS MARTINEAU BUSINESS 12.20.2019 08:21 PM

Facebook Removes Accounts With AI-Generated Profile Photos

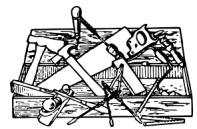
Researchers said it appears to be the first use of artificial intelligence to support an inauthentic social media campaign.



Profile pictures for Facebook accounts "Mary Keen" and "Jacobs Guillermo," admins on groups associated with The BL highlighted by Graphika. COURTESY OF GRAPHIKA

Ethical Challenges: Misuse

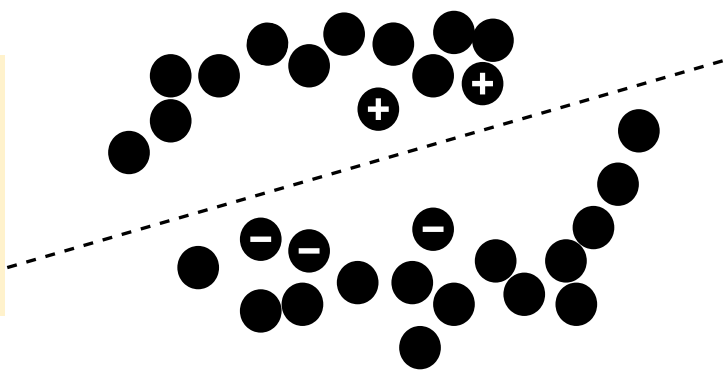
"Types" of Machine Learning



Semisupervised learning: a few labeled examples

- we can include structured problems such as
  - matrix completion (a few entries are observed)
  - link prediction

The algorithm sees few examples of the desired outputs



Netflix Awards $1 Million Prize and Starts a New Contest

BY STEVE LOHR SEPTEMBER 21, 2009 10:15 AM

Jason Kempin/Getty Images Netflix prize winners, f
Chabbert, Martin Piotte, Michael Jahrer, Andreas Tos

Update | 1:45 p.m. *Adding details announced close finish to the contest.*

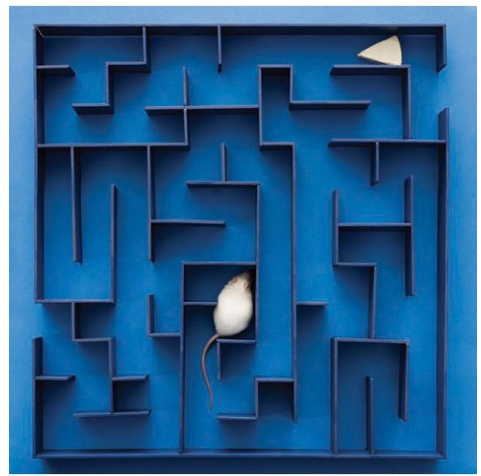
RYAN SINGEL SECURITY 03.12.10 02:48 PM

NetFlix Cancels Recommendation Contest After Privacy Lawsuit

Netflix is canceling its second $1 million Netflix Prize to settle a legal challenge that it breached customer privacy as part of the first contest's race for a better movie-recommendation engine.

Ethical Challenges: Privacy of Users

"Types" of Machine Learning



Reinforcement Learning:

- weak supervision through the reward signal
- sequential decision making
- biologically motivated

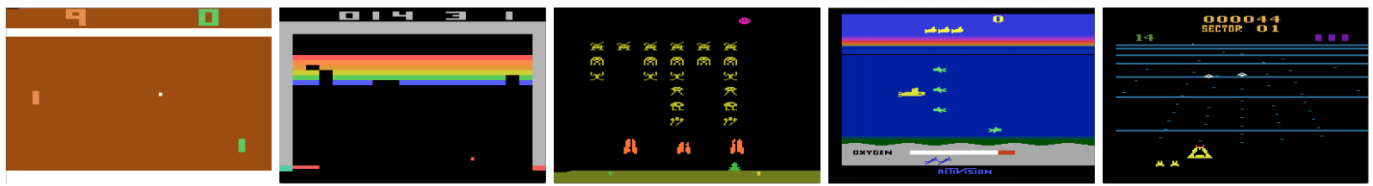
also related:

imitation learning: learning from demonstrations

- behavior cloning (is supervised learning!)
- inverse reinforcement learning (learning the reward function)

Reinforcement Learning: Example

Playing Atari like a pro 2015



Human Level Control Through Deep Reinforcement Learning

Abstract

The theory of reinforcement learning provides a normative account deeply rooted in psychological and neuroscientific perspectives on animal behaviour, of how agents may optimize their control of an environment. To use reinforcement learning successfully in situations approaching real-world complexity, however, agents are confronted with a difficult task: they must derive efficient representations of the environment from high-dimensional sensory inputs, and use these to generalize

Reinforcement Learning: Example

Google's AlphaGo Defeats Chinese Go Master in Win for A.I.

点击查看本文中文版

By **PAUL MOZUR** MAY 23, 2017

Ke Jie, the world's top Go player, reacting during his match on Tuesday against AlphaGo, artificial intelligence software developed by a Google affiliate. China Stringer Network, via Reuters

RELATED COVERAGE


A.I. Is Doing Legal Work. But It Won't Replace Lawyers, Yet. MARCH 19, 2017


China's Intelligent Weaponry Gets Smarter FEB. 3, 2017


THE FUTURE OF WORK
The Future of Not Working FEB. 23, 2017


Master of Go Board Game Is Walloped by Google Computer Program MARCH 9, 2016

Generalization Challenge

A model trained to play Go can not perform if we change a few rules

Ockham's Razor principle:

> *Entities should not be multiplied without necessity.*

The less complex model {between those performing similar on observed data}, the more likely that the model will generalize {to unseen data from the same distribution}

Pedro Domingos:

> *Simplicity does not imply accuracy but simplicity is a virtue in its own right*

source

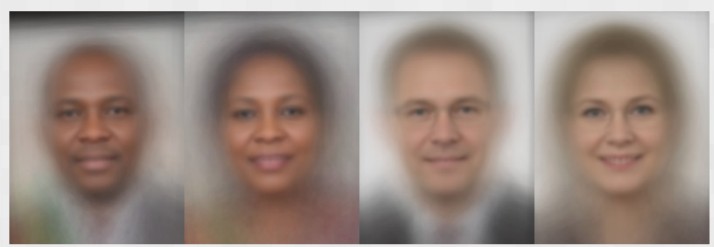
Generalization Challenge

The model learns from the distribution of the input data

the demographic and phenotypic composition of training and benchmark datasets are important

Face-recognition software is perfect – if you're a white man

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%



CHANDAN KHANNA/AFP/Getty Images

Bias and Fairness Challenge

The model learns from the distribution of the input data and its biases

Bias and Fairness Challenge

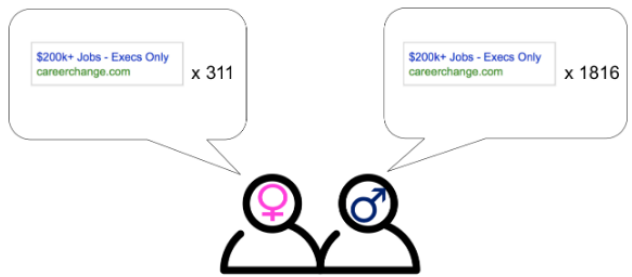
Growing use, growing concerns

The algorithm decides to show high-income jobs to men much more often than to women.

Google's online ad algorithm

The algorithm decides not to invite women to interview.

Amazon's hiring algorithm



BUSINESS NEWS OCTOBER 9, 2018 / 11:12 PM / A YEAR AGO

Amazon scraps secret AI recruiting tool that showed bias against women

Jeffrey Dastin 8 MIN READ

SAN FRANCISCO (Reuters) - Amazon.com Inc's (AMZN.O) machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

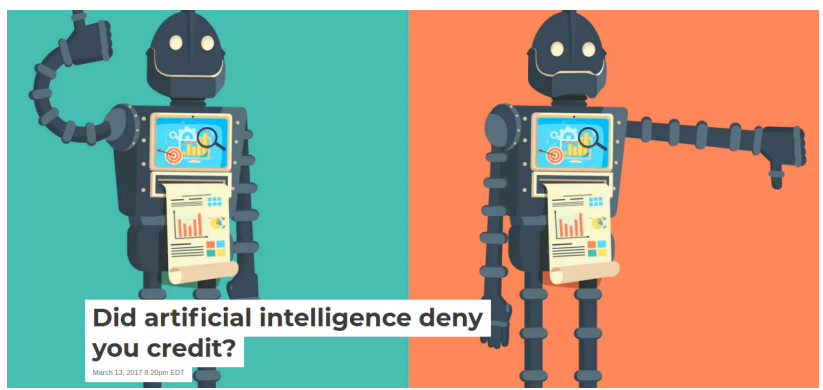
Growing use, growing concern

Inequality

Rise of the racist robots – how AI is learning all our worst impulses

There is a saying in computer science: garbage in, garbage out. When we feed machines data that reflects our prejudices, they mimic them – from antisemitic chatbots to racially biased software. Does a horrifying future await people forced to live at the mercy of algorithms?

Ethical Challenges: Bias



Growing use, growing concern

WIRED BUSINESS CULTURE GEAR IDEAS MORE SIGN IN SUBSCRIBE

TOM SIMONITE BUSINESS 10.24.2019 02:00 PM

A Health Care Algorithm Offered Less Care to Black Patients

A study shows the risks of making decisions using data that reflects inequities in American society.

nature

UPDATE 26 OCTOBER 2019

Millions of black people affected by racial bias in health-care algorithms

Study reveals rampant racism in decision-making software used by US hospitals – and highlights ways to correct it.

Heidi Ledford

PDF version

RELATED ARTICLES

A fairer way forward for AI in health care

Bias detectives: the researchers striving to make algorithms fair

Can we open the black box of AI?

SUBJECTS

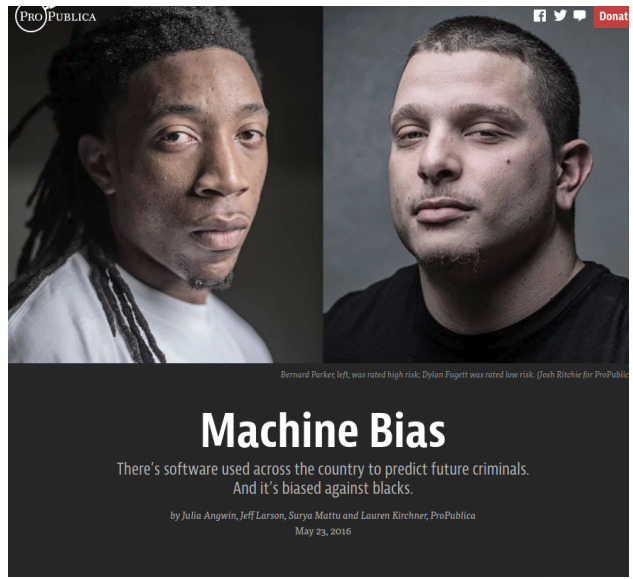
Computer science Health care Policy Society

Black people with complex medical needs were less likely than equally ill white people to be referred to programmes that provide more personalized care. Credit: Ed Kashi/VII/Redux/eyevine

An algorithm widely used in US hospitals to allocate health care to patients has been systematically discriminating against

Health

Freedom



PRO PUBLICA Donat

Bernard Parker, left, was rated high risk; Dylan Fugett was rated low risk. (Josh Ritchie for ProPublica

Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica

May 23, 2016

Ethics in ML

Be informed and mindful when using ML in practice.

Accuracy should not be our only goal.

Summary

Supervised Learning: we have labeled data

- classification
- regression

Unsupervised Learning: only unlabeled data

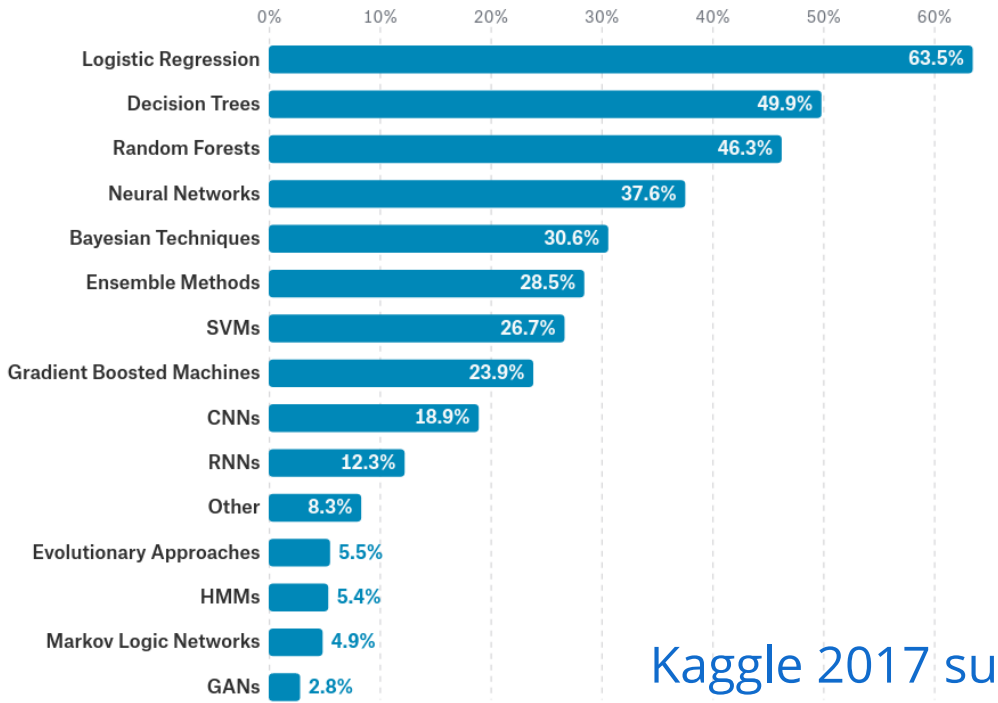
- clustering
- density estimation / generative modeling

Semisupervised learning: a few labeled examples

Reinforcement Learning: reward signal

Challenges: privacy, generalization, bias, misuse

What is used the most in practice?



Kaggle 2017 survey