Approximating Intelligence:

An introduction to Machine Learning driven Artificial Intelligence.

Spandan Madan,
SEAS Fellow @ Harvard
Research Assistant @ MIT
spandan_madan@g.harvard.edu
smadan@mit.edu
Github - https://goo.gl/zyycR8

Overview

- AI v/s Machine Learning v/s Deep Learning
- Why is machine learning used in Artificial Intelligence?
- Popular paradigms in machine learning, and challenges in implementing them.
- Deep learning adresses (one) such challenge very well
- The intuition behind popular Deep Learning models.
- What can they really achieve? What can they NOT achieve?

What is Artificial Intelligence?

- Translating languages?
- Detecting images of cats online?
- Writing poetry?
- Generating music?

Do these “tasks” define artificial intelligence?

The goal is to build agents that have the same *characteristics* as an intelligent being, and can do *similar* tasks.

Capabilities

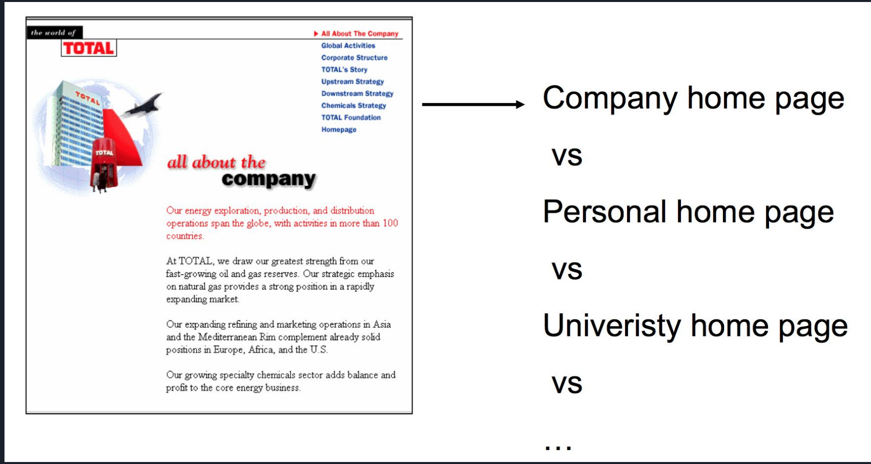
1. Recognizing a known thing (cockroach)
2. Communicating
3. Transporting a terrain
4. Planning how to stack
5. Recognizing an unknown thing (undergarments)

What is Machine Learning?

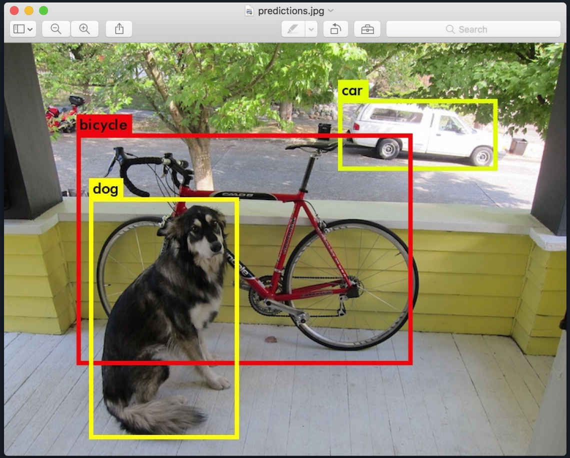
Making models that

- **<u>improve</u>** their performance
- at some **<u>specific task</u>**
- with **<u>experience</u>**

Example Task 1 :



Example Task 2 :



ML is a “toolkit” often used for AI

Analogy-

- AI : The goal of building architecture
- ML: The toolkit used (ex. Hammer, carpenter’s square, screwdriver)
- Deep Learning: A specific Tool (ex. Hammer)

If we define AI as the capability to do certain tasks, we can use a Machine Learning algorithm to learn to perform these tasks!

However, the tools may be used for other purposes (including non-human activities):

- Interactions between proteins and drugs
- Predicting how light reflects based on a material’s properties.

Machine Learning

- Consider 2 entities (ex: words on a website and kind of website).
- **<u>Assumption</u>** : There is a relationship between the entities.
- **<u>Problem</u>** : We don't know it!
- **<u>Solution</u>** : Approximate is using Machine Learning!

<u>Mathematically</u> :- Entities are X (input) and Y (output)

- **<u>Assumption</u>** : There is a function f which maps X to Y.
- **<u>Problem</u>** : We don't know f(X).
- **<u>Solution</u>** : Find best function g(X), which approximates f.

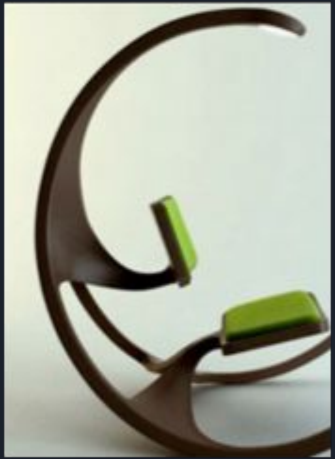
Challenge 1: help a robot recognize a chair

Pics: Zoya Bylinskii

But… aren't these also chairs? :/

- Shape definitely doesn't seem to work

- Maybe, it's a surface on the ground that we sit on?



Pics: Zoya Bylinskii

Surface on the ground used to sit?

Naah, this does doesn’t work either – None of these are chairs!

Pics: Zoya Bylinskii

Different kinds of learning paradigms (inspired by human learning)

How would you learn to play the guitar?

1. From a **teacher**
2. By playing yourself, **exploring** to see what sounds good.
3. By listening to more music, **identifying patterns** that sound good. Using the patterns later to learn faster.

Popular paradigms in ML.

Learning through example of teacher: Supervised learning

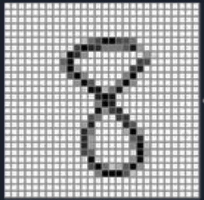
Learning by exploring to get reward: Reinforcement learning

Learning by observing patterns: Unsupervised learning

Supervised Learning
(Ex. Reading handwritten digits)

We have pairs of images and labels.

We want to learn a model which takes in an image, and tells what digit is written in the image



28 x 28
784 pixels

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	4	62	146	182	254	254	181	176	139	15	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	34	186	253	217	208	136	136	136	166	232	99	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	61	242	208	111	3	0	0	0	0	0	18	32	107	43	0	0	0	0	0	0
0	0	0	0	0	0	0	0	156	242	23	0	0	0	0	0	0	0	13	191	181	6	0	0	0	0	0	0
0	0	0	0	0	0	0	0	121	255	98	3	0	0	0	0	0	8	194	225	12	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	169	253	120	3	0	0	0	0	128	247	51	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	3	111	244	169	19	0	14	131	249	117	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	59	241	235	72	142	229	66	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	25	218	254	231	36	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	133	253	221	33	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	19	237	111	196	217	19	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	174	138	0	23	193	204	18	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	96	224	0	0	0	25	218	169	3	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	215	138	0	0	0	0	86	253	99	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	215	97	0	0	0	0	3	162	214	11	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	215	97	0	0	0	0	0	118	253	68	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	185	157	0	0	0	0	0	40	254	98	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	50	244	61	0	0	0	0	112	244	58	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	174	251	142	59	83	167	244	111	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	6	133	253	253	253	169	61	3	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

ML Model

8

Reinforcement Learning
(Learning to play flappy bird)

Given status quo, take action

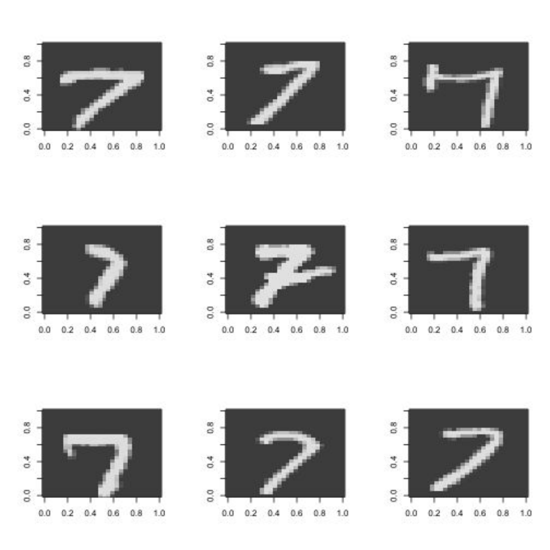
No example pairs given. But, they are accessible by exploring. For ex. Playing.

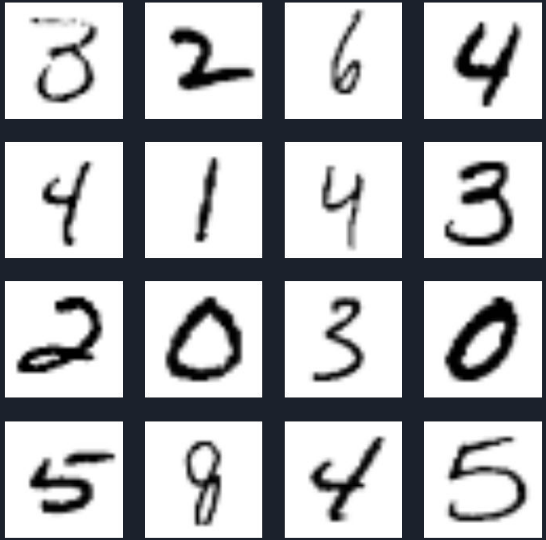
However, can't always play!
For ex: mars rover navigation?

Goal: Given a position of your bird and trees, deciding when to jump.

Demo!

Unsupervised Learning
(Identifying images of the same digit)

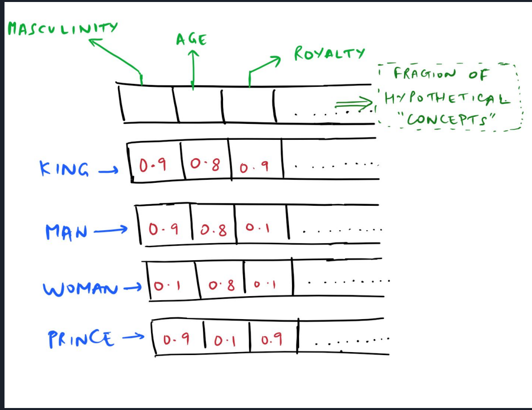




Again, no pairs. Not even a way to get some samples.

We have given ONLY images, NO labels.

Another example - quantifying similarity of words



similarity ('tremendous' , 'enormous') = 0.75

similarity('tremendous' , 'neglegible') = 0.38

Most similar words to 'king' -> queen, prince, royal.

Guess the approach!
(Supervised v/s Unsupervised v/s Reinforcement Learning)

Vision:

- Detecting cars?
- Identifying emotions on faces?
- Recognizing faces of important people automatically in videos?

Robotics:

- walking on new terrain
- grasping bottles of different kinds

Some other tasks the AI agent might need

Planning:

- Playing a game of tetris?
- Wall-E planning how to stack garbage together?

NLP:

- Predicting positive/negative review of a movie.
- Automatically generating headline of a news article.

Speech:

- understanding dialects languages

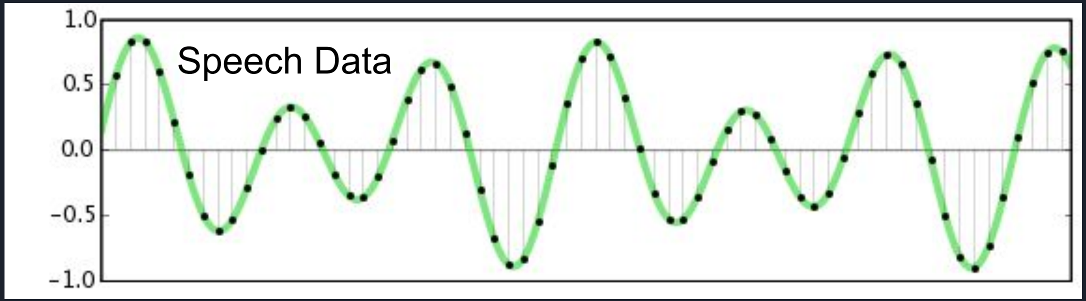
So, how do we feed in all this kind of information?

Machine learning is a "toolkit" often used for AI

v/s

77 97 99 104 105 110 101 32 108 101 97 114 110 105 110 103 32 105 115 32 97 32 34 116 111 111 108 107 105 116 34 32 111 102 116 101 110 32 117 115 101 100 32 102 111 114 32 65 73

0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0



If we used raw data directly…

The right representation is the key.
(And it's a very hard thing to find)

Conventionally - researchers would sit down and find interesting ways to represent a “cat”.

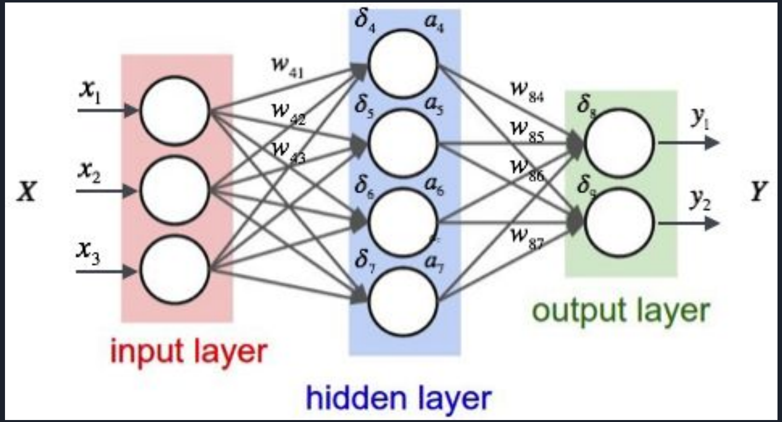
For ex: shape? Shape of ears/eyes? Stance?

Next phase was identifying “important points” in an image, and comparing them across images.

If only we could automate this....

That’s what deep learning does!

It (intelligently) tries out a bunch of mathematical operations on an image. Ultimately it learns a series of operations such that in their output, semantically similar images have similar representations

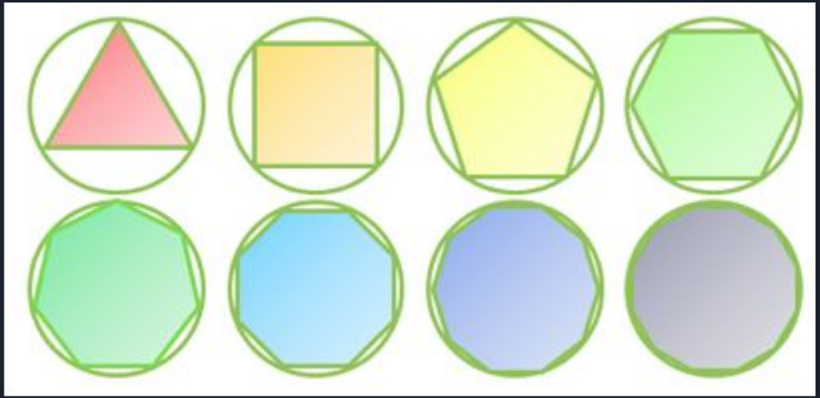


0.3

0.9

0.85

Deep Learning

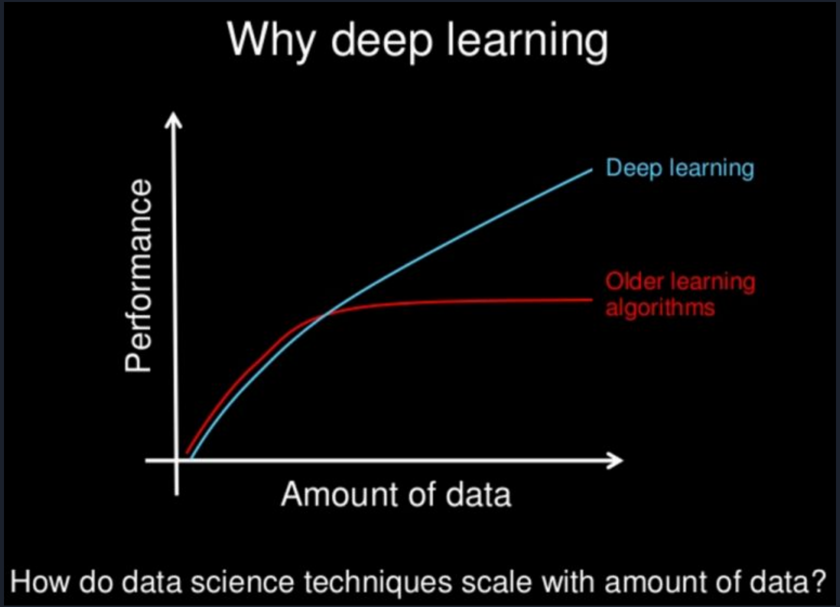


Approximating complicated relationships by carefully combining very simple relationships!

Each layer = simple computations. Chain = can approximate any function [Theorem]

“Deep” because lots of layers. Conventionally NN’s had only 1 layer.

Old techniques, revived due to powerful computers.



- Techniques been around since 90s.
- Requires LOTS of computations, and picking and choosing which series of computations is useful.
- But computers were not strong enough to do enough computations in reasonable time.

Two popular class of deep nets - CNNs and RNNs

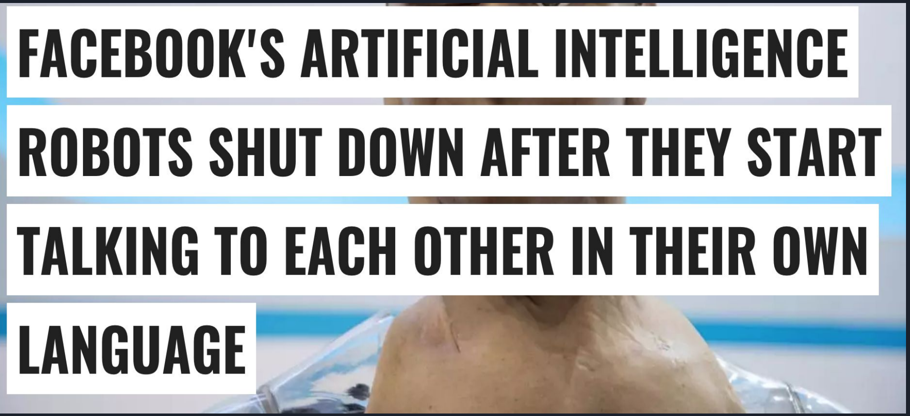
Convolutional Neural Networks (CNNs) -

- Learn the right curves/edges.
- Learn the right order to put them together in to make an object.

Recurrent Neural Networks (RNNs) -

- Learn what chunks repeat in time (syllables for ex)
- Learn what order to put them together in to make speech.

Media blowing things out of proportion sometimes...



Watch a Human Try to Fight Off Boston Dynamics' Door-Opening ...
https://www.wired.com/.../watch-a-human-try-to-fight-off-that-door-opening-robot-d...
Feb 20, 2018 - Hey, remember that **dog**-like robot, SpotMini, that **Boston Dynamics** showed off last week, the one that opened a **door** for its robot friend? Well, the company just dropped a new video starring the canine contraption. In this week's episode, a human with a hockey stick does everything in his power to stop the ...

Twitter taught Microsoft's AI chatbot to be a racist in less than ...
https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist
Mar 24, 2016 - The more you chat with Tay, said **Microsoft**, the smarter it gets, learning to engage people through "casual and playful conversation." Unfortunately, the conversations didn't stay playful for long. Pretty soon after Tay launched, people starting tweeting the **bot** with all sorts of misogynistic, **racist**, and Donald ...

Testing Robustness

Artificial General Intelligence

AI v/s AGI : Sometimes referred to as Narrow v/s Strong AI.

Narrow AI :

- Machine intelligence that equals or exceeds human intelligence, BUT
- **Task specific, narrow scope**
- Human input necessary at multiple stages.

Strong AI/ AGI:

- Broad spectrum capabilities.
- Still very much in the science fiction domain.
- A conscious being capable of cross-domain complex tasks.

Some necessary requirements for AGI

- reason, use strategy, solve puzzles, and make judgments under uncertainty;
- represent knowledge, including commonsense knowledge;
- plan;
- learn;
- communicate in natural language;
- and integrate all these skills towards common goals.

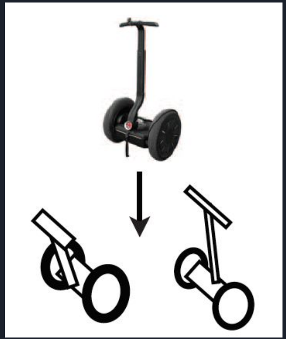
Some shortcomings of existing deep learning systems

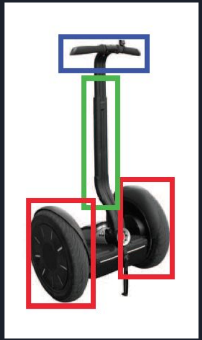
Ideas mostly borrowed from Josh Tenenbaum, gary marcus and Yann LeCun

We learn richer information with much lesser examples

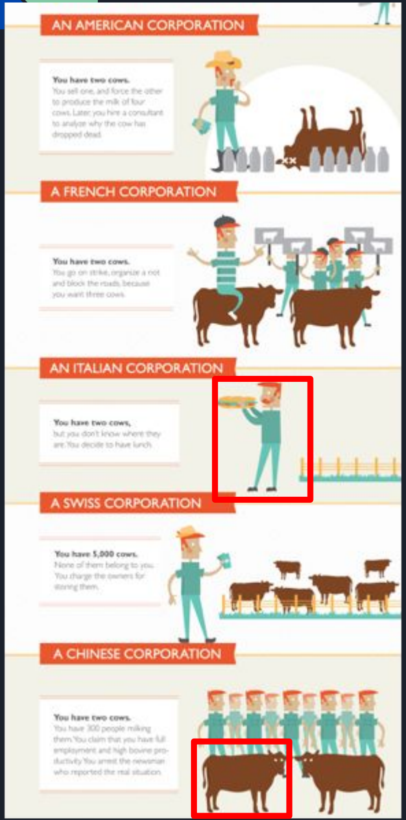
1. **Data** Hungry: We can learn to recognize this segway with just 1-2 samples.

2. We learn knowledge (representations) that can be reused much more:
   i. Parse the parts
   ii. Create new examples
   iii. Creating abstractions of the concept





Learned solutions don't easily transfer to new problems



Problems often share sub-structure, i.e. they are related.

For ex: Knowing that you're looking at a key helps you detect a key in a picture.

Networks can't transfer learned information across problems easily.

Black-box, non-interpretable solutions

Problems:

- Can lead to ethical constraints: Loan rejection bias?
- Hard to know when it will work, or why it isn’t working?

No notion of causality

Correlations (patterns) are not causation!

An understanding of the reasons of these scenes help us understand them better. Existing systems can only understand objects and their spatial relationships.

a woman riding a horse on a dirt road

an airplane is parked on the tarmac at an airport

a group of people standing on top of a beach

Some ongoing research to address these points (very narrow)

- Josh Tenenbaum (MIT) talks about need to incorporate the following -
    - Causality: An understanding of the underlying “story”
    - Compositionality: The capability to break into simpler tasks
    - Learning to Learn: The capability to learn and reuse these simpler tasks.

Summary

- Machine Learning is toolkit that is often used to solve tasks that we believe an artificially intelligent agent should be able to do.
- ML is more ubiquitous, than revolutionary!
- As we saw, task as simple as detecting a chair can’t be described in terms of simple rules, ML helps out there.
- The popular paradigms in ML are inspired by human learning (from teacher/exploration/observation)
- A big challenge in implementing these models is the right representation for input. That’s where deep learning comes in.
- Existing systems are great at identifying patterns in space (CNN) and time (RNN).
- There are some major shortcomings, but work on them is evolving fast.

Simple self-checks

- If you’re using the word “data” too often…

- Every time you hear the phrase “Some one built an AI that does something”, re-read it in your mind as “someone made an ML model that …”

- Every time you read too polarizing an article on AI (utopian or dystopian), take a moment to read the original source. Chances are high that it’s not as serious as portrayed.

Good Resources to Get started with coding

Tiny demo if you just want intuition - https://teachablemachine.withgoogle.com/

My In depth tutorial on Machine learning and Deep Learning - https://goo.gl/mube7u

Specific Coding Tutorials -

Image Classification - https://goo.gl/E2hR4i

General introduction to PyTorch - https://github.com/jcjohnson/pytorch-examples

Text Classification - https://github.com/Shawn1993/cnn-text-classification-pytorch

Sentiment analysis - https://einstein.ai/research/learning-when-to-skim-and-when-to-read

More Resources

A team project - https://goo.gl/3PpZ7c

A good video for neural network's intuition - https://www.youtube.com/watch?v=aircAruvnKk

Thank you!

Slides - https://goo.gl/Fjp9oB

Me - https://goo.gl/zyycR8