Introduction to Machine Learning

Master in Data Science and Advanced Analytics
BA and DS majors

Roberto Henriques

A Few Quotes

- "*Artificial intelligence will reach human levels by around 2029. Follow that out further to, say, 2045, we will have multiplied the intelligence, the human biological machine intelligence of our civilization a billion-fold.*" **Ray Kurzweil, Inventor, entrepreneur and visionary**
- "*Machine intelligence is the last invention that humanity will ever need to make*." **Nick Bostrom, philosopher**
- *"Much of what we do with machine learning happens beneath the surface. Machine learning drives our algorithms for demand forecasting, product search ranking, product and deals recommendations, merchandising placements, fraud detection, translations, and much more. Though less visible, much of the impact of machine learning will be of this type — quietly but meaningfully improving core operations."* **Jeff Bezos, Amazon CEO**

A Few Quotes

- *“Everything that civilization has to offer is a product of human intelligence; we cannot predict what we might achieve when this intelligence is magnified by the tools that AI may provide, but the eradication of war, disease, and poverty would be high on anyone’s list. Success in creating AI would be the biggest event in human history. Unfortunately, it might also be the last.”* **Stephen Hawking, theoretical physicist, cosmologist**

- *“In the long term, artificial intelligence and automation are going to be taking over so much of what gives humans a feeling of purpose.”* **Matt Bellamy, Muse rock band.**

A Few Quotes

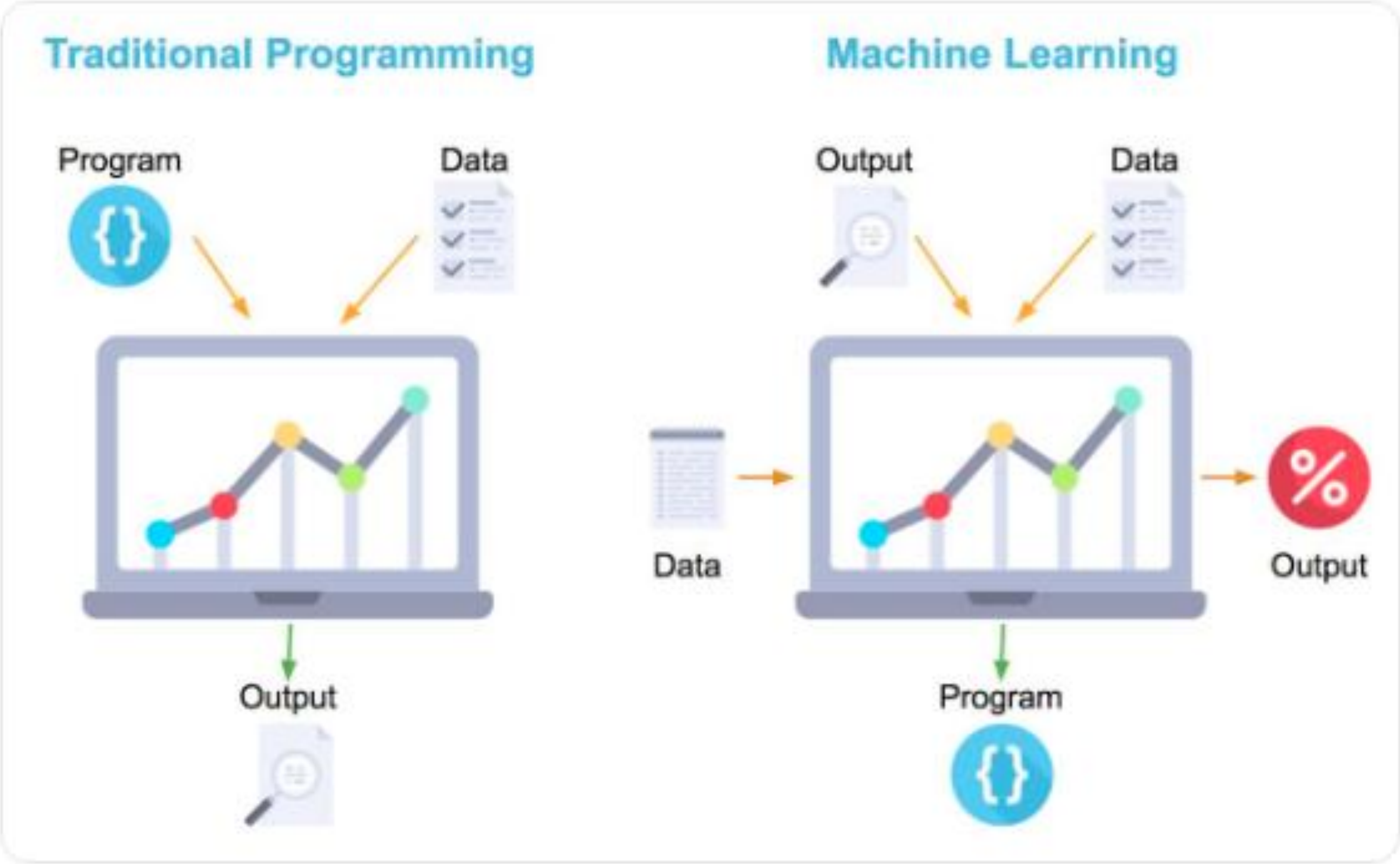
- "*Harnessing machine learning can be transformational, but for it to be successful, enterprises need leadership from the top. This means understanding that when machine learning changes one part of the business — the product mix, for example — then other parts must also change. This can include everything from marketing and production to supply chain, and even hiring and incentive systems*." **Erik Brynjolfsson, Director of the MIT Initiative on the Digital Economy**

- *"I am telling you, the world's first trillionaires are going to come from somebody who masters AI and all its derivatives, and applies it in ways we never thought of."* **Mark Cuban, American entrepreneur and investor. NBA Dallas Mavericks**

So, What Is Machine Learning?

- Inaccurate definition?
- Automation of automation
- Getting computers to program themselves
- Computer scientists → Machine learning?
- Data plays major role→ worker

New paradigm

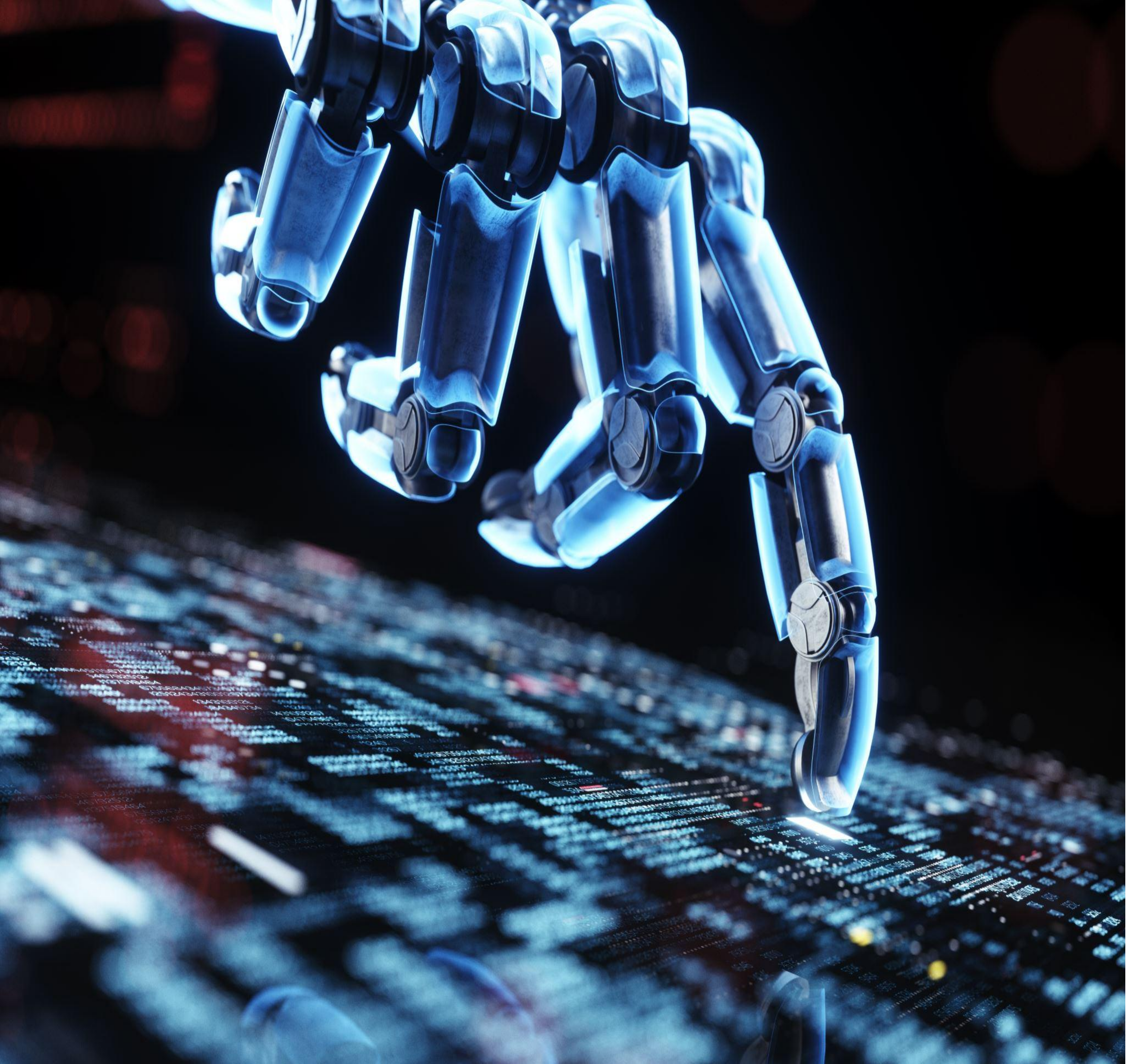


https://towardsdatascience.com/machine-learning-vs-traditional-programming-c066e39b5b17

How?

- **ML versus Gardening**

- Seeds → Algorithms
- Nutrients → Data
- Gardener → You
- Plants → Programs



Applications examples

- Web search
- Computational biology
- Finance
- E-commerce
- Space exploration
- Robotics
- Information extraction
- Social networks
- …

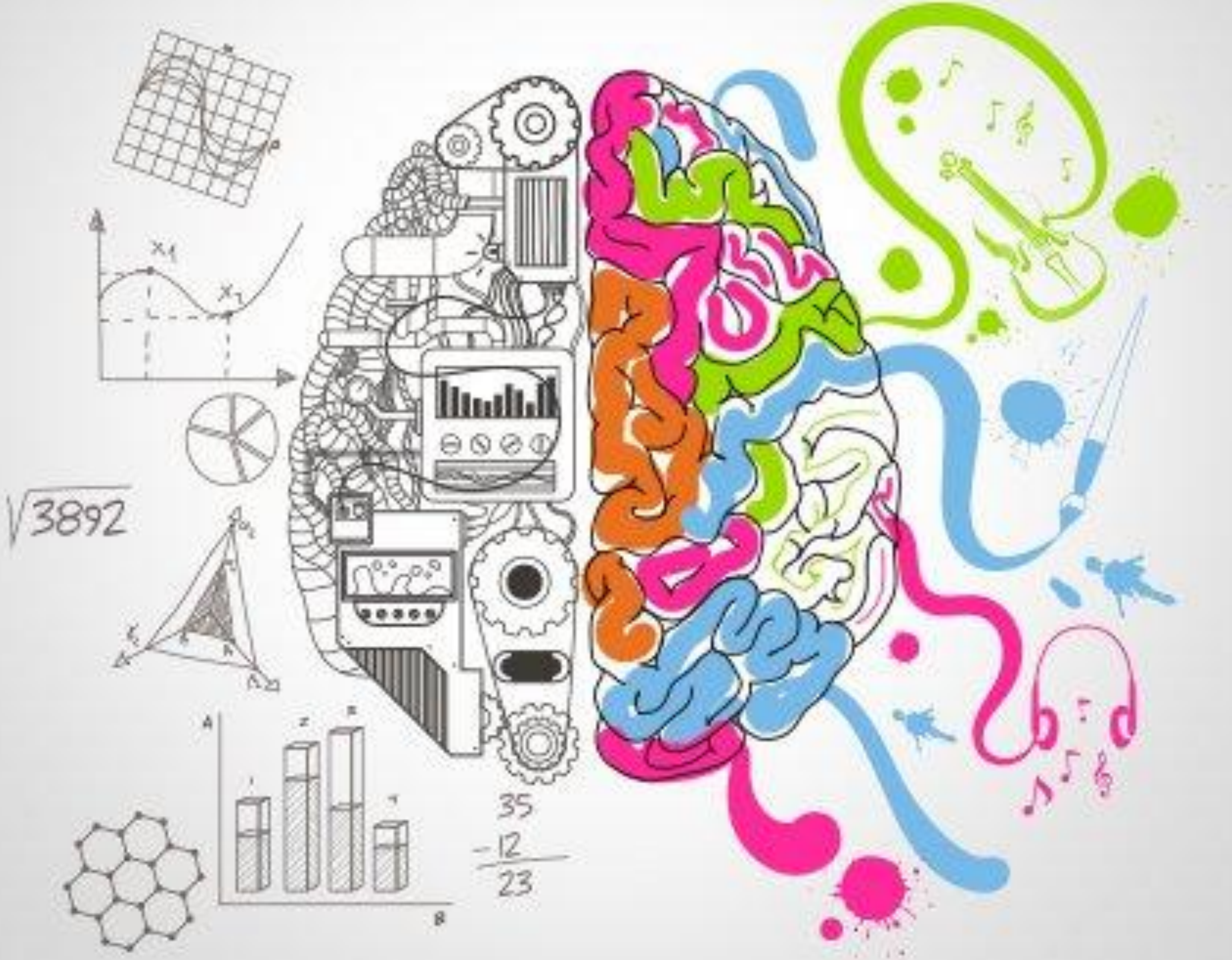
(Big) data analytics examples

How Obama's data-crunching prowess may get him re-elected

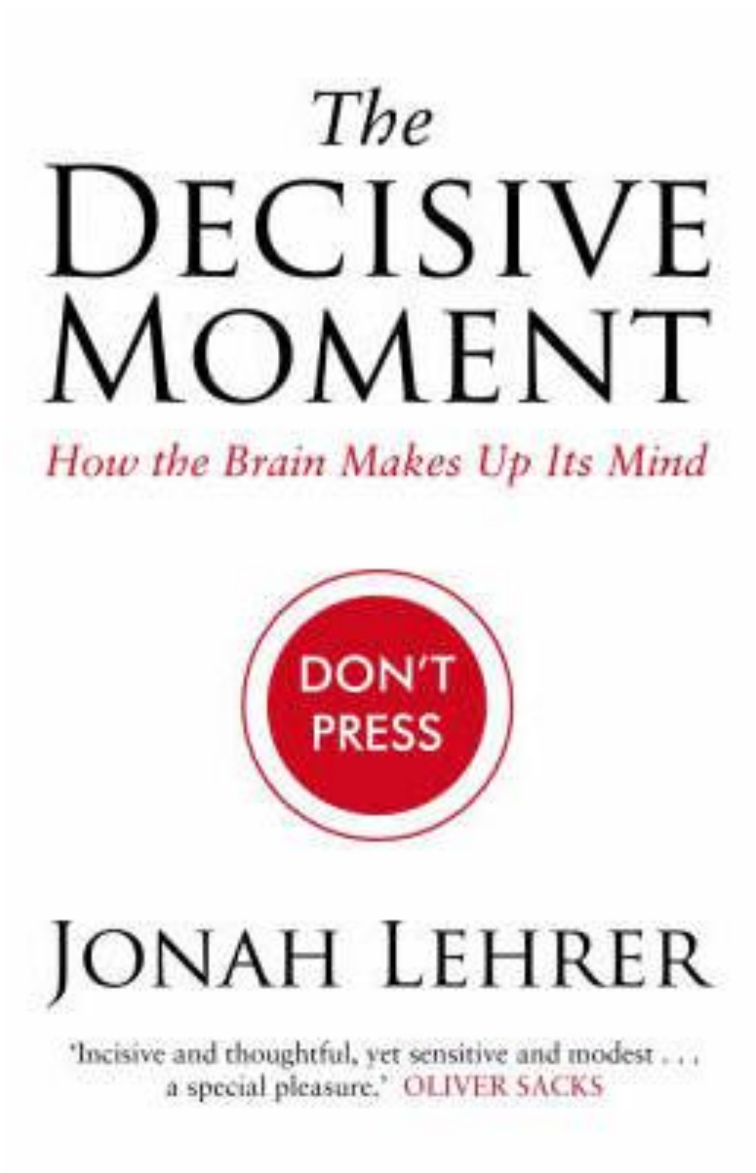
By **Micah Sifry**, Special to CNN
October 9, 2011 -- Updated 1250 GMT (2050 HKT) | Filed under: Innovations

President Obama posts a tweet during an online town hall meeting in July from the White House in Washington.

Data vs Gut



- "The best time to trust your gut is when you're making more complex decisions".
  - Buying a car or getting married are just the kind of decisions that seem to benefit the most from a more emotional, intuitive thought process.
- The other time it is a good idea to trust your gut is situations in which you've had a lot of experience.



Intuition

Great intuitive decision without consciously realizing why

Intuition

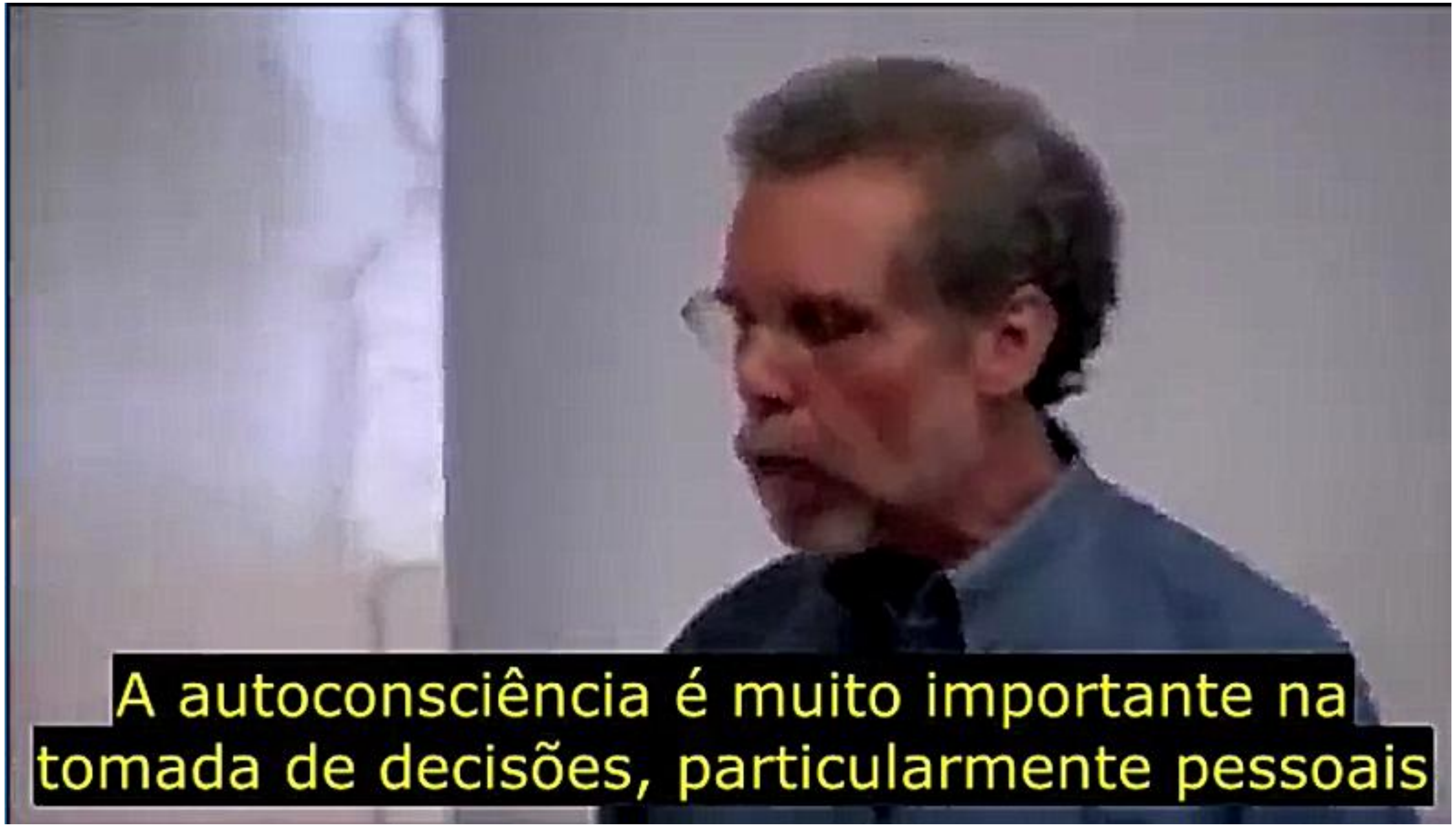
- Awash with emotion and need to engage our rational brain

Intuition

- Cognitive biases and related flaws in intuition
    - confirmation bias - people have a very hard time believing and remember evidence that contradicts their beliefs.
    - fallacy of centrality – people, especially those in authority, believe that if something important happens, they will know about it.

Emotional Intelligence and Intuition

- Daniel Goleman



Analytical models outperforming humans ...

1. **Image and object recognition**
2. **Video games**
   - Google's DeepMind uses a deep learning technique to play Atari game Breakout.
   - Google's DeepMind plays Go and wins to the world champion
3. **Voice generation and recognition**
   - Lipsync

EDITOR'S CHOICE

Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists FREE

H A Haenssle, C Fink, R Schneiderbauer, F Toberer, T Buhl, A Blum, A Kalloo, A Ben Hadj Hassen, L Thomas, A Enk, ... Show more

Annals of Oncology, Volume 29, Issue 8, 1 August 2018, Pages 1836–1842, https://doi.org/10.1093/annonc/mdy166

Published: 28 May 2018

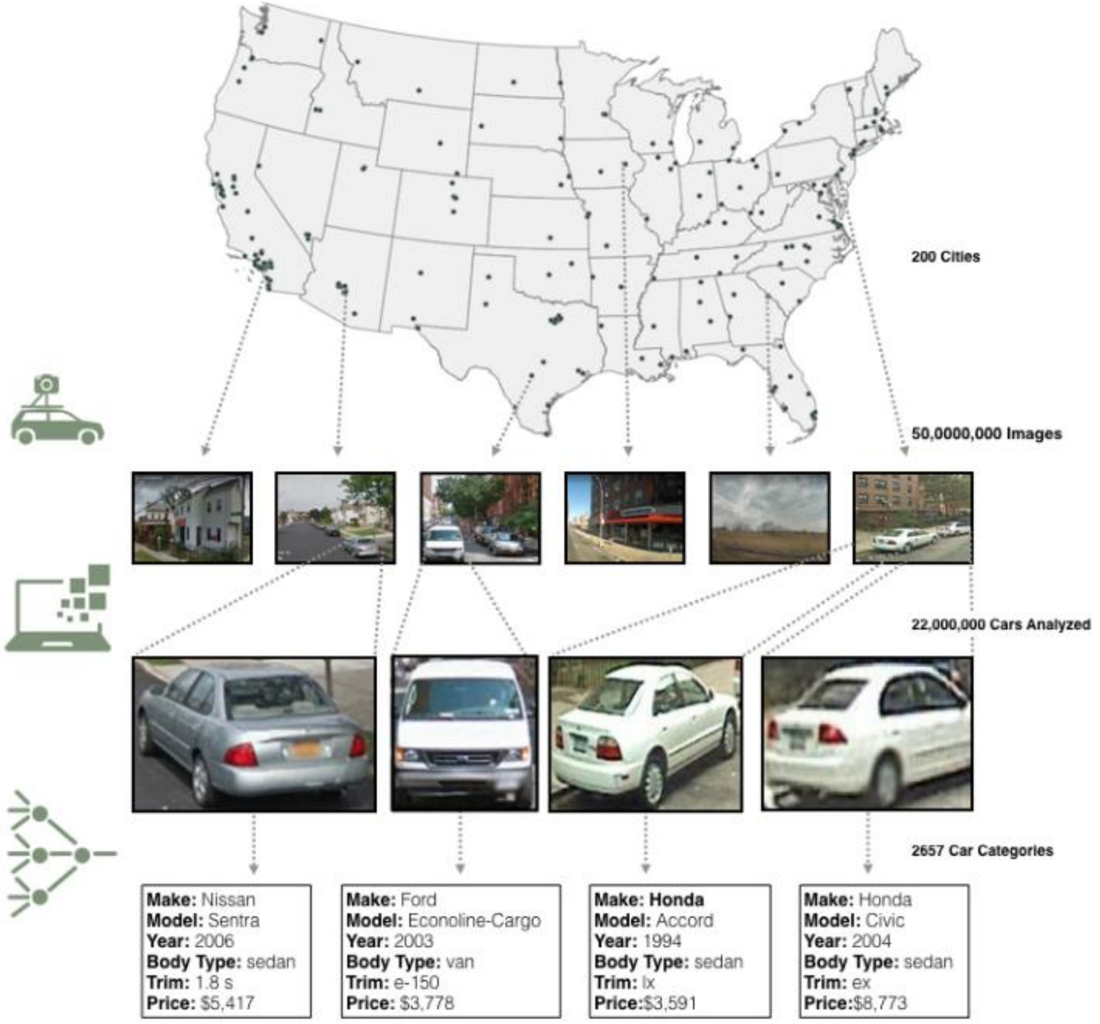
Analytical models outperforming humans ...

4. Art and style imitation

- DeepArt effects, uses the stylistic elements of one image to draw the content of another.

5. Predictions

- Timnit Gebru
- 50 million Google Street View images
- localize and recognize over 22 million cars (makes, models, body types, and years)
- if the number of sedans encountered during a 15-minute drive through a city is higher than the number of pickup trucks, the city is likely to vote for a Democrat during the next Presidential election (88% chance),”



Machine learning

- How to learn all machine learning algorithms? Thousands…
- Increasing the number every year
- Every machine learning algorithm has three components:
  - **Representation**
  - **Evaluation**
  - **Optimization**

Representation

- Decision trees
- Sets of rules / Logic programs
- Instances based
- Graphical models (Bayes/Markov nets)
- Neural networks
- Support vector machines
- Model ensembles
- …

Evaluation

- Accuracy
- Precision and recall
- Squared error
- Likelihood
- Posterior probability
- Cost / Utility
- Margin
- Entropy

Optimization

- Combinatorial optimization
  - E.g.: Greedy search
- Convex optimization
  - E.g.: Gradient descent
- Constrained optimization
  - E.g.: Linear programming

Types of Learning

- Supervised (inductive) learning
  - Training data includes desired outputs
- Unsupervised learning
  - Training data does not include desired outputs
- Semi-supervised learning
  - Training data includes a few desired outputs
- Reinforcement learning
  - Rewards from sequence of actions

Inductive Learning

- **Given** examples of a function $(x, f(x))$
- **Predict** function $f(x)$ for new examples x
  - Discrete $f(x)$: Classification
  - Continuous $f(x)$: Regression
  - $f(x)$ = $\mathrm{Probability}(x)$: Probability estimation

Supervised learning

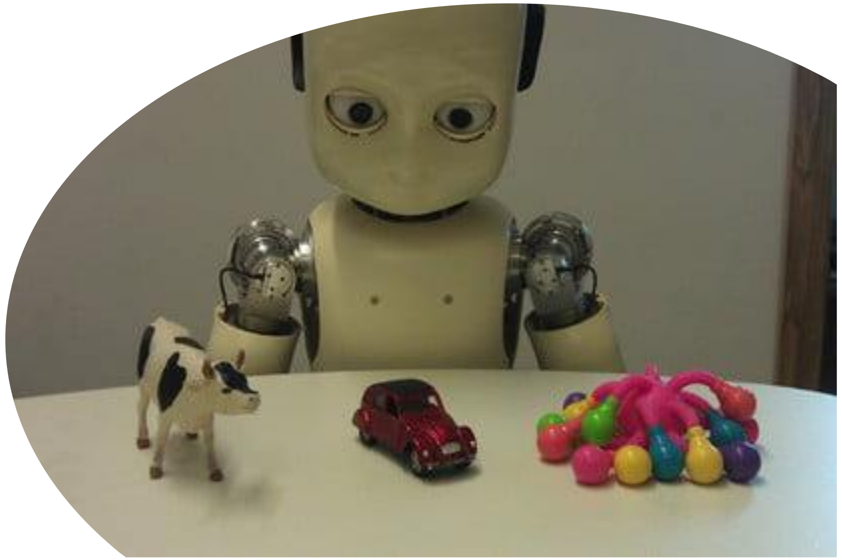
- Given: training examples $\left(x, f(x)\right)$ for some unknown function f
- Find: A good approximate to f
- **Examples**:
  - Credit Risk Assessment
    - x: properties of customer and proposed purchase
    - $f(x)$: Approve purchase or not
  - Face recognition
    - x: Bitmap feature of person's face
    - $f(x)$: name of the person
  - …

Suitable applications

- Situations where there is no human expert
- Situations where humans can perform the task but can't describe how they do it
- Situations where the desired function is changing frequently
- Situations where each user needs a customized function

What is Learning?

- Animals and Humans
  1. Learn using new experiences and prior knowledge
  2. Retain new knowledge from what is learned
  3. Repeat starting at 1.



https://icdn6.digitaltrends.com/image/digitaltrends/toddler1-416x416.jpg

- Essential to our survival and thriving

What is Learning?

- Inductive inference/modeling
  - Developing a general model/hypothesis from examples
  - Objective is to achieve good generalization for making estimates/predictions
- It’s like … Fitting a curve to data
  - Also considered modeling the data
  - Statistical modeling

Learning Bias

- (also known as inductive bias)
- set of assumptions that the learner uses to predict outputs given inputs that it has not encountered
- Inductive bias depends upon:
  - Having prior knowledge
  - Selection of most related knowledge

Learning Bias

- Universal heuristics - Occam's Razor
- Knowledge of intended use – Medical diagnosis
- Knowledge of the source - Teacher
- Knowledge of the task domain
- Analogy with previously learned tasks

Tom Mitchell, 1980

The model *g*(***x***)

- We need to choose what *kind* of model we want to learn

Input

x∈X

An item **x**

drawn from an instance space **X**

Learned Model

y = *g*(***x***)

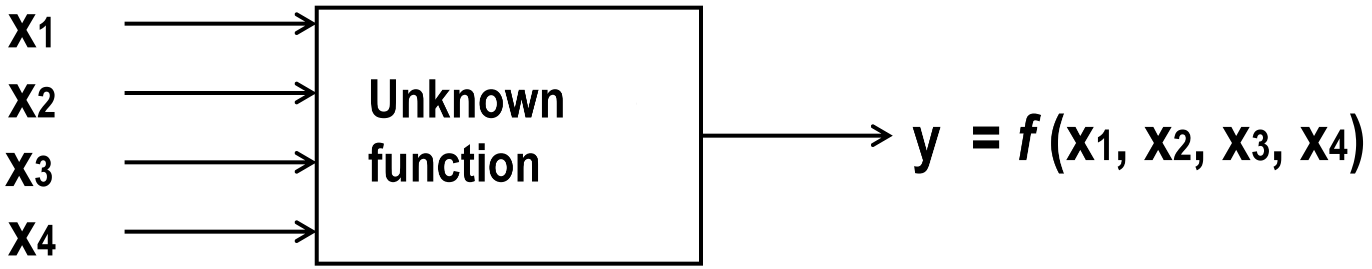
Output

y∈Y

An item **y**

drawn from a label space **Y**

A Learning Problem



Example	x_1	x_2	x_3	x_4	y
1	0	0	1	0	0
2	0	1	0	0	0
3	0	0	1	1	1
4	1	0	0	1	1
5	0	1	1	0	0
6	1	1	0	0	0
7	0	1	0	1	0

Can you learn this function? What is it?

Hypothesis Space: Complete Ignorance

- There are 2^{16} = 65536 possible functions over 4 input features.
- We can't figure out which one is correct until we've seen every possible input-output pair.
- After observing seven examples we still have 2^9 possibilities for *f*
- Is Learning Possible?

Example	x_1	x_2	x_3	x_4	y
1	0	0	0	0	?
2	0	0	0	1	?
	0	0	1	0	0
	0	0	1	1	1
	0	1	0	0	0
	0	1	0	1	0
	0	1	1	0	0
	0	1	1	1	?
	1	0	0	0	?
	1	0	0	1	1
	1	0	1	0	?
	1	0	1	1	?
	1	1	0	0	0
	1	1	0	1	?
	1	1	1	0	?
16	1	1	1	1	?

Hypothesis Space: Simple Rules

There are only 16 simple conjunctive rules of the form $y = x_i \wedge x_j \wedge x_k$

Variables	*Counterexample*
$\Rightarrow y$	1
$x_1 \Rightarrow y$	3
$x_2 \Rightarrow y$	2
$x_3 \Rightarrow y$	1
$x_4 \Rightarrow y$	7
$x_1 \wedge x_2 \Rightarrow y$	3
$x_1 \wedge x_2 \wedge x_3 \Rightarrow y$	3
$x_1 \wedge x_2 \wedge x_3 \wedge x_4 \Rightarrow y$	3

Example	x_1	x_2	x_3	x_4	y
1	0	0	1	0	0
2	0	1	0	0	0
3	0	0	1	1	1
4	1	0	0	1	1
5	0	1	1	0	0
6	1	1	0	0	0
7	0	1	0	1	0

No simple rule explains the data.

Hypothesis Space: *m-of-n* rules

At least *m-of the n* variables must be true. There are 32 possible rules.

Variables	***Counterexample***			
	1-of	*2-of*	*3-of*	*4-of*
$\{x_1\}$	3	-	-	-
$\{x_2\}$	2	-	-	-
…				
$\{x_1, x_2\}$	3	3	-	-
$\{x_1, x_3\}$	4	3	-	-
…				
$\{x_1, x_2, x_3\}$	1	3	3	-
$\{x_1, x_2, x_4\}$	2	3	3	-
…				
$\{x_1, x_3, x_4\}$	1	***	3	-
$\{x_1, x_2, , x_3, x_4\}$	1	5	3	3

Example	x_1	x_2	x_3	x_4	y
1	0	0	1	0	0
2	0	1	0	0	0
3	0	0	1	1	1
4	1	0	0	1	1
5	0	1	1	0	0
6	1	1	0	0	0
7	0	1	0	1	0

Views of Learning

- **Learning is the removal of our remaining uncertainty**
    - Suppose we knew that the unknown function was an m-of-n Boolean function, then we could use the training data to infer which function it is.
- **Learning requires guessing a good hypothesis class**:
    - We can start with a very small class and enlarge it until it contains an hypothesis that fits the data.
- **We could be wrong!**
  - **Our prior knowledge might be wrong**
  - **Our guess of the hypothesis space could be wrong**
    - The smaller the hypothesis class, the more likely are wrong

- Example (both are consistent with the training data):
  - $x_4 \wedge oneof\ \{x_1, x_3\} \Rightarrow y$
  - $x_4 \wedge -x_2 \Rightarrow y$
- If this is the unknown function, then we will make errors when we are given new examples, and are asked to predict the value of the function

Terminology

- **Training example**. An example of the form $\left(x, f(x)\right)$
- **Target function** (target concept): The true function f
- **Hypothesis**: A proposed function h, believed to be similar to f. The output of our learning algorithm.
- **Concept**: Boolean function. Example for which $f(x) = 1$ are positive examples; those for which $f(x) = 0$ are negative examples (instances)
- **Classifier**: A discrete valued function produced by the learning algorithm. The possible value of $f: \{1,2, \dots, K\}$ are the classes or class labels. (In most algorithms the classifier will actually return a real valued function that we'll have to interpret).
- **Hypothesis space**: The space of all hypotheses that can, in principle, be the output of the learning algorithm.
- **Version space.** The space of all hypothesis space that have not yet been rule out by a training example

Questions?