CAUSAL NETWORKS

Fabio Stella

Department of Informatics, Systems and Communication

University of Milan-Bicocca

Viale Sarca 336, 2016 Milan, ITALY

E-mail: fabio.stella@unimib.it

Twitter: FaSt@FabioAStella

In Memory of Luciano and Enrico.
Not a day goes by, that I don't think of you.

Artificial Intelligence

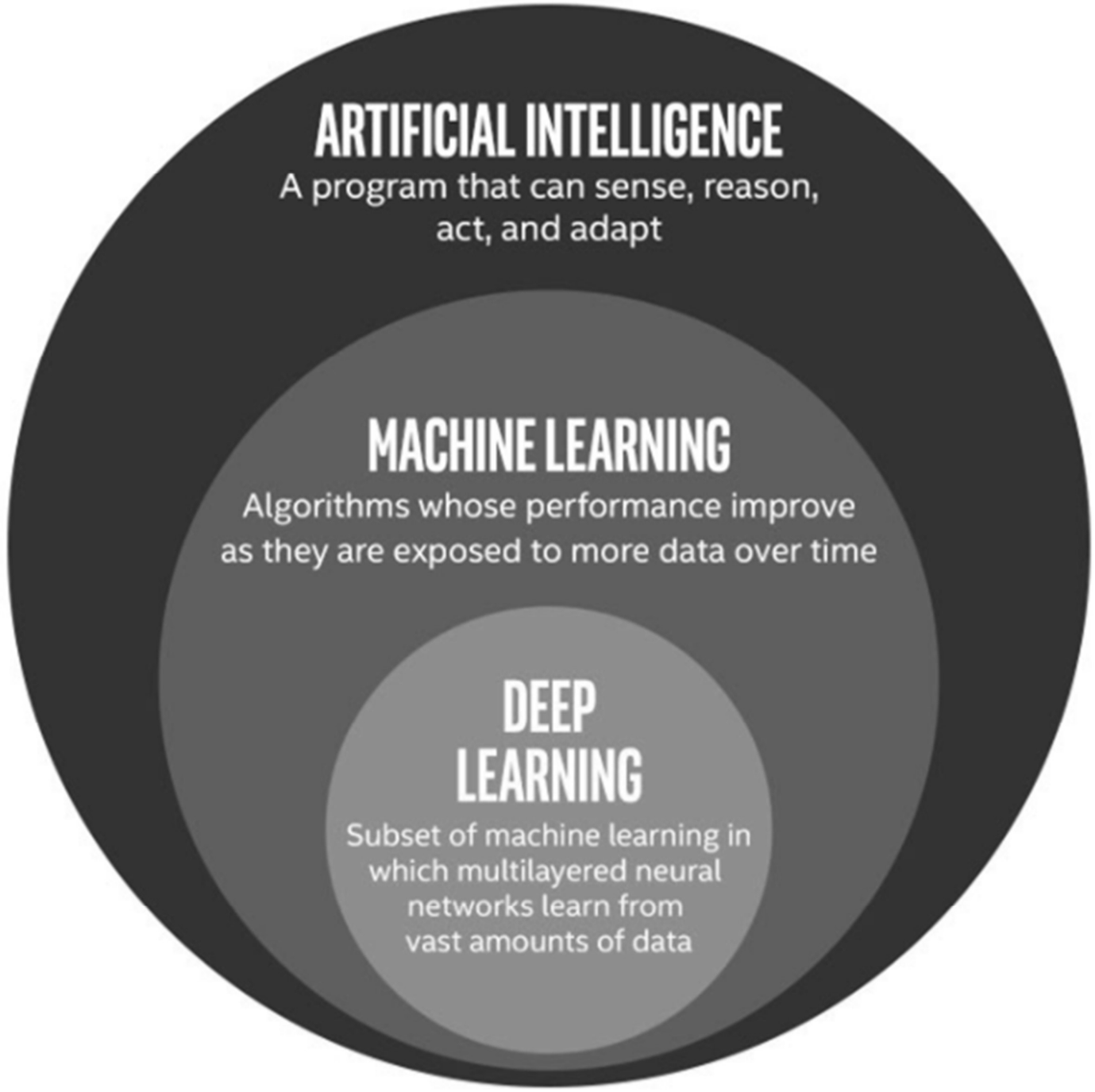
Intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans.

- Colloquially, the term Artificial Intelligence (**AI**) is used to describe machines/computers that mimic “cognitive” functions that humans associate with other human minds, such as "learning" and "problem solving".

- Two kinds of AI:

  - Weak

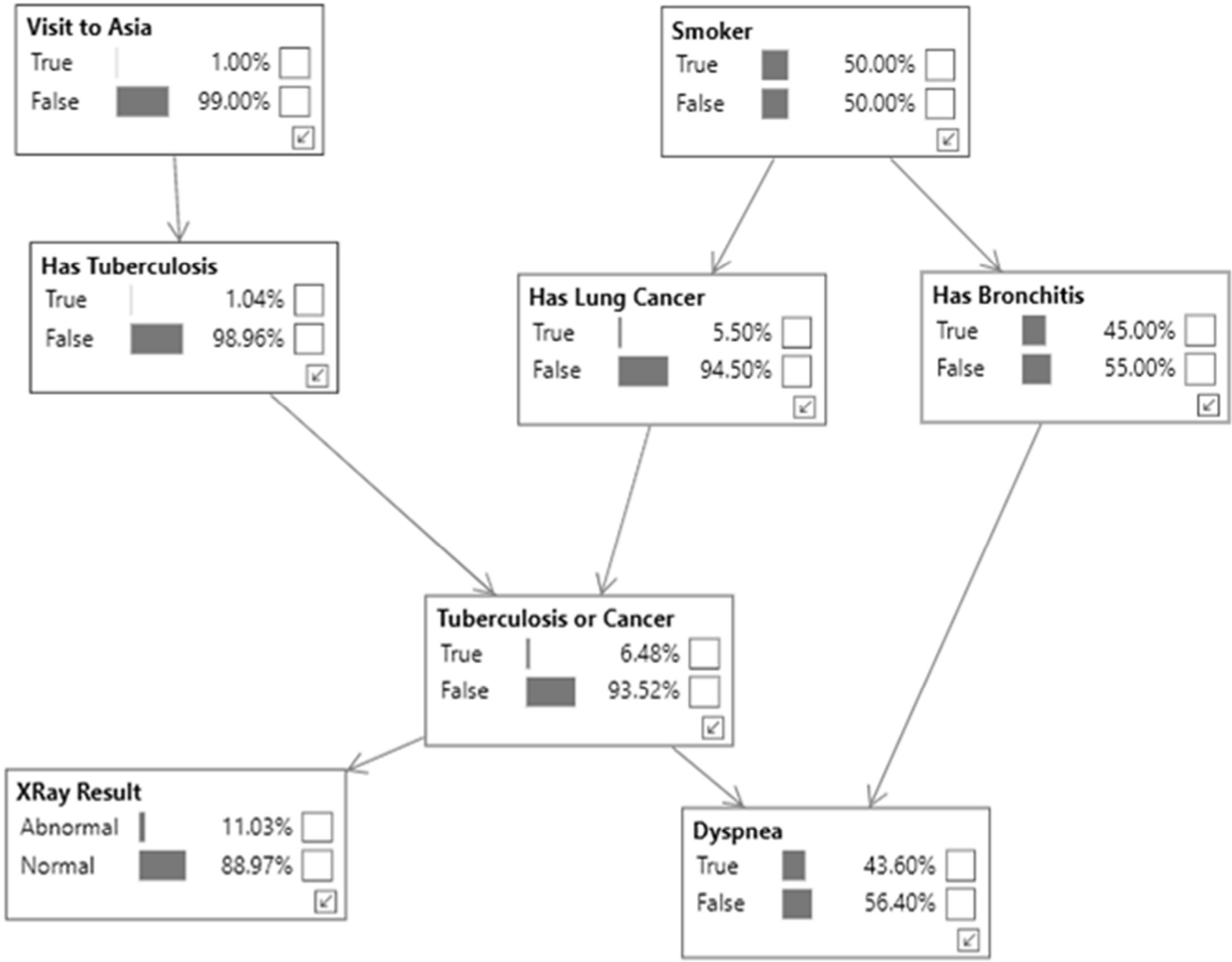
  - Strong



Artificial Intelligence

Bayesian Networks

- A type of statistical model that represents a set of variables and their conditional dependencies via a directed acyclic graph (DAG).
- Bayesian networks are ideal for taking an event that occurred and predicting the likelihood that any one of several possible known causes was the contributing factor.
- Structural Causal Models.

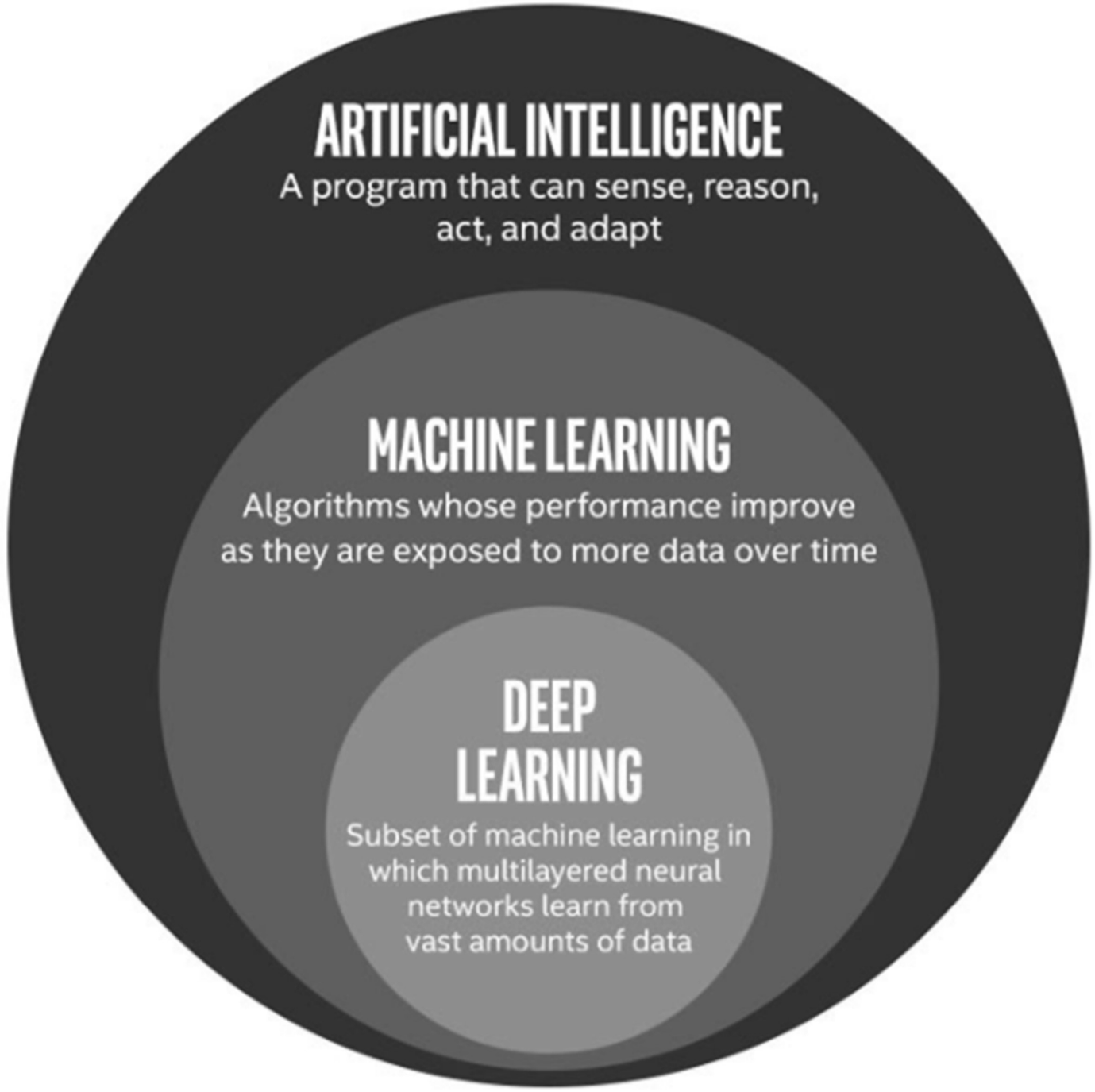


Machine Learning

Algorithms and statistical models that computer systems use in order to perform a specific task effectively without using explicit instructions, relying on patterns and inference instead.

Three kinds of ML:

- Supervised
- Self-Supervised
- Reinforcement Learning



Machine Learning

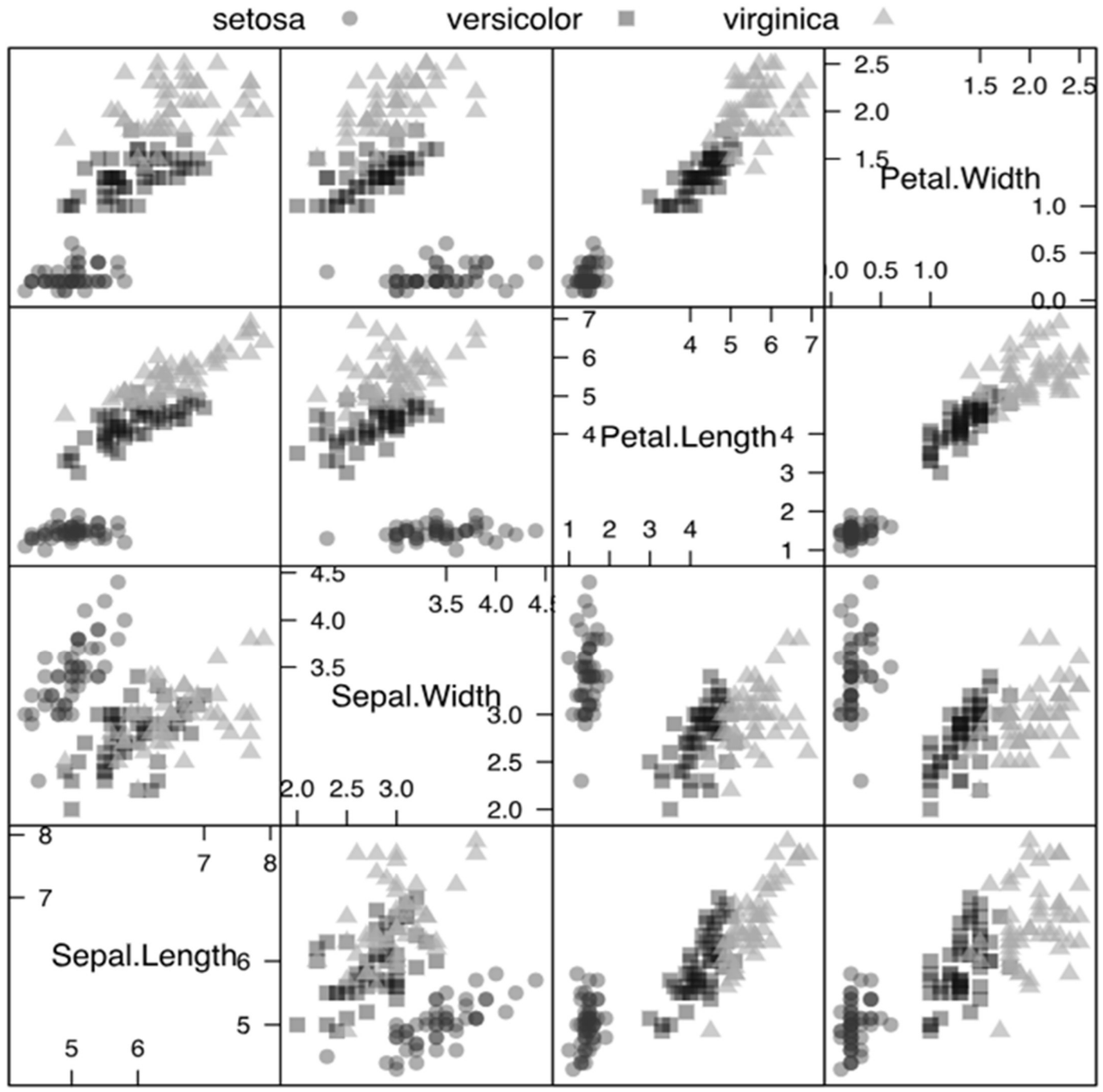
Supervised

- **Classification**

SETOSA

VERSICOLOR

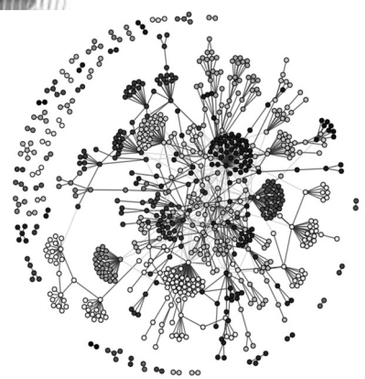
VIRGINICA



Machine Learning

Self-Supervised

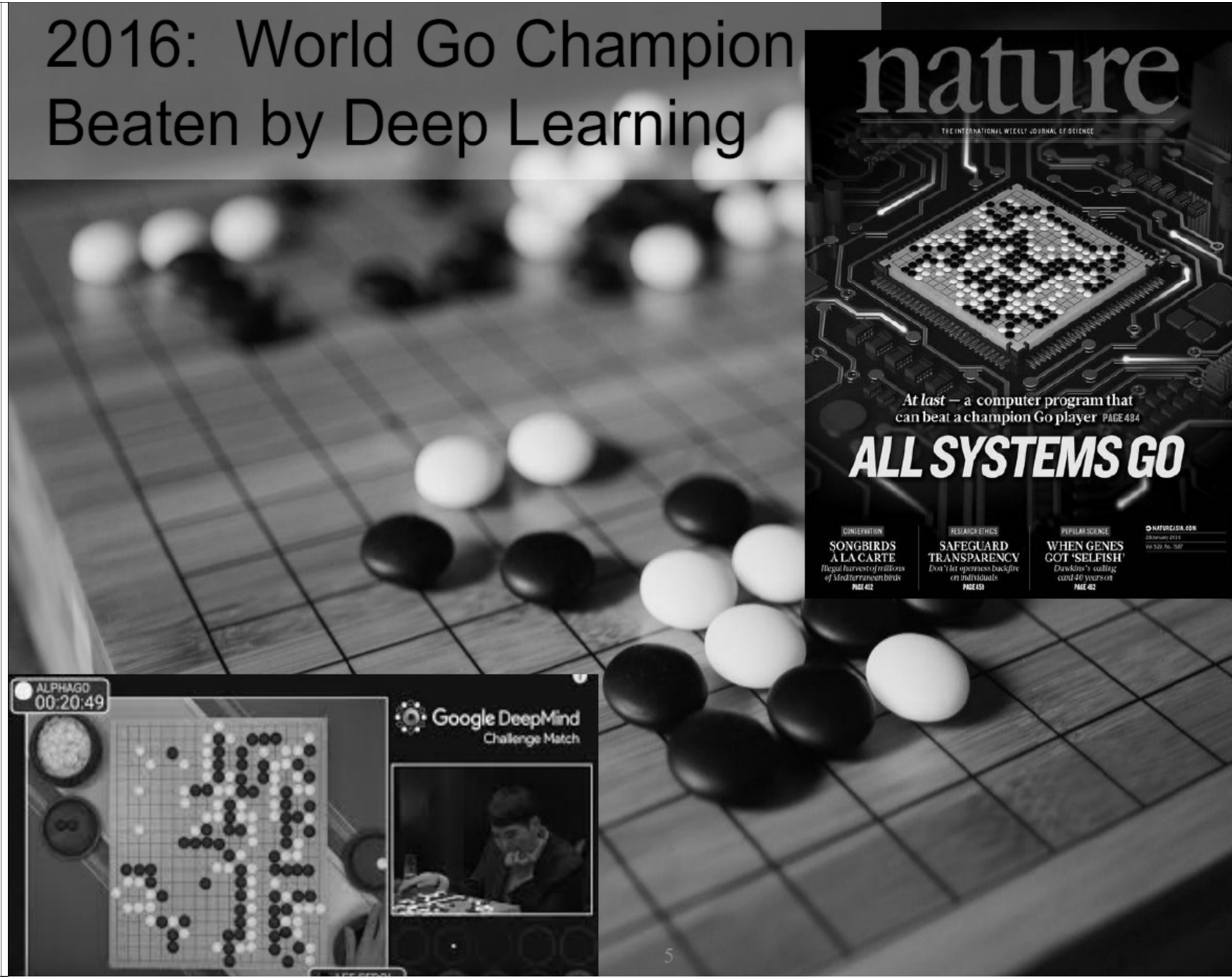
- Recommendation System
- Market Basket Analysis
- Social Network Analysis



Machine Learning

Reinforcement Learning

- Learn by interacting with the environment
- The environment reacts to our decisions/actions
- Sequential learning, only at the end of the game we know our performance (reward/punishment)

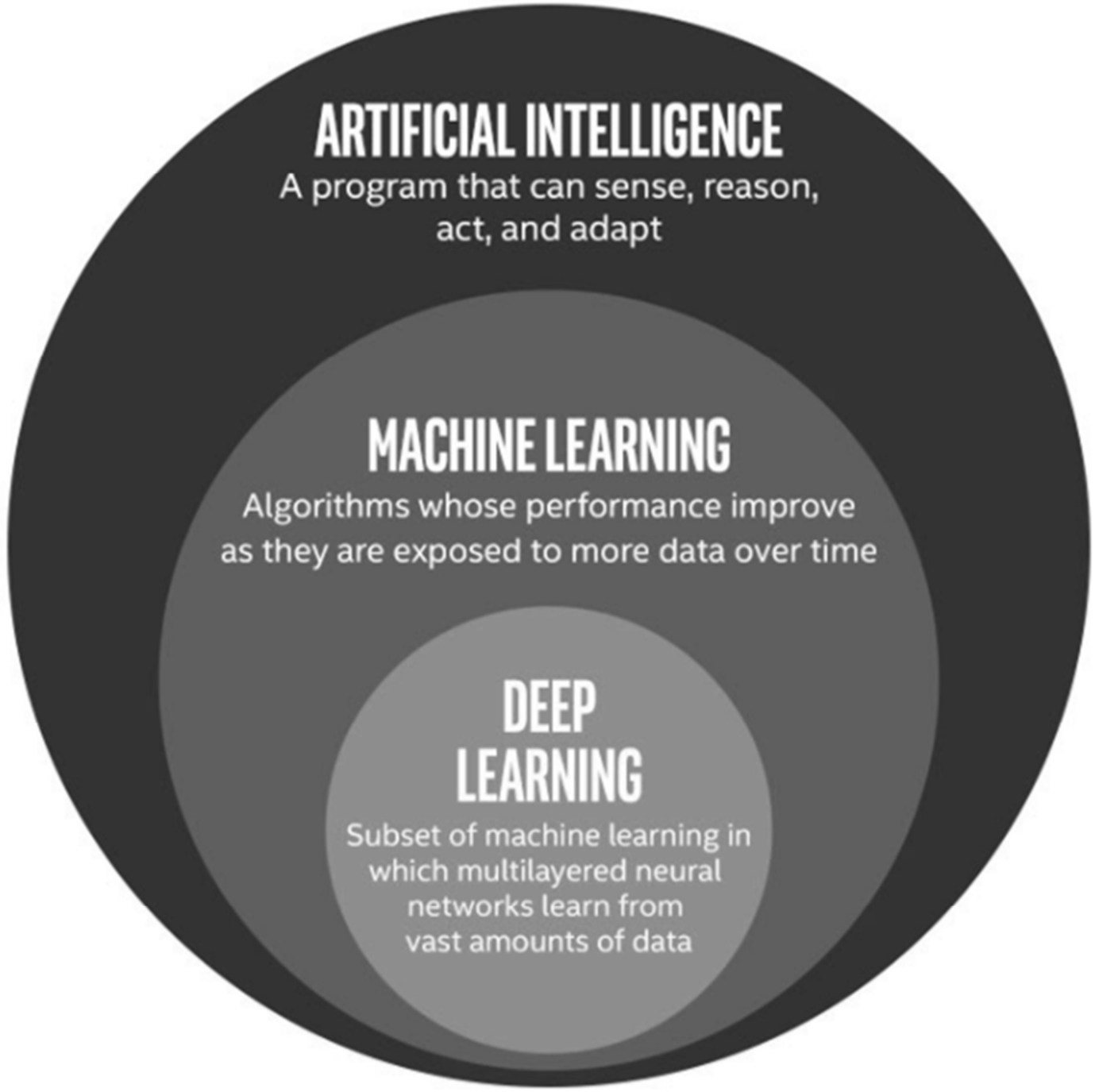


Deep Learning

Is part of a broader family of machine learning methods based on Artificial Neural Networks.

Three kinds of DL:

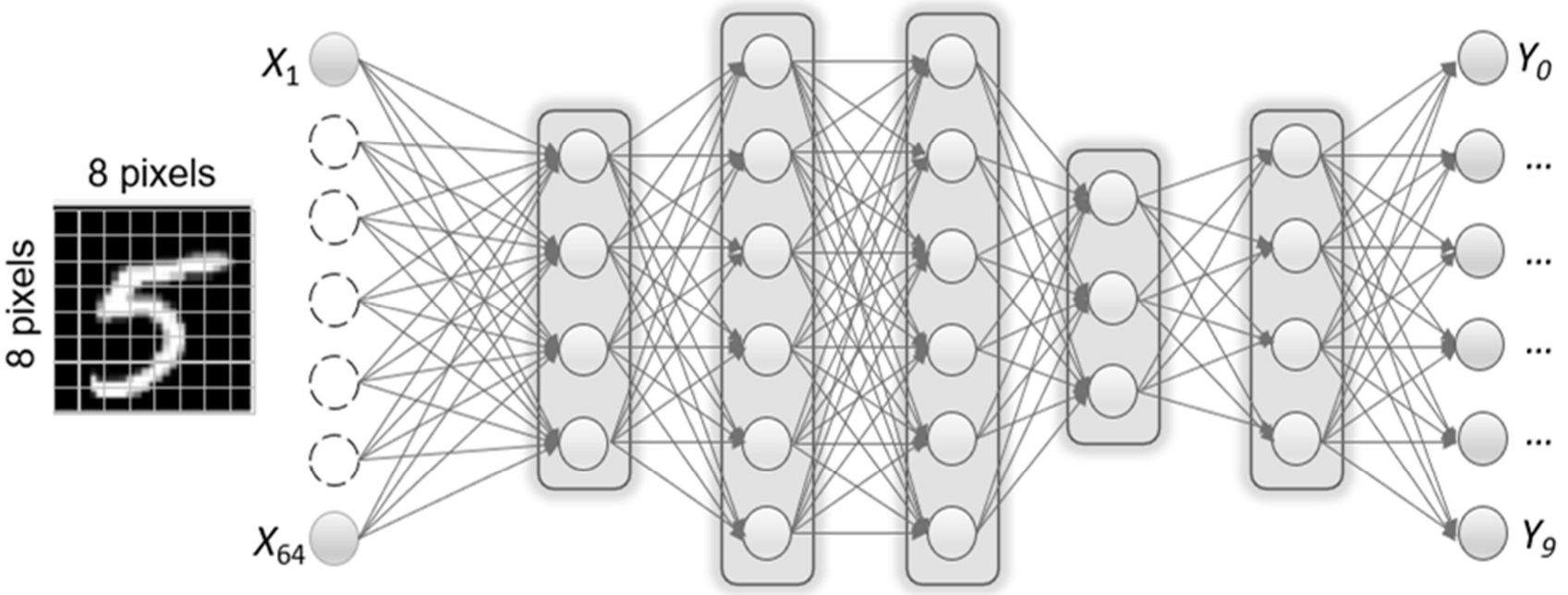
- Supervised
- Self-Supervised
- Reinforcement Learning



Deep Learning

Feedforward Neural Networks

- The first and simplest type of artificial neural network devised.
- The information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any) and to the output nodes.
- There are no cycles or loops in the network.



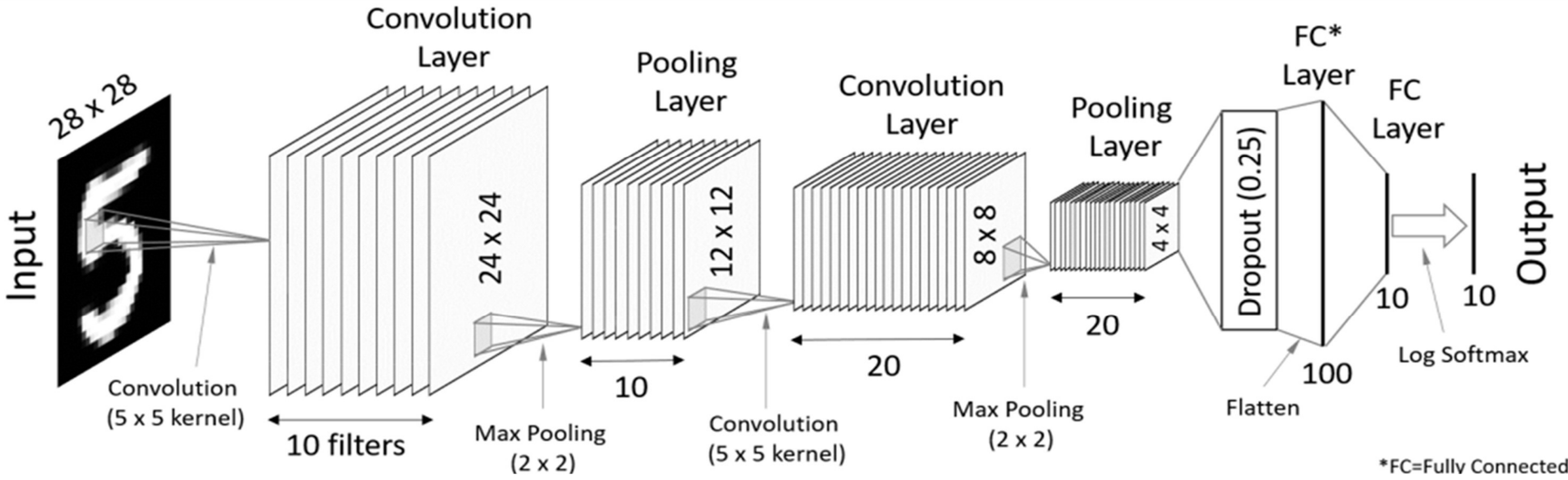
Deep Learning

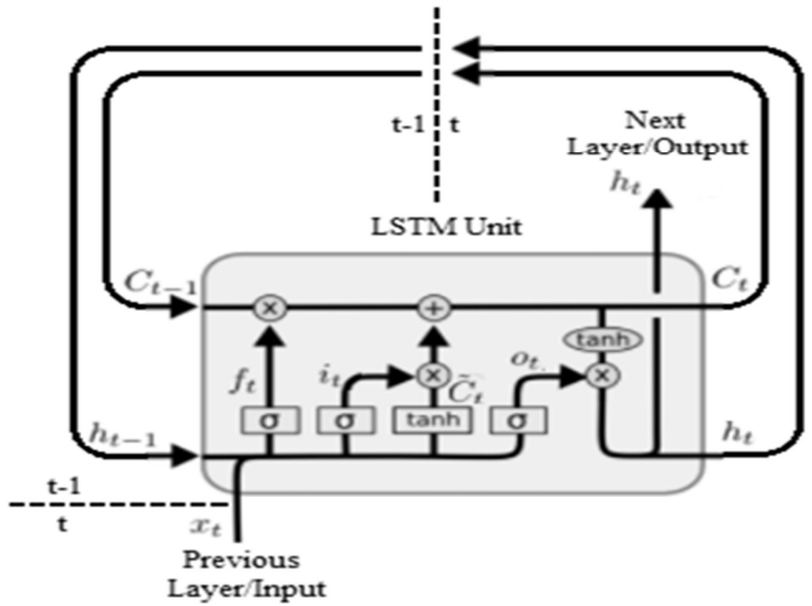
CNN

- Regularized versions of multilayer perceptrons which are fully connected and thus prone to overfitting the data.
- Regularization by adding some form of magnitude measurement of weights to the loss function.
- Different approach towards regularization: take advantage of the hierarchical pattern in data and assemble more complex patterns using smaller and simpler patterns.

LSTM

- An artificial Recurrent Neural Network architecture.
- Unlike standard feedforward neural networks, LSTM has feedback connections that make it a "general purpose computer" (it can compute anything that a Turing machine can).
- LSTM started to revolutionize speech recognition, outperforming traditional models in certain speech applications.

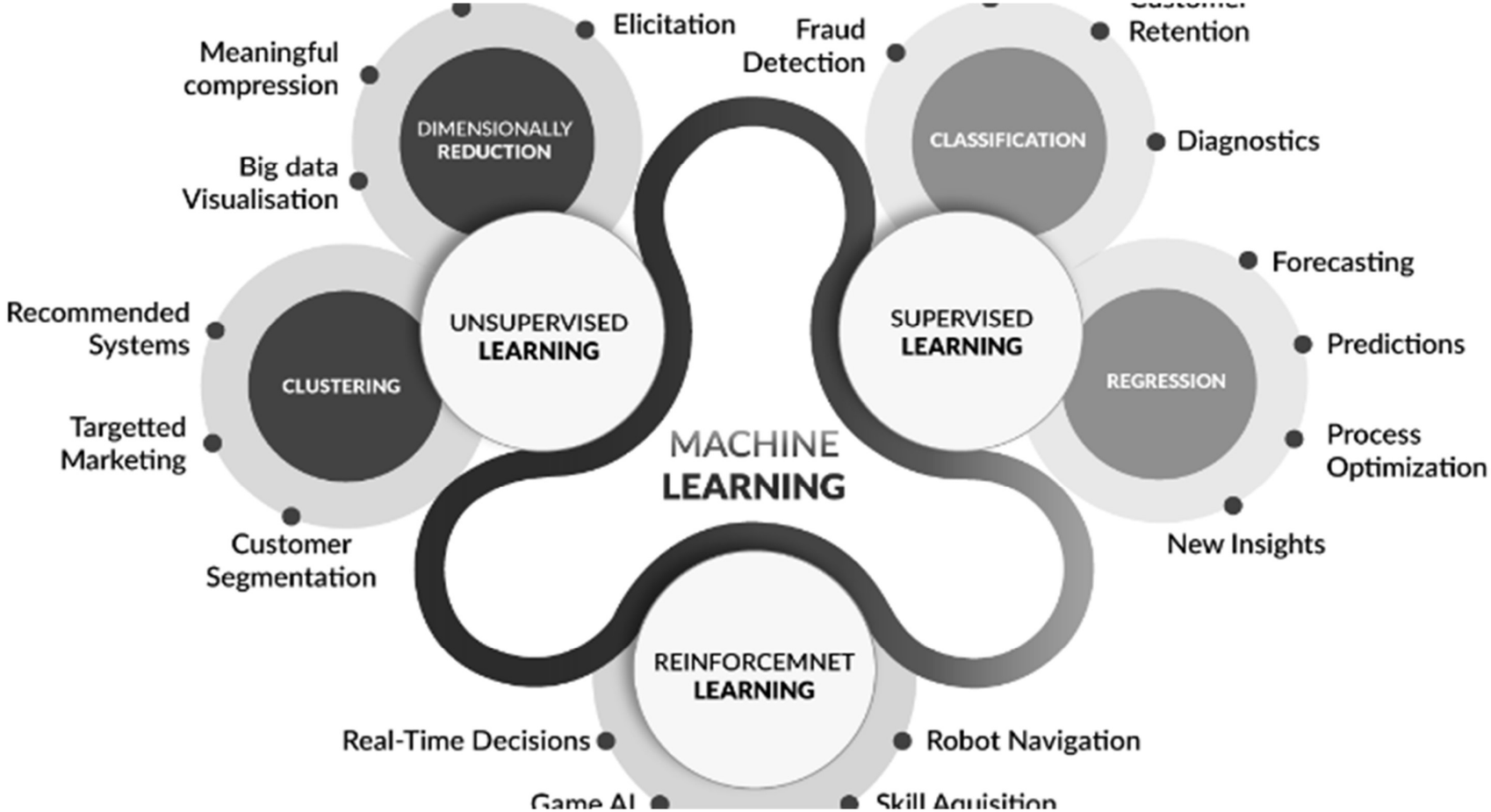




$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$
$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$
$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$
$$o_t = \sigma\left(W_o \; [h_{t-1}, x_t] \; + \; b_o\right)$$
$$h_t = o_t * \tanh\left(C_t\right)$$

Machine Learning

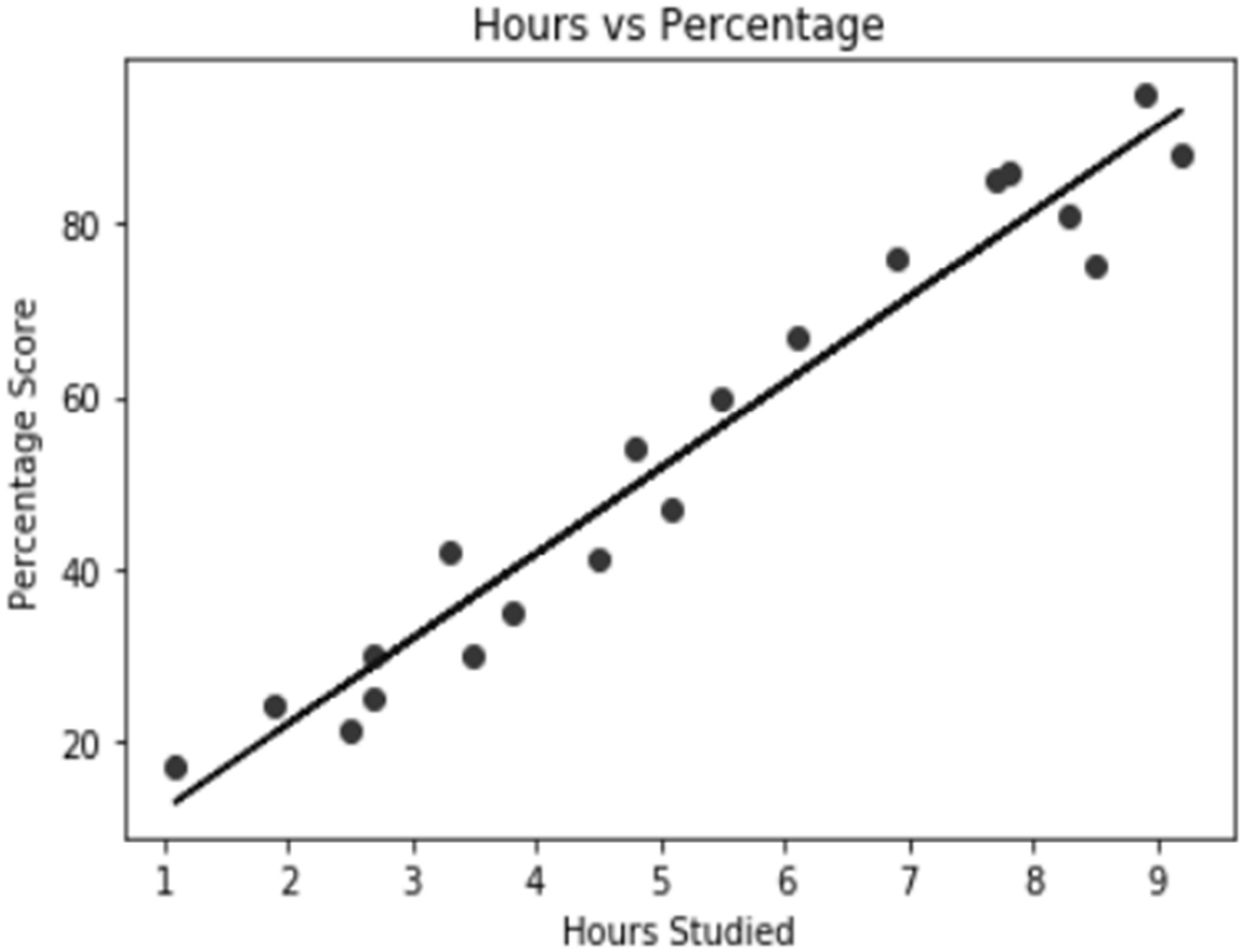
- Many different names for learning



But most of machine learning nowadays
is just **curve fitting**

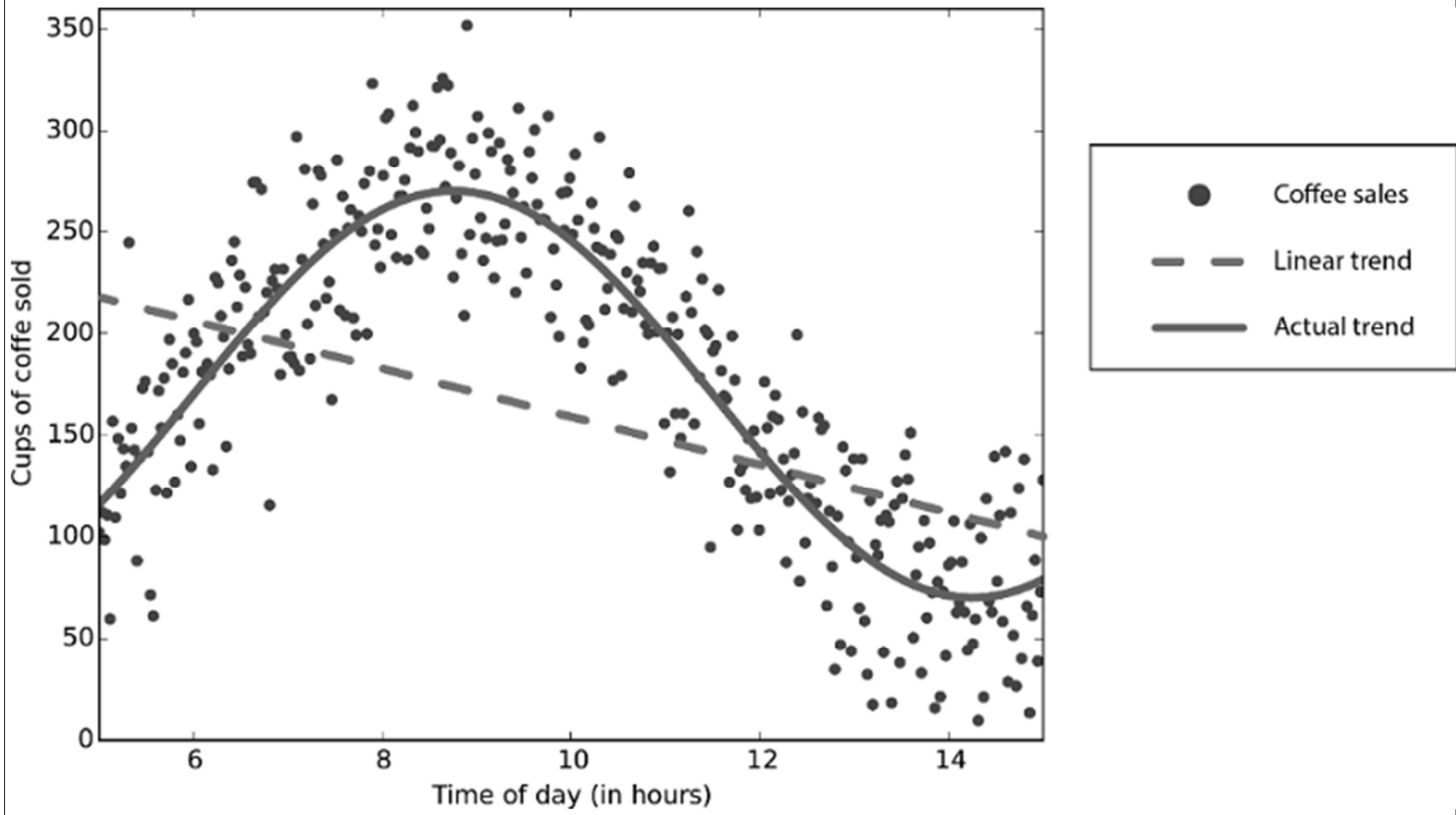
Machine Learning

- Curve fitting – linear (correlation)



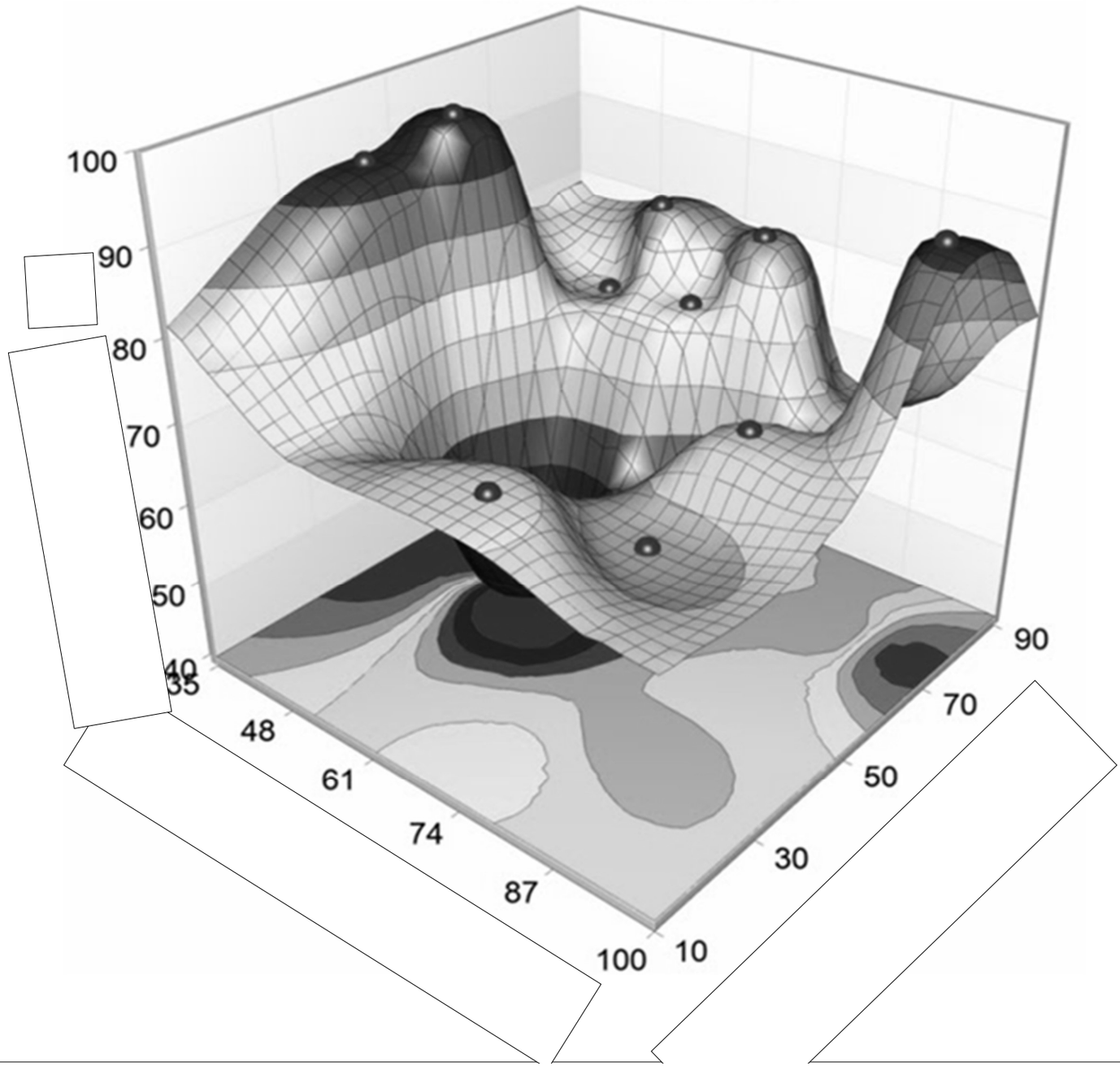
Machine Learning

- Curve fitting - nonlinear



Machine Learning

- Curve fitting - multidimensional



Machine Learning

- Deep Neural Networks

Highly dimensional, highly nonlinear

curve fitting

DOI:10.1145/3271625

What just happened in artificial intelligence and how it is being misunderstood.

BY ADNAN DARWICHE

Human-Level Intelligence or Animal-Like Abilities?

"The vision systems of the eagle and the snake outperform everything that we can make in the laboratory, but snakes and eagles cannot build an eyeglass or a telescope or a microscope."

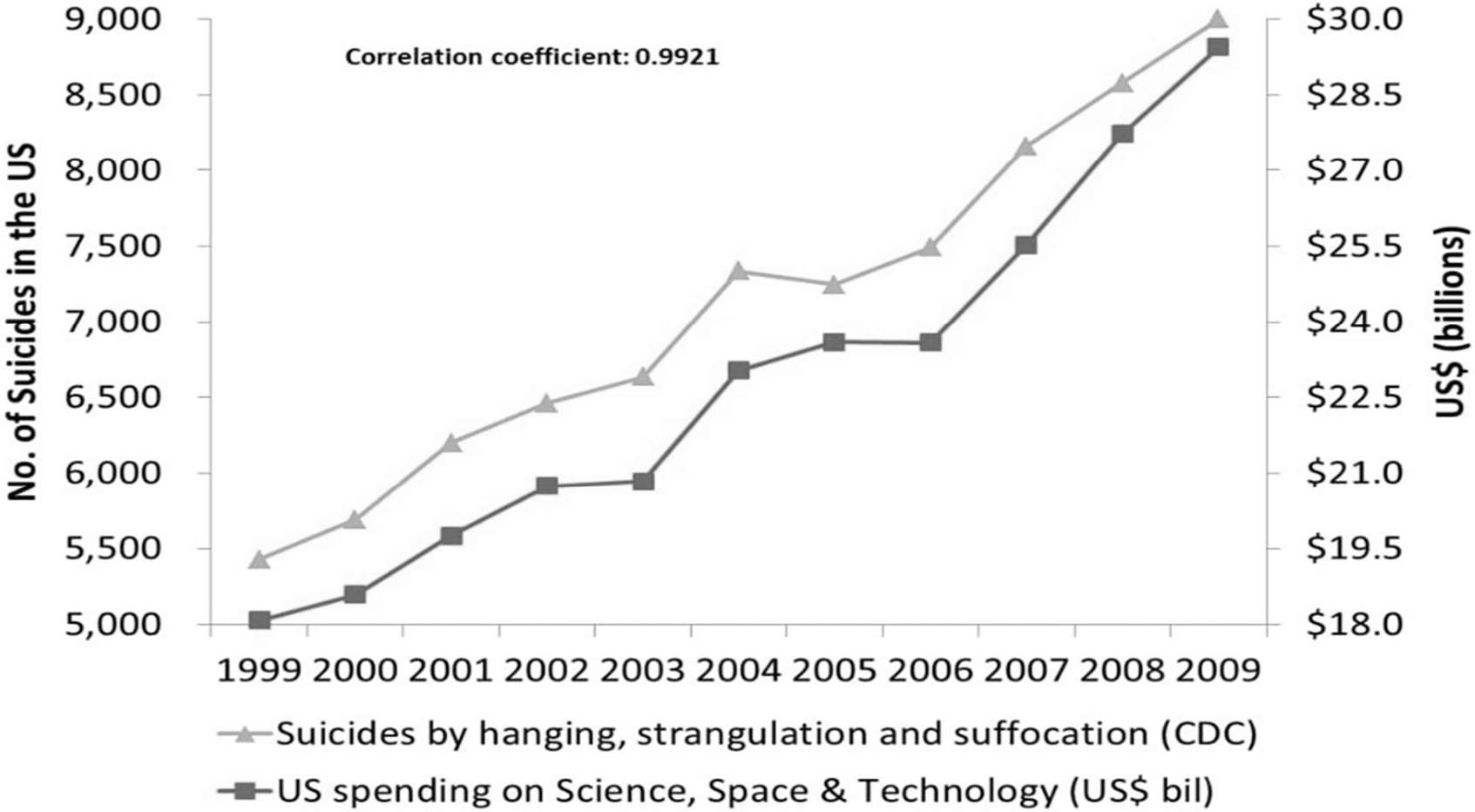
"The vision systems of the eagle and the snake outperform everything that we can make in the laboratory, but snakes and eagles cannot build an eyeglass or a telescope or a microscope."

— *Judea Pearl*

Machine Learning

- Correlation works very well in many cases, but what if …
- Spurious Correlations

Fitting can be highly misleading



Spurious Correlations

http://www.tylervigen.com/spurious-correlations

WHY CAUSALITY MATTERS?

Why study causation?

- To make sense of data
  - effect of smoking on lung cancer?
  - effect of education on salaries?
  - effect of carbon emissions on the climate?
- To understand how we have an effect
  - malaria caused by mosquitos or by mal-air?

Why study causation?

- To make sense of data
  - effect of smoking on lung cancer?
  - effect of education on salaries?
  - effect of carbon emissions on the climate?
- To understand how we have an effect
  - malaria caused by mosquitos or by mal-air?
- To guide actions and policies
  - pack mosquito nets or use breathing masks?
  - reduce CO_2 emissions?
  - have a degree?
  - stop smoking?


It’s my
moment

TEST 1

- A group of sick patients are given the option to try a new drug (**TREATMENT**)
- Among those who took the drug (**TREATMENT GROUP**), a lower percentage recovered (**OUTCOME**) than among those who did not (**CONTROL GROUP**)
- However, when we partition by gender (**COVARIATE**), we see that:
  - *more* men taking the drug recover than do men are not taking the drug, and
  - *more* women taking the drug recover than do women are not taking the drug!

We record the number of recoveries of 700 patients who were given access to the drug.

A total of 350 patients chose to take the drug and 350 patients did not.

The results of the study are shown in the following Table.

	Drug		No Drug	
	patients	recovered	patients	recovered
Men	87	81	270	234
Women	263	192	80	55
Combined data	350	273	350	289

Should a doctor prescribe this drug or not?

Simpson's Paradox

- Named after Edward Simpson (born 1922), statistician
- A group of sick patients are given the option to try a new drug (**TREATMENT**)
- Among those who took the drug (**TREATMENT GROUP**), a lower percentage recovered (**OUTCOME**) than among those who did not (**CONTROL GROUP**)
- However, when we partition by gender (**COVARIATE**), we see that:
  - *more* men taking the drug recover than do men are not taking the drug, and
  - *more* women taking the drug recover than do women are not taking the drug!

The drug appears to help men and women, but hurt the general population

Simpson's Paradox

- Named after Edward Simpson (born 1922), statistician
- A group of sick patients are given the option to try a new drug (**TREATMENT**)
- Among those who took the drug (**TREATMENT GROUP**), a lower percentage recovered (**OUTCOME**) than among those who did not (**CONTROL GROUP**)
- However, when we partition by gender (**COVARIATE**), we see that:
  - *more* men taking the drug recover than do men are not taking the drug, and
  - *more* women taking the drug recover than do women are not taking the drug!

We record the recovery rates of 700 patients who were given access to the drug. A total of 350 patients chose to take the drug and 350 patients did not. The results of the study are shown in **Table 1.1**.

Table 1.1 Results of a study into a new drug, with gender being taken into account

	Drug			No Drug		
	patients	recovered	% recovered	patients	recovered	% recovered
Men	87	81	93%	270	234	87%
Women	263	192	73%	80	55	69%
Combined data	350	273	78%	350	289	83%

Simpson's Paradox

- Named after Edward Simpson (born 1922), statistician
- A group of sick patients are given the option to try a new drug (**TREATMENT**)
- Among those who took the drug (**TREATMENT GROUP**), a lower percentage recovered (**OUTCOME**) than among those who did not (**CONTROL GROUP**)
- However, when we partition by gender (**COVARIATE**), we see that:
  - *more* men taking the drug recover than do men are not taking the drug, and
  - *more* women taking the drug recover than do women are not taking the drug!

We record the recovery rates of 700 patients who were given access to the drug. A total of 350 patients chose to take the drug and 350 patients did not. The results of the study are shown in **Table 1.1**.

Table 1.1 Results of a study into a new drug, with gender being taken into account

	Drug			No Drug		
	patients	recovered	% recovered	patients	recovered	% recovered
Men	87	81	93%	270	234	87%
Women	263	192	73%	80	55	69%
Combined data	350	273	78%	350	289	83%

Drug vs non-drug takers recovery rates:

- 93% vs 87% male

Simpson's Paradox

- Named after Edward Simpson (born 1922), statistician
- A group of sick patients are given the option to try a new drug (**TREATMENT**)
- Among those who took the drug (**TREATMENT GROUP**), a lower percentage recovered (**OUTCOME**) than among those who did not (**CONTROL GROUP**)
- However, when we partition by gender (**COVARIATE**), we see that:
  - *more* men taking the drug recover than do men are not taking the drug, and
  - *more* women taking the drug recover than do women are not taking the drug!

We record the recovery rates of 700 patients who were given access to the drug. A total of 350 patients chose to take the drug and 350 patients did not. The results of the study are shown in **Table 1.1**.

Table 1.1 Results of a study into a new drug, with gender being taken into account

	Drug			No Drug		
	patients	recovered	% recovered	patients	recovered	% recovered
Men	87	81	93%	270	234	87%
Women	263	192	73%	80	55	69%
Combined data	350	273	78%	350	289	83%

Drug vs non-drug takers recovery rates:

- 93% vs 87% male
- 73% vs 69% female
- 78% vs 83% general population!

Should a doctor prescribe the drug; to whom?

Should a policy maker approve the drug for use?

Simpson's Paradox

- Named after Edward Simpson (born 1922), statistician
- A group of sick patients are given the option to try a new drug (**TREATMENT**)
- Among those who took the drug (**TREATMENT GROUP**), a lower percentage recovered (**OUTCOME**) than among those who did not (**CONTROL GROUP**)
- However, when we partition by gender (**COVARIATE**), we see that:
  - *more* men taking the drug recover than do men are not taking the drug, and
  - *more* women taking the drug recover than do women are not taking the drug!

We record the recovery rates of 700 patients who were given access to the drug. A total of 350 patients chose to take the drug and 350 patients did not. The results of the study are shown in **Table 1.1**.

Table 1.1 Results of a study into a new drug, with gender being taken into account

	Drug			No Drug		
	patients	recovered	% recovered	patients	recovered	% recovered
Men	87	81	93%	270	234	87%
Women	263	192	73%	80	55	69%
Combined data	350	273	78%	350	289	83%

Drug vs non-drug takers recovery rates:

- 93% vs 87% male
- 73% vs 69% female
- 78% vs 83% general population!

Should a doctor prescribe the drug; to whom?

Should a policy maker approve the drug for use?

Simpson's Paradox

- Named after Edward Simpson (born 1922), statistician
- A group of sick patients are given the option to try a new drug (**TREATMENT**)
- Among those who took the drug (**TREATMENT GROUP**), a lower percentage recovered (**OUTCOME**) than among those who did not (**CONTROL GROUP**)
- However, when we partition by gender (**COVARIATE**), we see that:
  - *more* men taking the drug recover than do men are not taking the drug, and
  - *more* women taking the drug recover than do women are not taking the drug!

We record the recovery rates of 700 patients who were given access to the drug. A total of 350 patients chose to take the drug and 350 patients did not. The results of the study are shown in **Table 1.1**.

Table 1.1 Results of a study into a new drug, with gender being taken into account

	Drug			No Drug		
	patients	recovered	% recovered	patients	recovered	% recovered
Men	87	81	93%	270	234	87%
Women	263	192	73%	80	55	69%
Combined data	350	273	78%	350	289	83%

Drug vs non-drug takers recovery rates:

- 93% vs 87% male
- 73% vs 69% female
- 78% vs 83% general population!

Should a doctor prescribe the drug; to whom?

Should a policy maker approve the drug for use?

Understand the causal story behind the data

- What mechanism generated the data?
- Suppose estrogen has a negative effect on recovery
  - women less likely to recover than men, regardless of the drug

From the data:

Table 1.1 Results of a study into a new drug, with gender being taken into account

	Drug			No Drug		
	patients	recovered	% recovered	patients	recovered	% recovered
Men	87	81	93%	270	234	87%
Women	263	192	73%	80	55	69%
Combined data	350	273	78%	350	289	83%

Conclusion: the drug appears to be harmful but it is not

- If we select a drug taker at random, that person is more likely to be a woman
- Hence less likely to recover than a random person who doesn't take the drug

Causal Story

- Being a woman is a common cause of both drug taking and failure to recover.
- To assess the effectiveness we need to compare subjects of the same gender.

(Ensures that any difference in recovery rates is not ascribable to estrogen)


I’m back

TEST 2

- Consider a drug affecting recovery by lowering blood pressure (BP)
- Unfortunately, it has also a toxic effect

	No Drug		Drug	
	patients	recovered	patients	recovered
Low BP	87	81	270	234
High BP	263	192	80	55
Combined data	350	273	350	289

Should a doctor prescribe this drug or not?

Data Segregation

- We have solved the problem using gender-segregated data
- Then let's just segregate the data whenever possible, right?

WRONG!!!

- Consider a drug affecting recovery by lowering blood pressure (BP)
- Unfortunately, it has also a toxic effect

Table 1.2 Results of a study into a new drug, with posttreatment blood pressure taken into account

	No Drug			Drug		
	patients	recovered	% recovered	patients	recovered	% recovered
Low BP	87	81	93%	270	234	87%
High BP	263	192	73%	80	55	69%
Combined data	350	273	78%	350	289	83%

Should a doctor prescribe this drug or not?

Data Segregation

- We have solved the problem using gender-segregated data
- Then let's just segregate the data whenever possible, right?

WRONG!!!

- Consider a drug affecting recovery by lowering blood pressure (BP)
- Unfortunately, it has also a toxic effect

Table 1.2 Results of a study into a new drug, with posttreatment blood pressure taken into account

	No Drug			Drug		
	patients	recovered	% recovered	patients	recovered	% recovered
Low BP	87	81	93%	270	234	87%
High BP	263	192	73%	80	55	69%
Combined data	350	273	78%	350	289	83%

Should a doctor prescribe this drug or not?

Once again, the answer follows from the way the data were generated.

Data Segregation

- We have solved the problem using gender-segregated data
- Then let's just segregate the data whenever possible, right?

WRONG!!!

- Consider a drug affecting recovery by lowering blood pressure (BP)
- Unfortunately, it has also a toxic effect

Table 1.2 Results of a study into a new drug, with posttreatment blood pressure taken into account

	No Drug			Drug		
	patients	recovered	% recovered	patients	recovered	% recovered
Low BP	87	81	93%	270	234	87%
High BP	263	192	73%	80	55	69%
Combined data	350	273	78%	350	289	83%

Should a doctor prescribe this drug or not?

Once again, the answer follows from the way the data were generated.

- In the general population, the drug might improve recovery rates because of its effect on blood pressure.

Data Segregation

- We have solved the problem using gender-segregated data
- Then let's just segregate the data whenever possible, right?

WRONG!!!

- Consider a drug affecting recovery by lowering blood pressure (BP)
- Unfortunately, it has also a toxic effect

Table 1.2 Results of a study into a new drug, with posttreatment blood pressure taken into account

	No Drug			Drug		
	patients	recovered	% recovered	patients	recovered	% recovered
Low BP	87	81	93%	270	234	87%
High BP	263	192	73%	80	55	69%
Combined data	350	273	78%	350	289	83%

Should a doctor prescribe this drug or not?

Once again, the answer follows from the way the data were generated.

- In the general population, the drug might improve recovery rates because of its effect on blood pressure.
- in subpopulations—the group of people whose posttreatment BP is high and the group whose posttreatment BP is low—we, of course, would not see that effect; we would only see the drug's toxic effect.

Data Segregation

- We have solved the problem using gender-segregated data
- Then let's just segregate the data whenever possible, right?

WRONG!!!

- Consider a drug affecting recovery by lowering blood pressure (BP)
- Unfortunately, it has also a toxic effect

Table 1.2 Results of a study into a new drug, with posttreatment blood pressure taken into account

	No Drug			Drug		
	patients	recovered	% recovered	patients	recovered	% recovered
Low BP	87	81	93%	270	234	87%
High BP	263	192	73%	80	55	69%
Combined data	350	273	78%	350	289	83%

Should a doctor prescribe this drug or not?

Once again, the answer follows from the way the data were generated.

- In the general population, the drug might improve recovery rates because of its effect on blood pressure.
- in subpopulations—the group of people whose posttreatment BP is high and the group whose posttreatment BP is low—we, of course, would not see that effect; we would only see the drug's toxic effect.

Data Segregation

- We have solved the problem using gender-segregated data
- Then let's just segregate the data whenever possible, right?

WRONG!!!

- Consider a drug affecting recovery by lowering blood pressure (BP)
- Unfortunately, it has also a toxic effect

Table 1.2 Results of a study into a new drug, with posttreatment blood pressure taken into account

	No Drug			Drug		
	patients	recovered	% recovered	patients	recovered	% recovered
Low BP	87	81	93%	270	234	87%
High BP	263	192	73%	80	55	69%
Combined data	350	273	78%	350	289	83%

Should a doctor prescribe this drug or not? **YES**

- Only by BP-segregating the data we can see the toxic effect
- It makes no sense to segregate the data; we should use the combined data

Data Segregation

- We have solved the problem using gender-segregated data
- Then let's just segregate the data whenever possible, right?

WRONG!!!

- Consider a drug affecting recovery by lowering blood pressure (BP)
- Unfortunately, it has also a toxic effect

Table 1.1 Results of a study into a new drug, with gender being taken into account

	Drug			No Drug		
	patients	recovered	% recovered	patients	recovered	% recovered
Men	87	81	93%	270	234	87%
Women	263	192	73%	80	55	69%
Combined data	350	273	78%	350	289	83%

Table 1.2 Results of a study into a new drug, with posttreatment blood pressure taken into account

	No Drug			Drug		
	patients	recovered	% recovered	patients	recovered	% recovered
Low BP	87	81	93%	270	234	87%
High BP	263	192	73%	80	55	69%
Combined data	350	273	78%	350	289	83%

Note that the data are the same of Simpson's paradox.

Lessons Learned

- Information that allowed us to make a correct decision
  - the timing of the measurements
  - that the treatment affects blood pressure
  - that blood pressure affects recovery
- All this information was not in the data
  - as statisticians rightly say, **CORRELATION IS NOT CAUSATION**
  - hence there is no method that can determine the causal story from data alone
  - whence no ML method can aid in our decision
- The same holds for Simpson’s paradox
  - the paradox arises out of our conviction that treatment cannot affect sex
  - if it could, we could explain it as in our blood pressure case
  - but we cannot test the assumption using the data

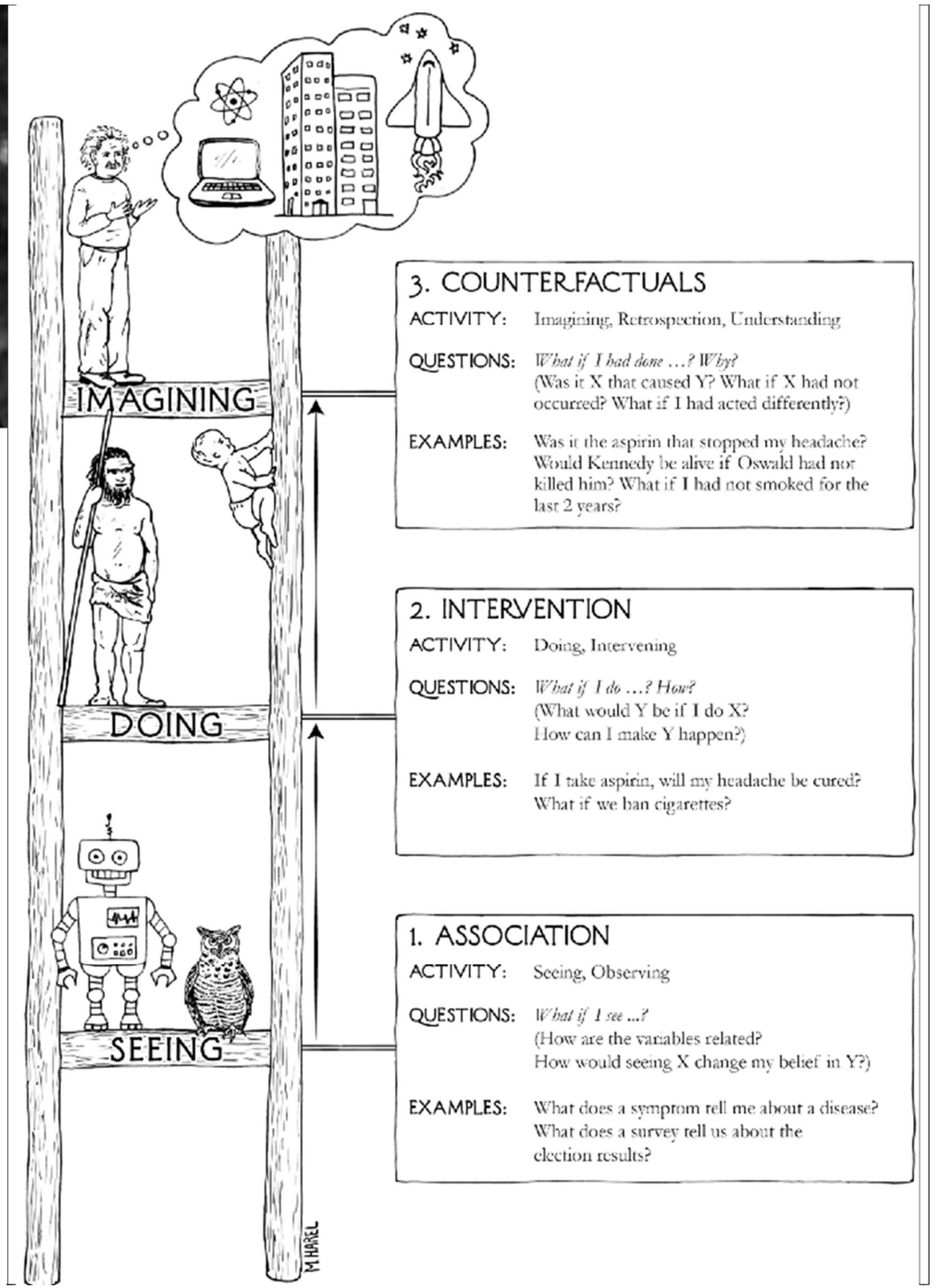
JUDEA PEARL

WINNER OF THE TURING AWARD

AND DANA MACKENZIE

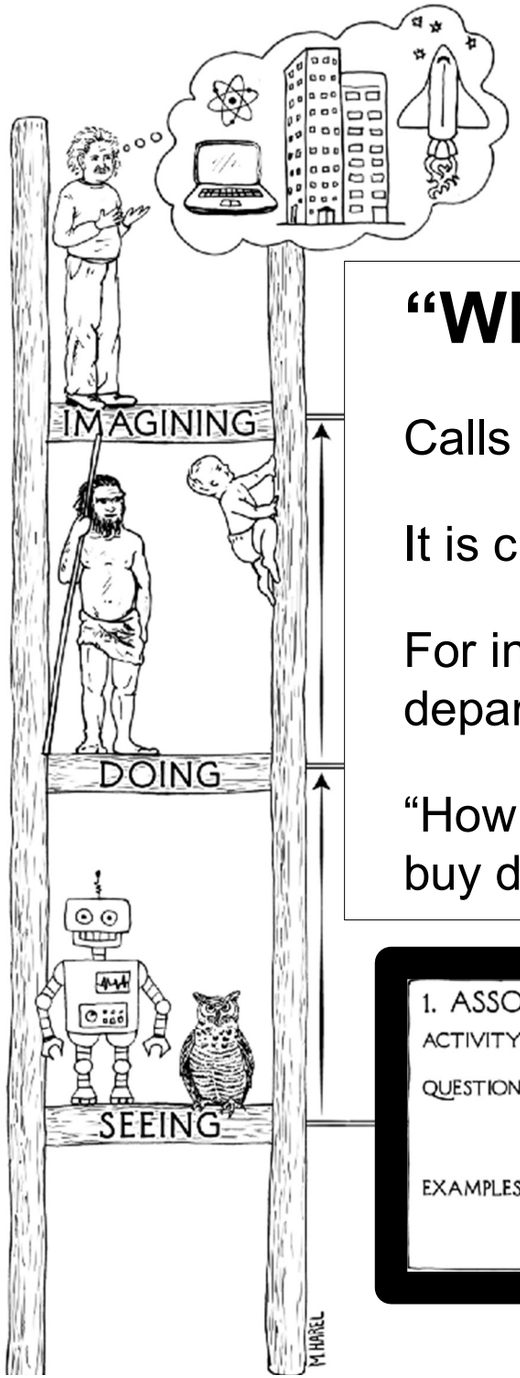
THE BOOK OF WHY

THE NEW SCIENCE OF CAUSE AND EFFECT



The Ladder of Causation

Seeing; we are looking for regularities in observations.



“What if I see …?”

Calls for predictions based on passive observations.

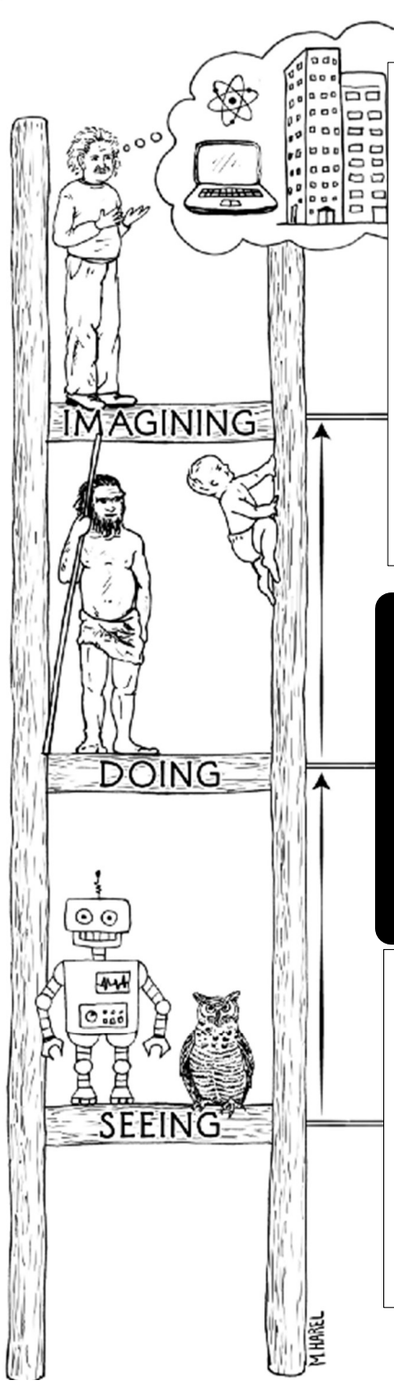
It is characterized by the question “What if I see …?”

For instance, imagine a marketing director at a department store who asks,

“How likely is a customer who bought toothpaste to also buy dental floss?”

1. ASSOCIATION	
ACTIVITY:	Seeing, Observing
QUESTIONS:	*What if I see ...?* (How are the variables related? How would seeing X change my belief in Y?)
EXAMPLES:	What does a symptom tell me about a disease? What does a survey tell us about the election results?

Intervention; ranks higher than association because it involves not just seeing but changing what is.



"What if do …?" & "How?"

We step up to the next level of causal queries when we begin to change the world. A typical question for this level is

"What will happen to our floss sales if we double the price of toothpaste?"

2. INTERVENTION

ACTIVITY: Doing, Intervening

QUESTIONS: *What if I do …? How?*
(What would Y be if I do X?
How can I make Y happen?)

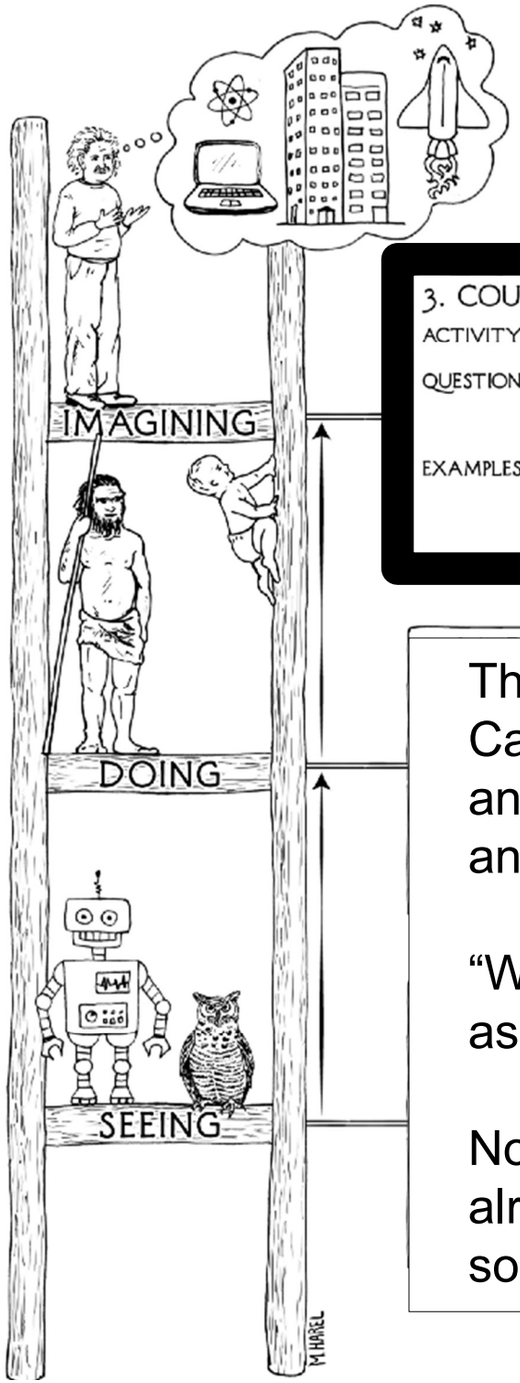
EXAMPLES: If I take aspirin, will my headache be cured?
What if we ban cigarettes?

This already calls for a new kind of knowledge, absent from the data, which we find at rung two of the Ladder of Causation, **Intervention**.

Many scientists have been quite traumatized to learn that none of the methods they learned in statistics is sufficient even to articulate, let alone answer, a simple question like

"What happens if we double the price?"

Counterfactuals; ranks higher than intervention because it involves **imagining**, **retrospection** and **understanding**.



"What if I had done …?" & "Why?"

3. COUNTERFACTUALS	
ACTIVITY:	Imagining, Retrospection, Understanding
QUESTIONS:	*What if I had done …? Why?* (Was it X that caused Y? What if X had not occurred? What if I had acted differently?)
EXAMPLES:	Was it the aspirin that stopped my headache? Would Kennedy be alive if Oswald had not killed him? What if I had not smoked for the last 2 years?

We might wonder, My headache is gone now, but

- Why?
- Was it the aspirin I took?
- The food I ate?
- The good news I heard?

These queries take us to the top rung of the Ladder of Causation, the level of **Counterfactuals**, because to answer them we must go back in time, change history, and ask,

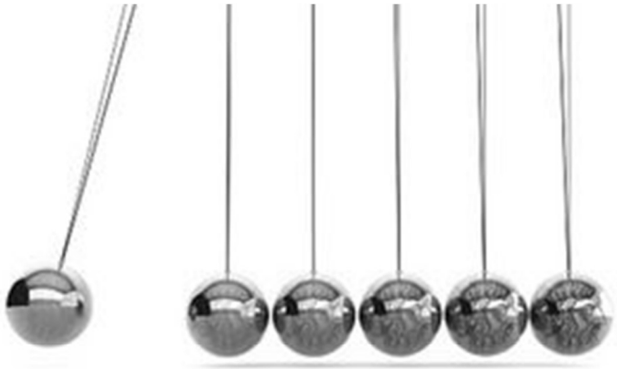
"What would have happened if I had not taken the aspirin?"

No experiment in the world can deny treatment to an already treated person and compare the two outcomes, so we must import a whole new kind of knowledge.

CONTENTS

CONTENTS

References



CAUSAL INFERENCE IN STATISTICS

A Primer

Judea Pearl
Madelyn Glymour
Nicholas P. Jewell

WILEY

DOI:10.1145/3271625

What just happened in artificial intelligence and how it is being misunderstood.

BY ADNAN DARWICHE

Human-Level Intelligence or Animal-Like Abilities?

"The vision systems of the eagle and the snake outperform everything that we can make in the laboratory, but snakes and eagles cannot build an eyeglass or a telescope or a microscope."

JUDEA PEARL

WINNER OF THE TURING AWARD

AND DANA MACKENZIE

THE BOOK OF WHY

THE NEW SCIENCE OF CAUSE AND EFFECT