Toronto Metropolitan University | TED ROGERS SCHOOL OF MANAGEMENT | Centre for Digital Enterprise Analytics and Leadership

# Business Analytics and Data Driven Decision Making

## Session#16: Lecture#31: Predictive Analytics: Machine Learning Cases


**Ravi Vatrapu**

*Director, Centre for Digital Enterprise Analytics and Leadership (DEAL)*

*Loretta Rogers Research Chair in Digital Enterprise*

*Professor,* Department of Information Technology Management

Ted Rogers School of Management

Toronto Metropolitan University, Canada

*Founding Director*, Centre for Business Data Analytics (bda.cbs.dk)

*Professor (on leave)*, Copenhagen Business School, Denmark

*Visiting Professor*, Indian Institute of Management Visakhapatnam, India

*Visiting Professor*, GITAM Deemed University, India

*Visiting Professor*, University of Surrey, United Kingdom

Email: vatrapu@torontomu.ca

DEAL Website: https://www.torontomu.ca/tedrogersschool/digital-enterprise-analytics-and-leadership/

Faculty *Web*: https://www.torontomu.ca/information-technology-management/faculty-research/ravi-vatrapu/


IIM VISAKHAPATNAM

GITAM SERVE STRIVE THRIVE DEEMED TO BE UNIVERSITY

UNIVERSITY OF SURREY

CBS CENTRE FOR BUSINESS DATA ANALYTICS TRANSFORMING BIG DATA SETS INTO BUSINESS ASSETS


*Acknowledgements: Dr. Morteza Zihayat*

# AI in Business: Introduction

TED ROGERS SCHOOL OF MANAGEMENT
Centre for Digital Enterprise Analytics and Leadership

# In-Class Exercise: Personal Experiences of AI in Business

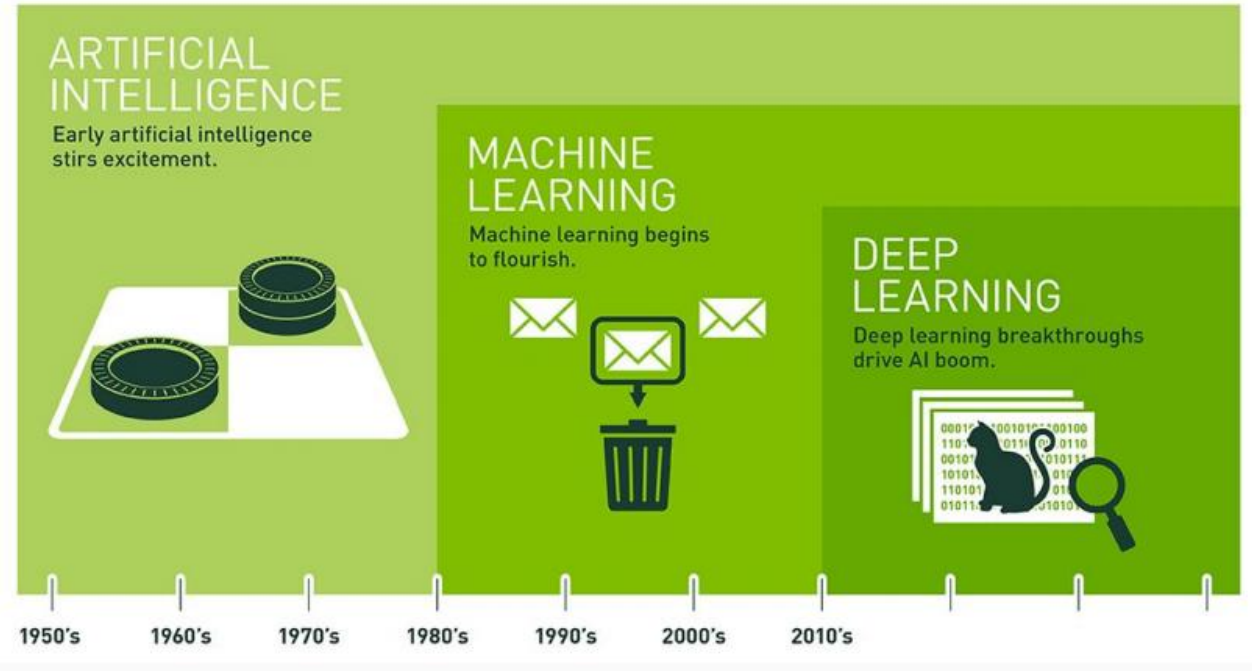
# Artificial Intelligence (AI)

Artificial intelligence is a machine that mimics a "cognitive" function of human mind

# The birth of AI: 1956

At the Dartmouth workshop of 1956's, Allen Newell, Herbert A. Simon and Cliff Shaw debuted with Logic Theorist

- Turing's test
- GameAI
  - In 1951, using the Ferranti Mark 1 machine of the University of Manchester, Christopher Strachey wrote a checkers program and Dietrich Prinz wrote one for chess



# The birth of AI

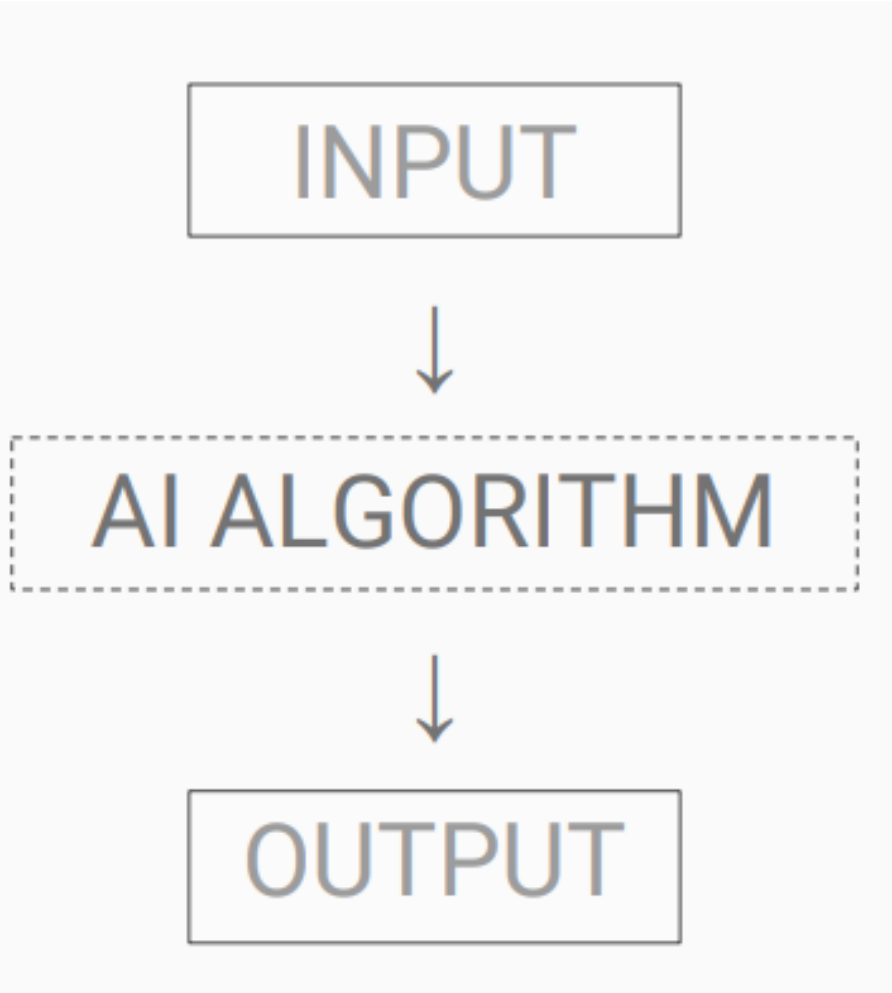
# Artificial intelligence

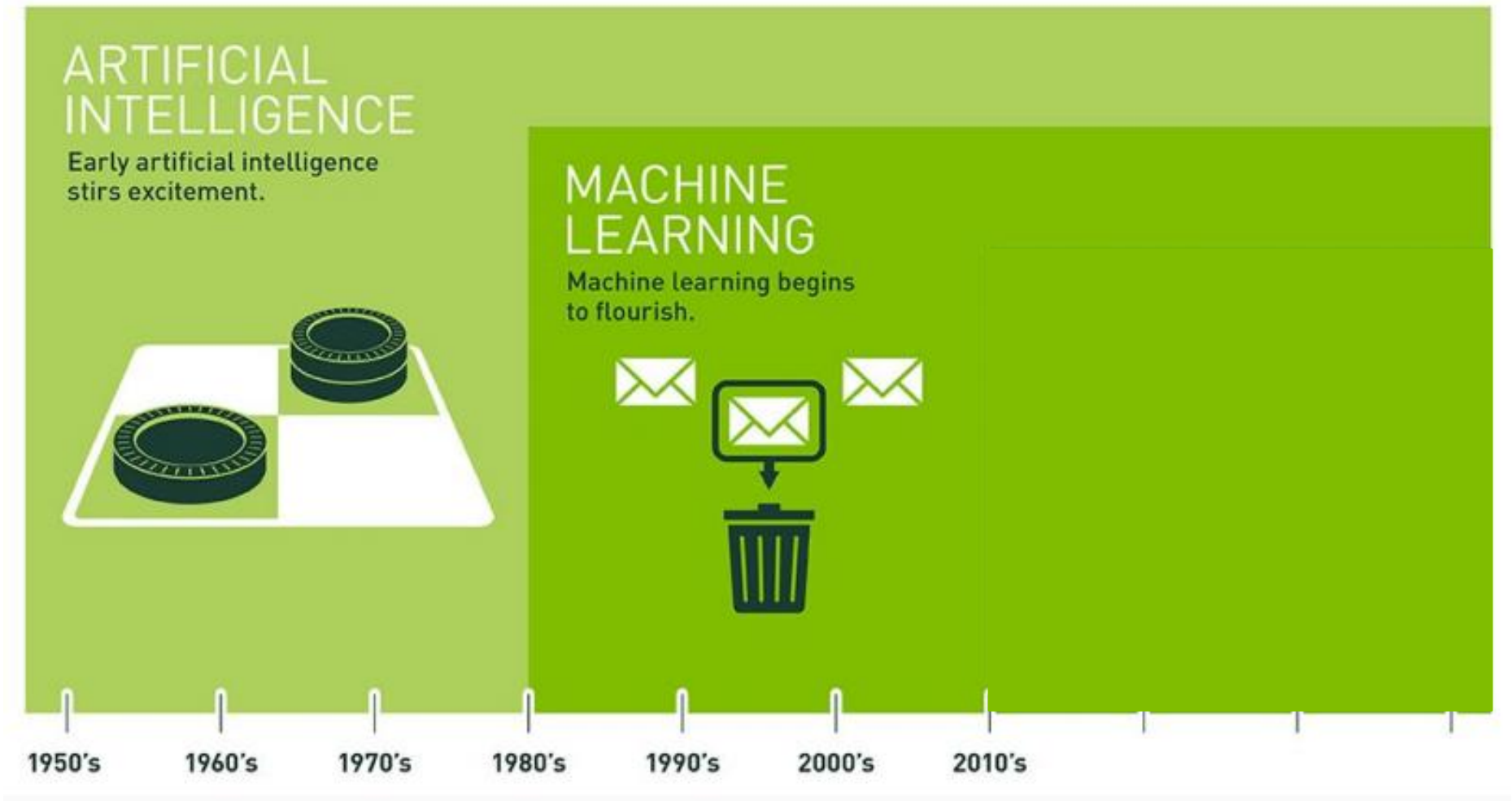
| | |
|---|---|
| "The exciting new effort to make computers think . . . *machines with minds,* in the full and literal sense" (Haugeland, 1985)<br>"[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning . . ." (Bellman, 1978) | "The study of mental faculties through the use of computational models" (Charniak and McDermott, 1985)<br>"The study of the computations that make it possible to perceive, reason, and act" (Winston, 1992) |
| "The art of creating machines that perform functions that require intelligence when performed by people" (Kurzweil, 1990)<br>"The study of how to make computers do things at which, at the moment, people are better" (Rich and Knight, 1991) | "A field of study that seeks to explain and emulate intelligent behavior in terms of computational processes" (Schalkoff, 1990)<br>"The branch of computer science that is concerned with the automation of intelligent behavior" (Luger and Stubblefield, 1993) |

Figure 1.1 Some definitions of AI. They are organized into four categories:

| | |
|---|---|
| Systems that think like humans. | Systems that think rationally. |
| Systems that act like humans. | Systems that act rationally. |

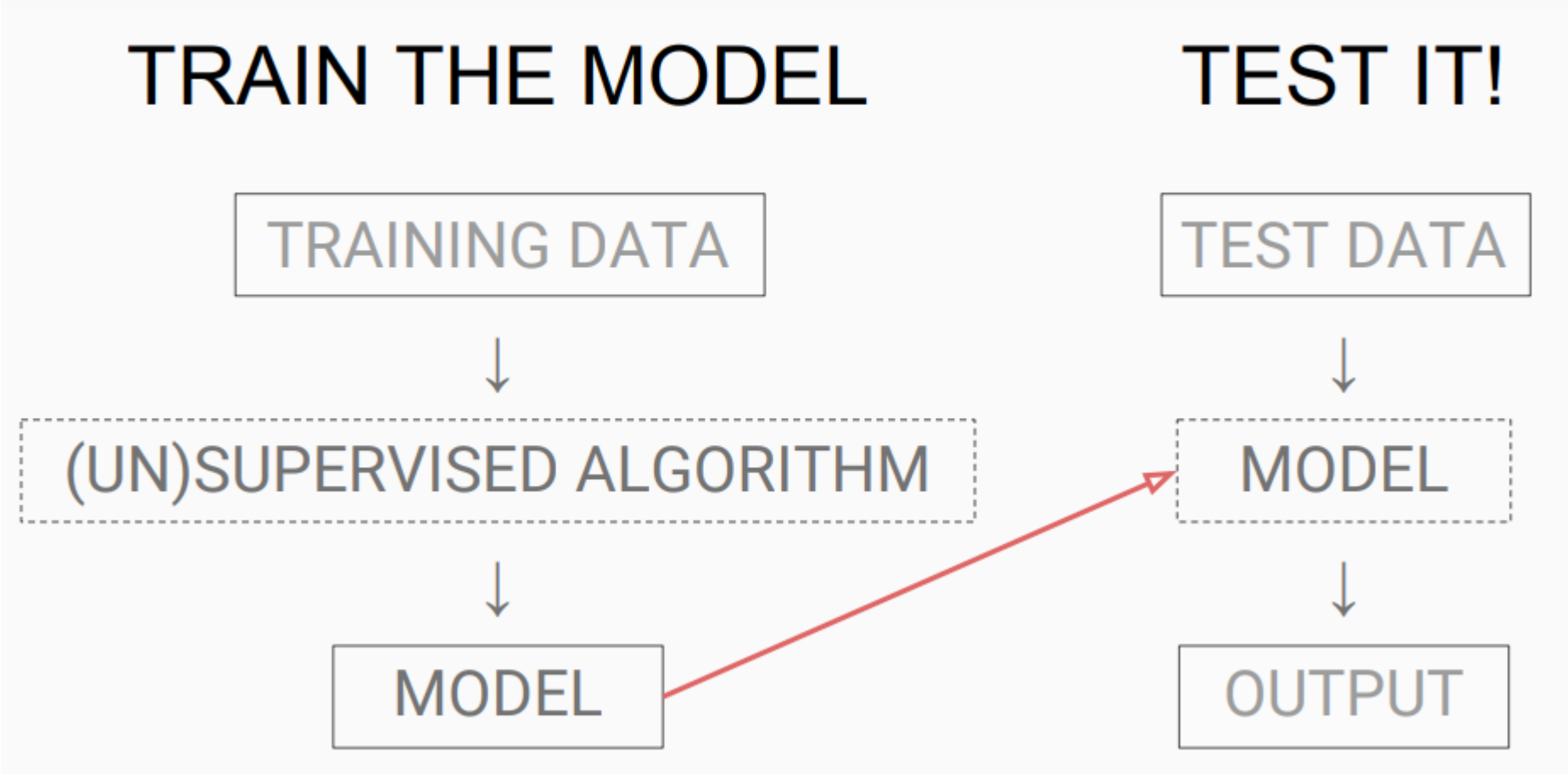
# Overview




ARTIFICIAL INTELLIGENCE
Early artificial intelligence stirs excitement.
MACHINE LEARNING
Machine learning begins to flourish.
1950's
1960's
1970's
1980's
1990's
2000's
2010's

# Machine Learning

Machine learning explores the study and construction of algorithms that can learn from and make predictions on data



# Refresh your mind first

# Terminology

## Model

- A simplified representation of reality created to serve a purpose

## Predictive Model

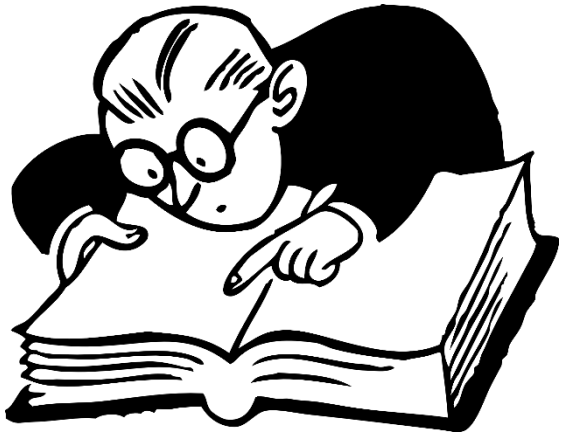
- A formula for estimating the unknown value of interest: **the target**
  - The formula can be mathematical, logical statement (e.g., rule), etc.

## Prediction

- Estimate an unknown value (i.e. the target)

## Instance / example

- Represents a fact or a data point
- Described by a set of **attributes** (fields, columns, variables, or features)



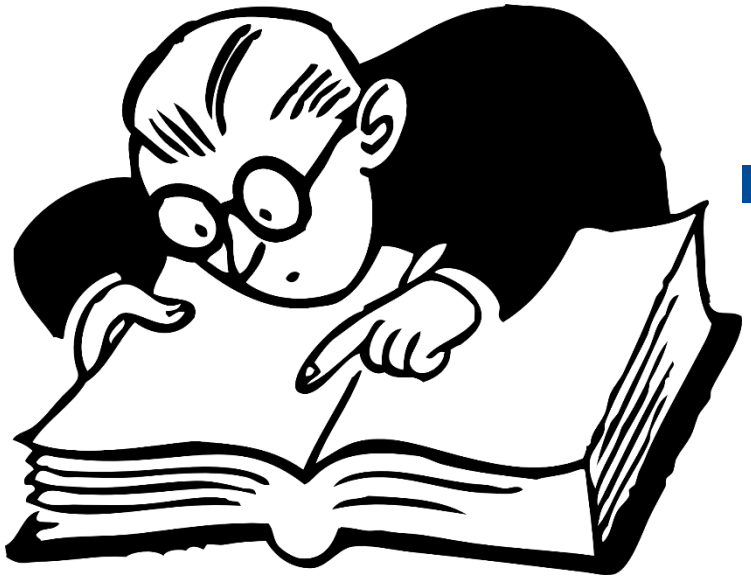
# Terminology

## Model induction:

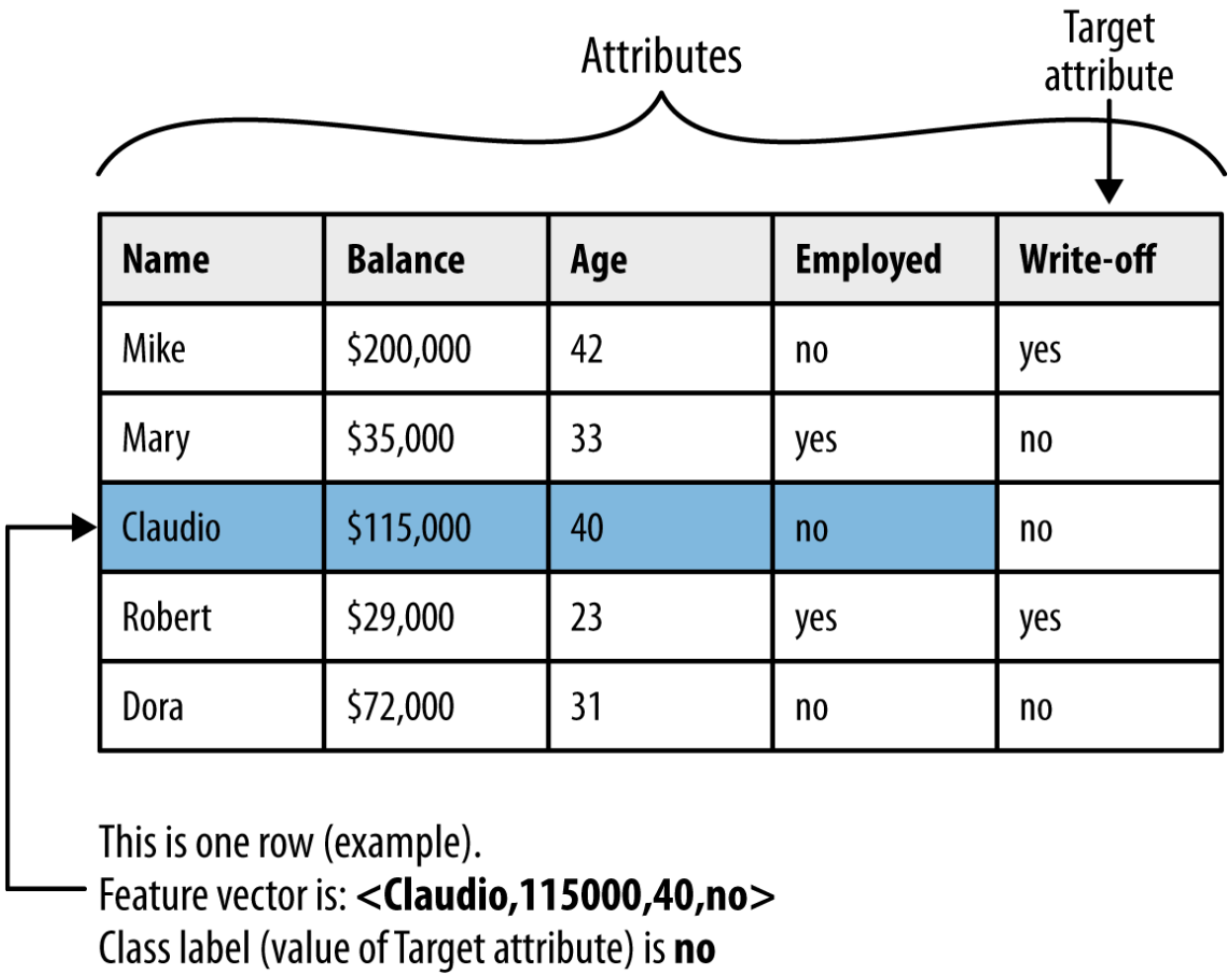
- The creation of **models** from data

## Training data:

- The input data for the induction algorithm



# Terminology




| Name | Balance | Age | Employed | Write-off |
|---|---|---|---|---|
| Mike | $200,000 | 42 | no | yes |
| Mary | $35,000 | 33 | yes | no |
| Claudio | $115,000 | 40 | no | no |
| Robert | $29,000 | 23 | yes | yes |
| Dora | $72,000 | 31 | no | no |

# What is a model?

A *simplified* representation* of reality created for a *specific purpose*

**based on some assumptions*

*Examples: map, prototype, Black-Scholes model, etc.*

*Data Mining example:*

- *"formula" for predicting probability of customer attrition at contract expiration*

→ *"classification model" or "class-probability estimation model"*

# Feature Types

**Numeric**: anything that has some order

- Numbers (that mean numbers)
- Dates (that look like numbers …)
- *Dimension is 1*

**Categorical**: stuff that does not have an order

- Text
- *Dimension is the number of possible values*

Food for thought: Names, Ratings

# Dimensionality of the data?

Attributes / Features

| Name | Balance | Age | Default |
|---|---|---|---|
| Mike | $123,000 | 30 | Yes |
| Mary | $51,100 | 40 | Yes |
| Bill | $68,000 | 55 | No |
| Jim | $74,000 | 46 | No |
| Mark | $23,000 | 47 | Yes |
| Anne | $100,000 | 49 | No |

**Dimensionality of a dataset** is the sum of the dimensions of the features

- The sum of the number of numeric features and ~ the number of values of categorical features

# Common Machine Learning Tasks

## **Classification** and class probability estimation

- **How likely** is this consumer to respond to our campaign?

## Regression

- **How much** will she use the service?

## Similarity Matching

- Can we find consumers **similar** to my best customers?

## Clustering

- Do my customers form **natural groups**?


## Co-occurrence Grouping and Association Rules

- Also known as **frequent itemset mining**, association rule discovery, and market-basket analysis
- What items are commonly purchased together?

# Common Machine learning Tasks

**Profiling** (behavior description)

- What does “normal behavior” look like? (for example, as baseline to detect fraud)

**Data Reduction**

- Which latent dimensions describe the consumer taste preferences?

**Link Prediction**

- Since John and Jane share 2 friends, should John become Jane’s friend?

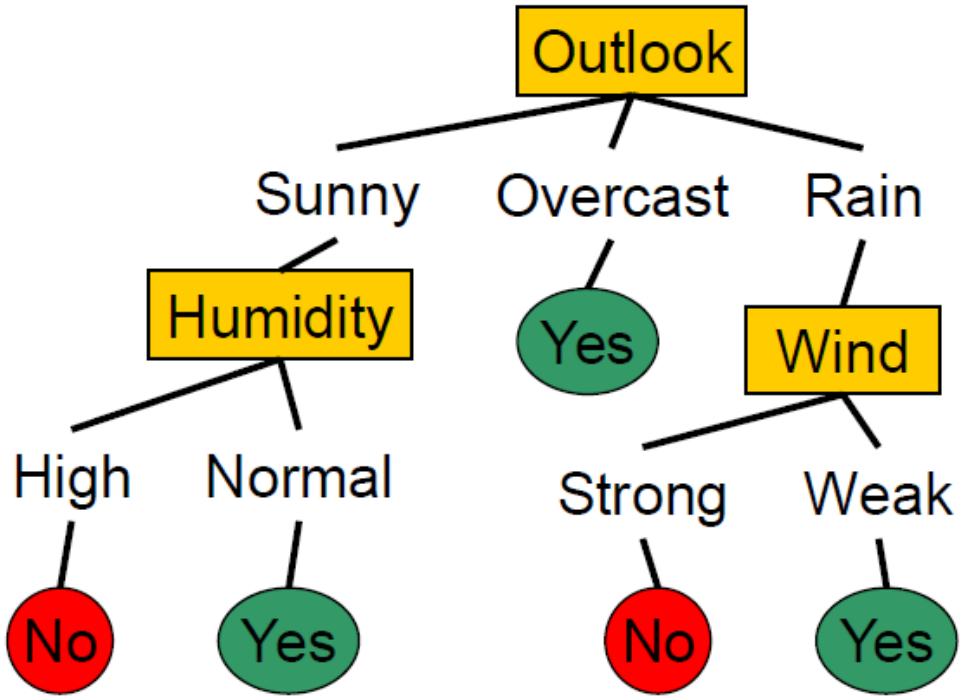
**Causal Modeling**

- Understand what events or actions ***actually influence*** others
  - We use predictive modeling to target advertisements to some customers
  - The targeted customers buy at a higher rate after receiving the ads
    - Was this because the ads influenced the customers to buy more?
    - Or this particular group of customers would have bought more anyway?!

# Classification Models

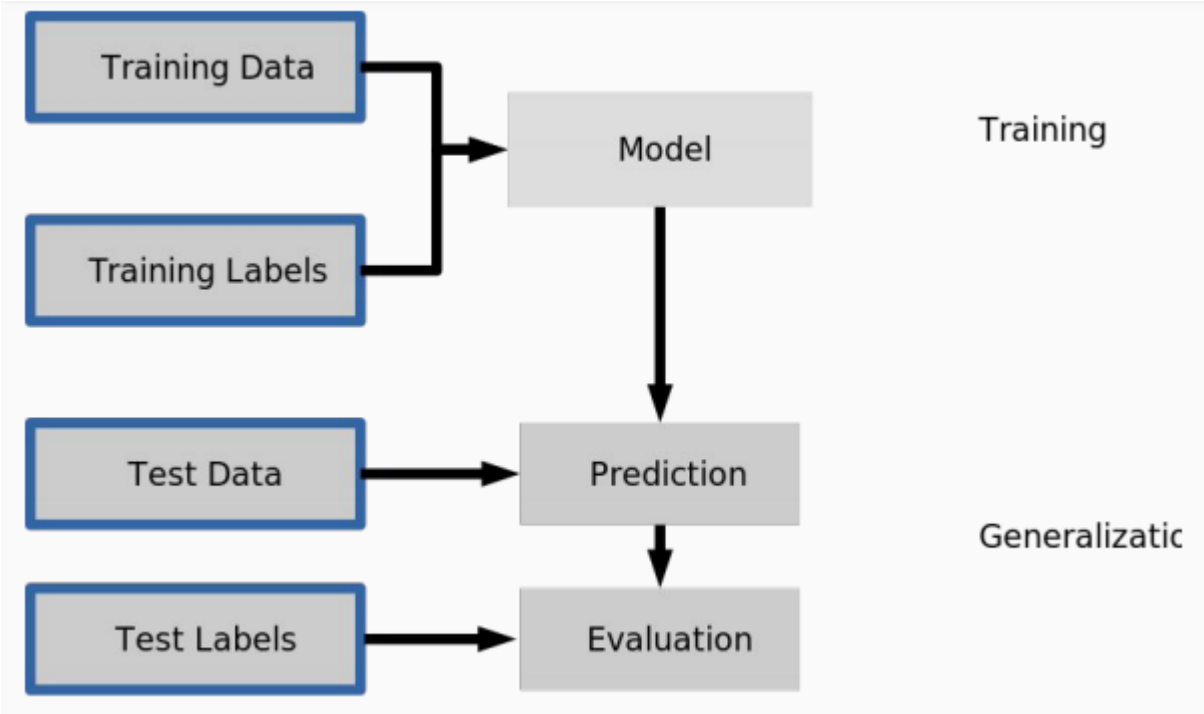
## A decision tree for *PlayTennis*

| Day | Outlook | Temp | Humid | Wind | PlayTennis |
|---|---|---|---|---|---|
| D1 | Sunny | Hot | High | Weak | No |
| D2 | Sunny | Hot | High | Strong | No |
| D3 | Overcast | Hot | High | Weak | Yes |
| D4 | Rain | Mild | High | Weak | Yes |
| D5 | Rain | Cool | Normal | Weak | Yes |
| D6 | Rain | Cool | Normal | Strong | No |
| D7 | Overcast | Cool | Normal | Strong | Yes |
| D8 | Sunny | Mild | High | Weak | No |
| D9 | Sunny | Cool | Normal | Weak | Yes |
| D10 | Rain | Mild | Normal | Weak | Yes |
| D11 | Sunny | Mild | Normal | Strong | Yes |
| D12 | Overcast | Mild | High | Strong | Yes |
| D13 | Overcast | Hot | Normal | Weak | Yes |
| D14 | Rain | Mild | High | Strong | No |



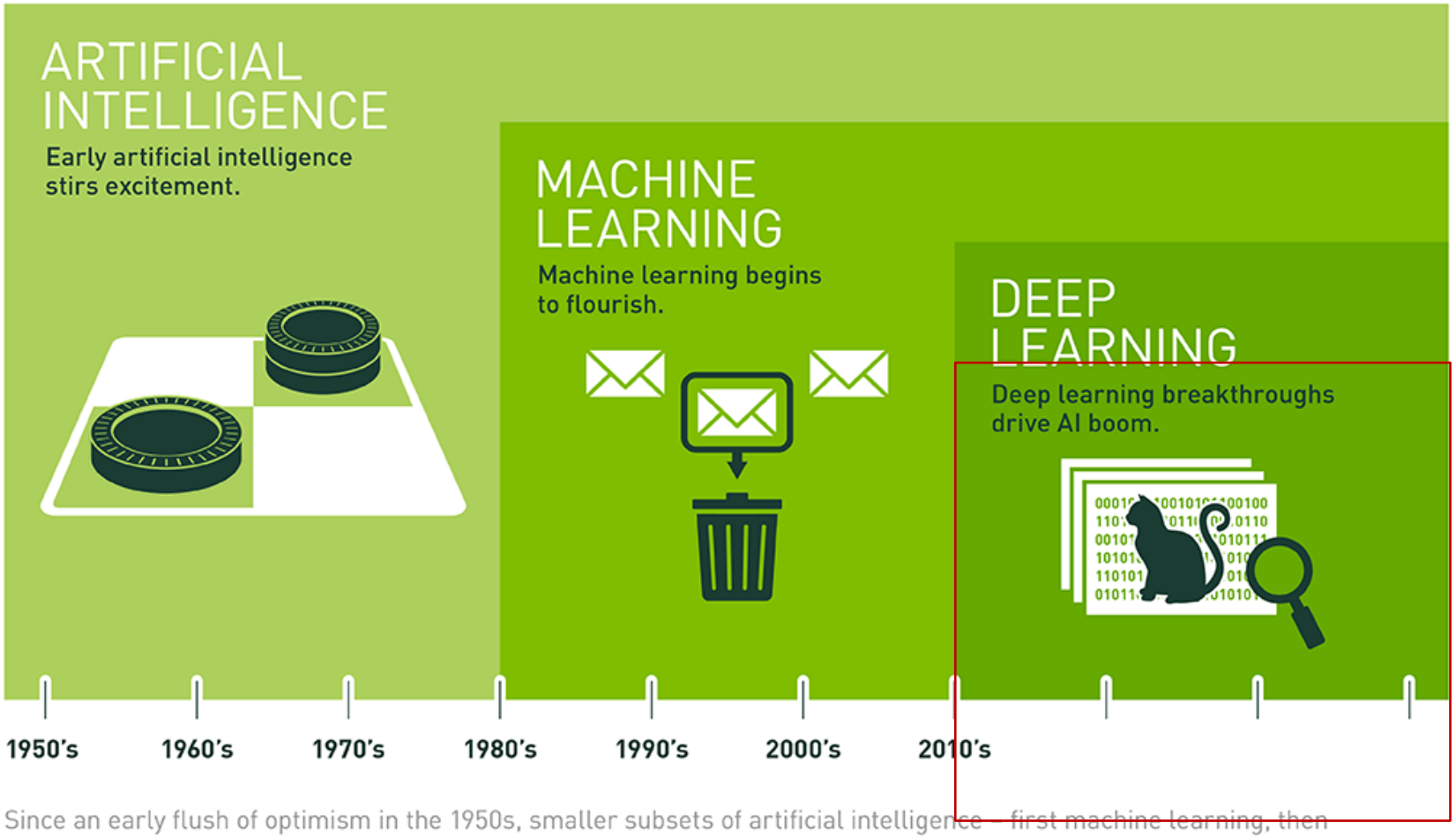
# Supervised learning

Supervised learning is the Data mining task of inferring a function from **labeled** training data.



# Unsupervised learning

Unsupervised machine learning is the machine learning task of inferring a function to describe hidden structure from "unlabeled" data




Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

Source: http://bit.ly/2sufGd

# Why Deep Learning?

## The world's best Dota 2 players just got destroyed by a killer AI from Elon Musk's startup

- Tonight, during Valve's yearly *Dota 2* tournament, a surprise segment introduced what could be the best new player in the world -- a bot from Elon Musk-backed startup OpenAI.
- https://www.theverge.com/2017/8/11/16137388/dota-2-dendi-open-ai-elon-musk

## Elon Musk-funded *Dota 2* bots spank top-tier humans, and they know how to trash talk

- A team of computer-controlled bots took on a team of top human players in a *Dota 2* show match this weekend, and the computer didn't just beat the humans—it *crushed* them.
- https://arstechnica.com/gaming/2018/08/elon-musks-dota-2-bots-spank-top-tier-humans-and-they-know-how-to-trash-talk/

# What is Deep Learning

Deep learning is a neural network with one or more hidden layers

Artificial neural networks are a computational model based on simple neural units

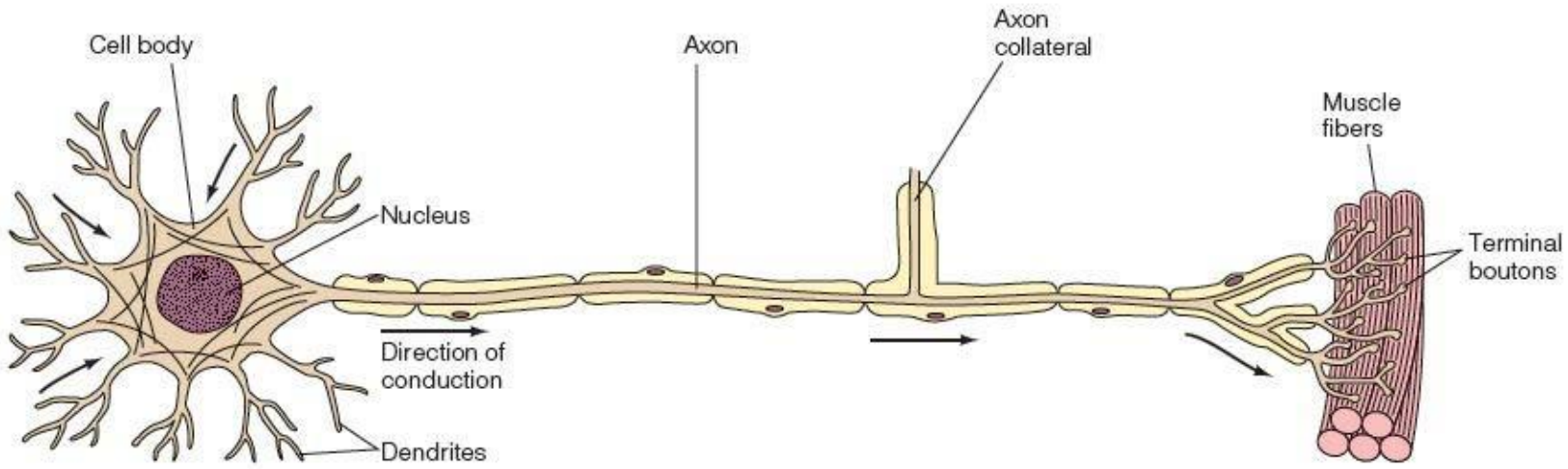
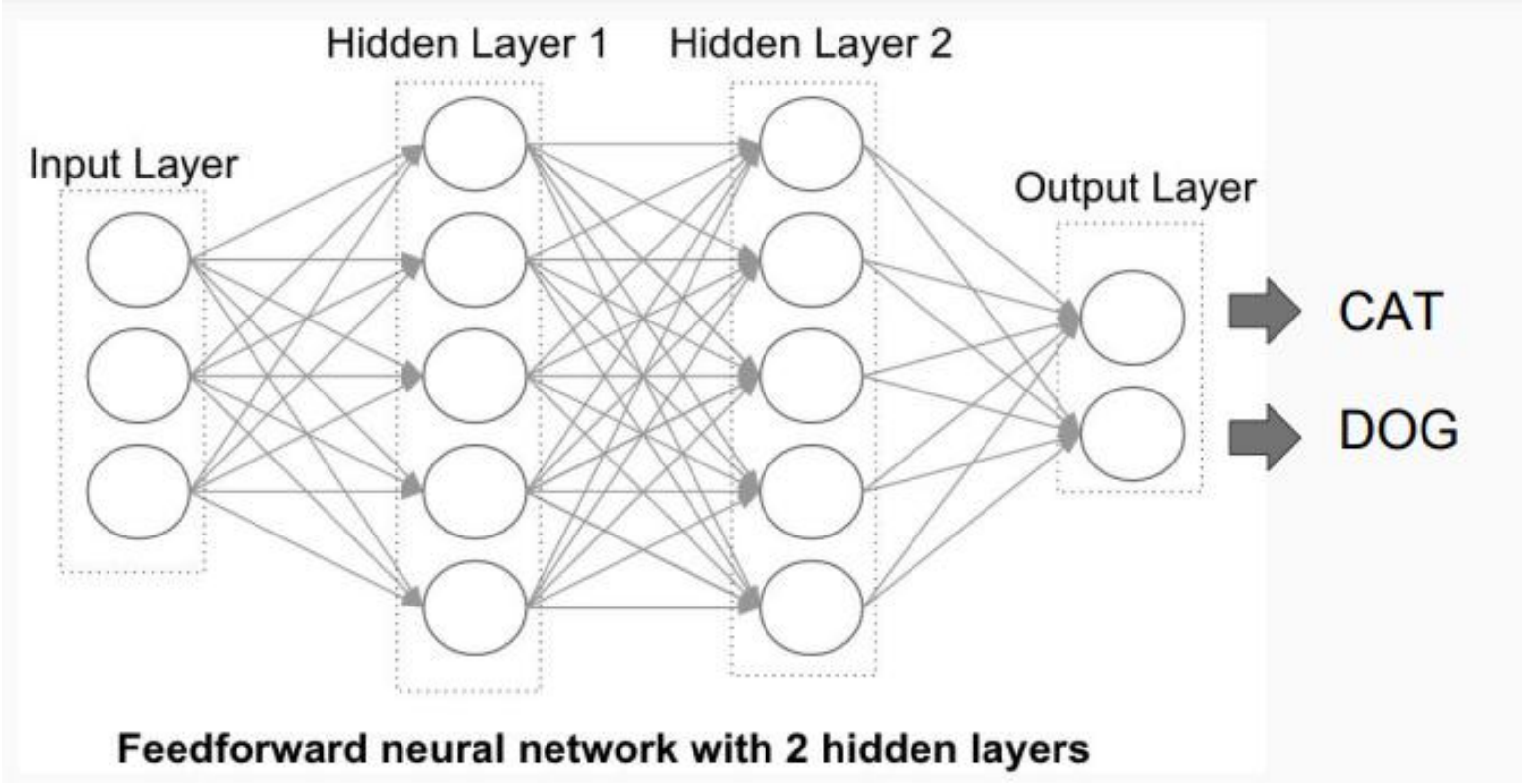



FIGURE 13.6 A biological neuron

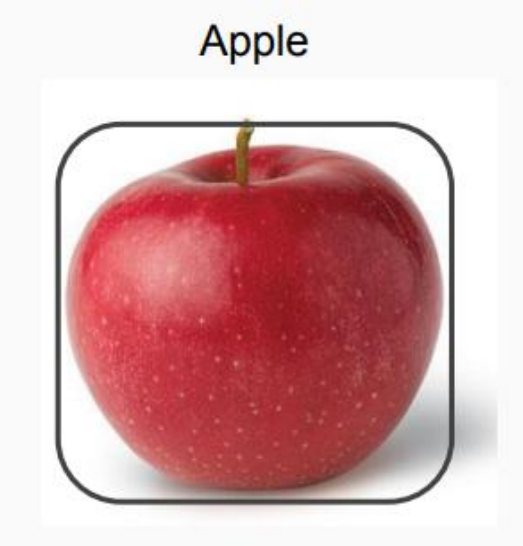
# Practical example!


**Feedforward neural network with 2 hidden layers**

# Practical example!

https://www.youtube.com/watch?v=Aut32pR5PQA

# Easy for Me, Hard for You

Computers are nothing more than calculators at heart.

Adding up numbers really quickly - thousands, or even millions, a second - may be impressive but it isn't artificial intelligence

- It simply requires an ability to follow very basic instructions

We can process the quite large amount of information that the images contain, and very successfully process it to recognize what's in the image

kind of task isn't easy for computers - in fact it's incredibly difficult.

# Human vs Computers

We suspect image recognition needs human intelligence - something machines lack, however complex and powerful we build them, because they're not human

| Problem | Computer | Human |
|---|---|---|
| Multiply thousands of large numbers quickly | Easy | Hard |
| Find faces in a photo of a crowd of people | Hard | Easy |

But it is exactly these kinds of problems that we want computers to get better at solving because they're fast and don't get tired

And it these kinds of hard problems that artificial intelligence is all about.

# Artificial Intelligence

- Artificial intelligence is to find new kinds of recipes, or **algorithms** , which work in new ways to try to solve these kinds of harder problem

- Even if not perfectly well, but well enough to give an impression of a human like intelligence at work.

# Recent and future developments in AI

A practically important variant of supervised learning is called "transfer learning."

A complex neural network can be trained with large amounts of data, so that it learns to
discern important features of the data

The trained network can then be re-used for different pattern recognition tasks, when the underpinning features are similar enough.

Example

- A network can be trained to label human faces with millions of images
- When the network has learned to recognize the faces that have been used for its training, its deep layers become optimized for face recognition
- The top levels of the network can then relatively easily be trained to detect new faces that the system has not seen before

This drastically reduces the computational and data requirements.

# Supervised Learning Limitation

Supervised learning methods usually give more accurate predication

What is the limitation then?

- They can only see the world as a repetition of the past

Supervised learning models materialize and hardwire cultural beliefs that often remain otherwise unexplored

Supervised learning creates machines that are only able to perceive worlds where humans are put in pre-defined boxes

# Supervised Learning

Supervised learning assumes that we already know what categories input patterns can represent.

- For example, a self-driving car needs to know whether an object is a cyclist, truck, a train, or a child.
- Technically, supervised learning creates machines that map input patterns into a collection of output classes

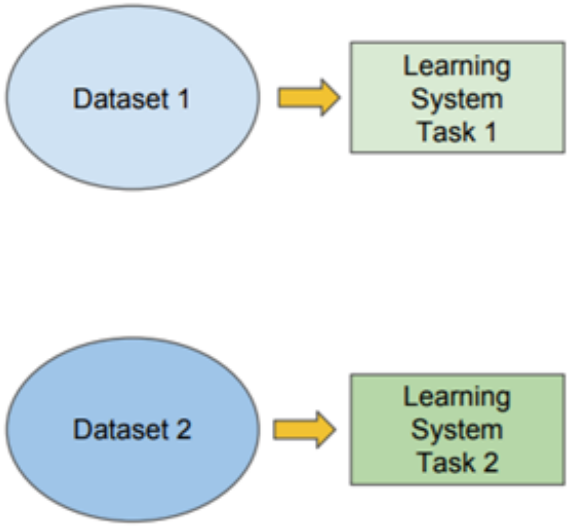
Their intelligence is similar to simplest living beings that can associate environmental conditions with learned behaviours.

In psychology, these learning models underpin the *Pavlovian theory* of reflexes and, for example, Skinnerian reinforcement learning

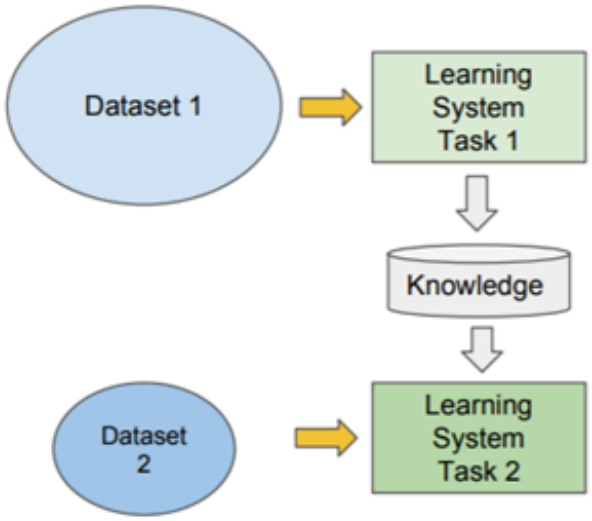
- This type of learning represents the developmentally **simplest model of learning**, and both pigeons and humans are well capable of it

# Traditional ML vs Transfer Learning

- Isolated, single task learning:
  - Knowledge is not retained or accumulated. Learning is performed w.o. considering past learned knowledge in other tasks




- Learning of a new tasks relies on the previous learned tasks:
  - Learning process can be faster, more accurate and/or need less training data




https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep-learning-212bf3b2f27a

# Reinforcement learning

Increasing computational power has also allowed researchers to use simple pattern matching networks as components in higher-level architectures

- Google's *AlphaZero* uses Reinforcement learning
  - System generates game simulations and adjusts network weights based on success in these games

Reinforcement learning amplifies behavior that leads to outcomes that are defined as positive

A variant of reinforcement learning is known as generative adversarial networks, or GANs

- One network tries to fool another to believe that the data it generates actually comes from the training data set
- Create synthetic images of artworks and human faces that an image recognition system cannot distinguish from real images
- It is also commercially used for product design, for example in the fashion industry.

https://www.youtube.com/watch?time_continue=13&v=QilHGSYbjDQ&feature=emb_logo

# Discussion

# Applications of AI in Finance

# *38% of companies already use AI in some capacity*

# Three main areas

1. Fraud detection and compliance
2. Banking chatbots and robo-advisory services
3. Algorithmic trading

# Fraud detection and compliance: Challenges

Fraud is a **latent variable**,

- It is not directly observable but must be inferred from data

Being too cautious about each transaction is costly!

- It is more challenging for AI algorithms to make accurate predictions of possible fraud than shopping decisions
  - **Transactions wrongly declined due to suspected fraud account for $118 billion in retail losses**

In the case of money laundering detection, a large number of false positive results are often created.

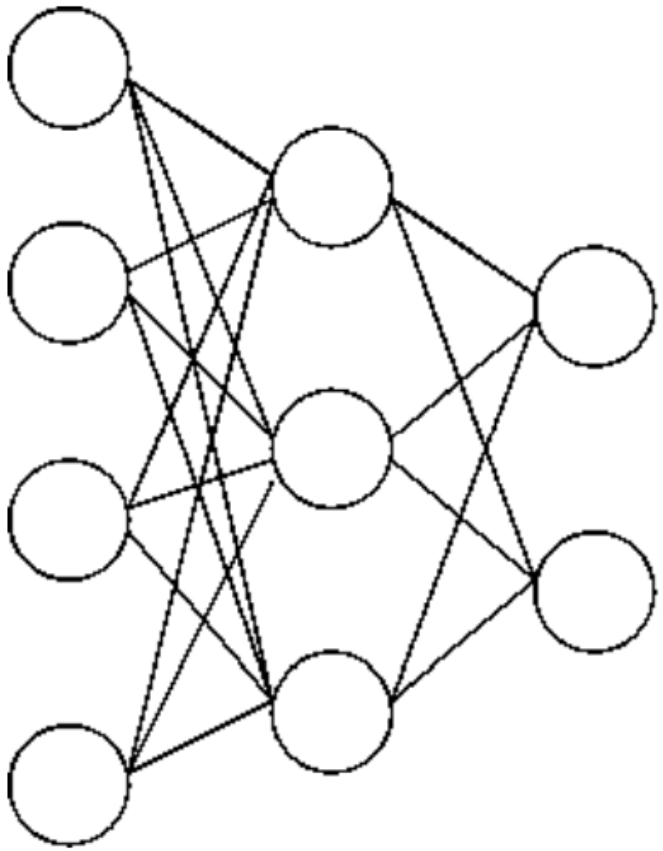
- ML in the anti-money laundering area has proved more challenging
- There is a lack of large public datasets that can be used to analyse money laundering prediction
- Conventional rules-based systems may not be sophisticated enough to detect sophisticated transactions

Age, Income, Occupation, of card Holder

Number of large purchases on the card

Frequency of large purchases

Location where large purchase took



Transaction OK

Transaction probably fraudulent

**Input Layer** **Hidden Layer** **Output Layer**

# Chatbots and Banking Industry

A robo-advisor is an algorithm based digital platform that offers automated financial advice or investment management services.

- Robo-advisors are algorithms built to calibrate a financial portfolio to the user's goals and risk tolerance
- Significant appeal with millennial consumers who do not need a physical advisor to feel comfortable investing

## Plum

- a chatbot that can be accessed through Facebook Messenger and helps a customer save money in small increments

Plum is connected to the customer's bank account and its AI engine then analyses customer income and spending history and then predicts how much they can afford to save

# Algorithmic trading

Also called “Automated Trading Systems,”

- Algorithmic trading is about implementing trading rules into a program
- Using the program to trade, [and AI trading] can be defined as an approach to machine learning that learns the structure of the data, and then tries to predict what will happen.

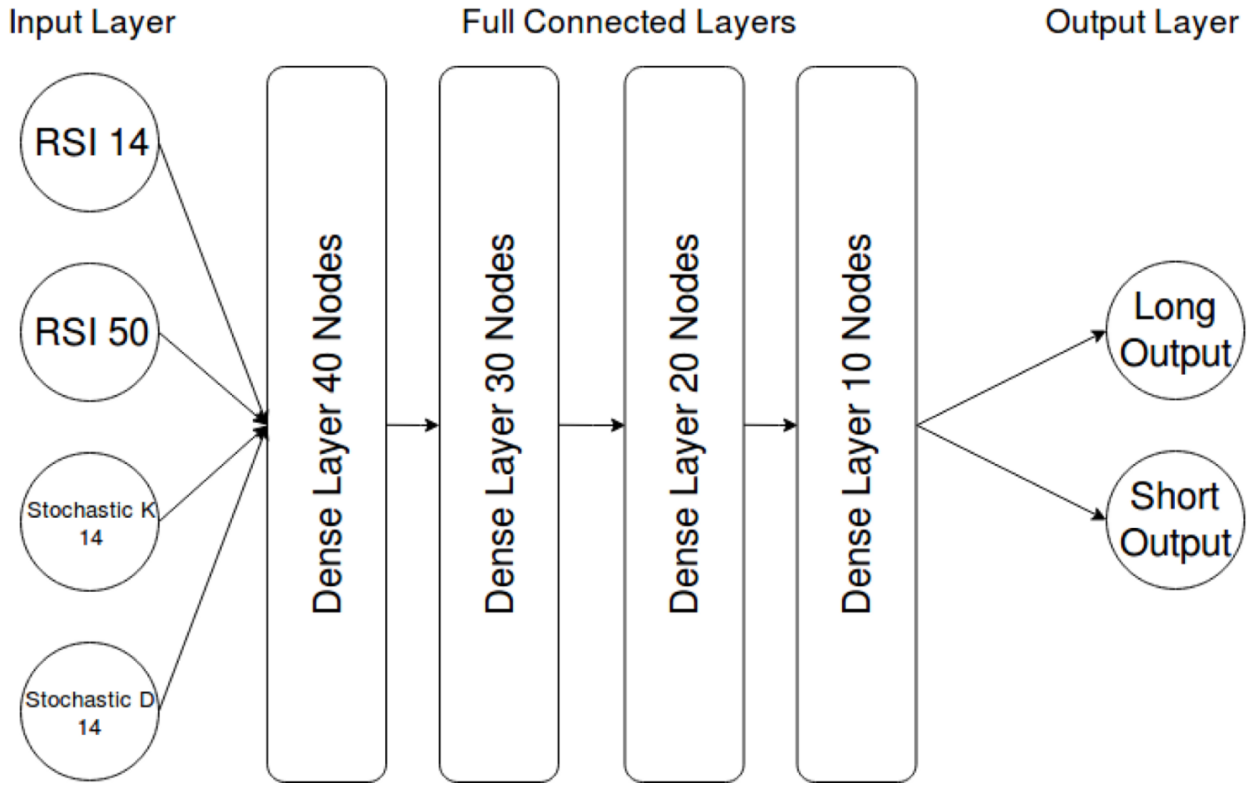
Computers generate 50-70% of equity market trades, 60% of futures trades and 50% of Treasuries

Benefits of AT

- The ability of trades to be executed at the best possible prices
- Increased accuracy and a reduced likelihood of mistakes
- The ability to automatically and simultaneously check multiple market conditions
- Human errors caused by psychological or emotional conditions are likely to be reduced.

# Neural Network Modeling

The “dense” layers within the architecture mean that each neuron is connected to the outputs of all the neurons in the layer below



# Applications of AI in Healthcare

# Motivation

Sophisticated algorithms to ‘learn’ features from a large volume of healthcare data

- Use the obtained insights to assist clinical practice

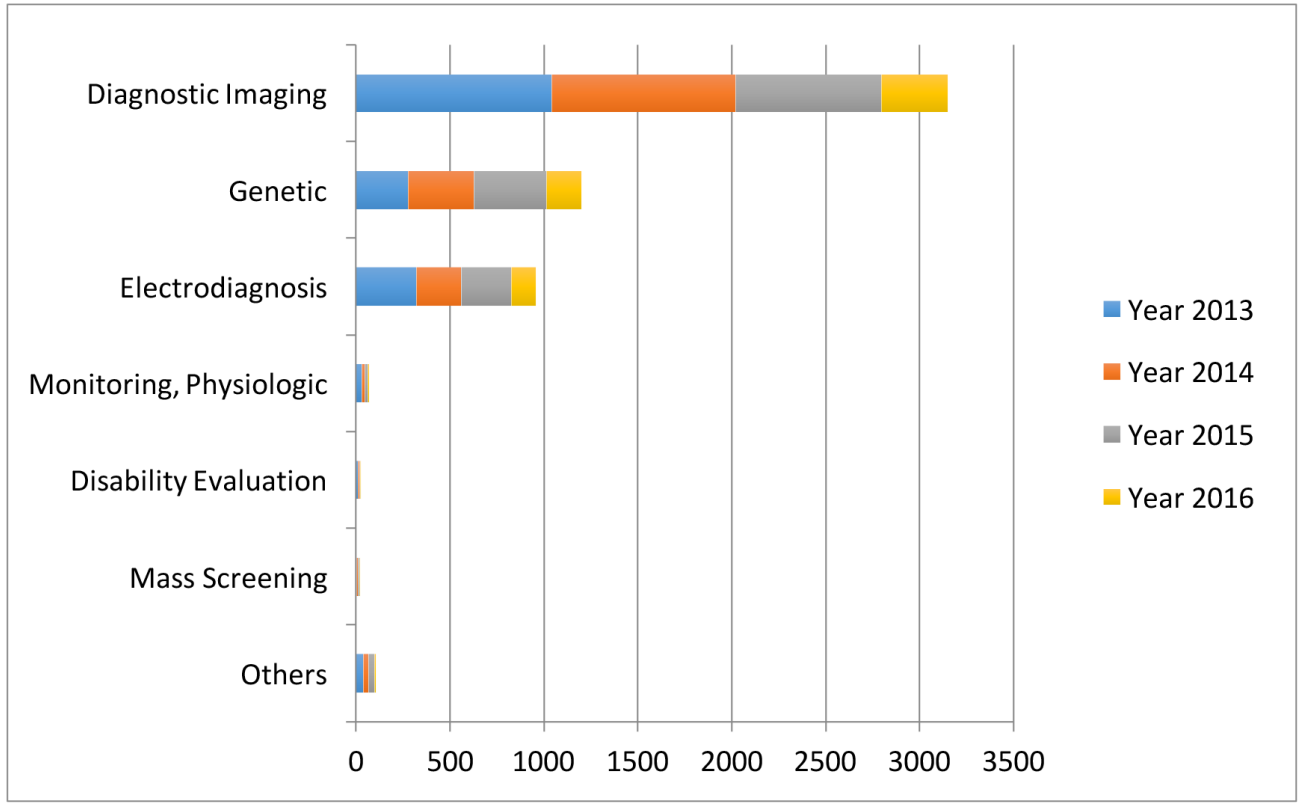
It can be equipped with **learning** and **self-correcting abilities** to improve its accuracy based on feedback

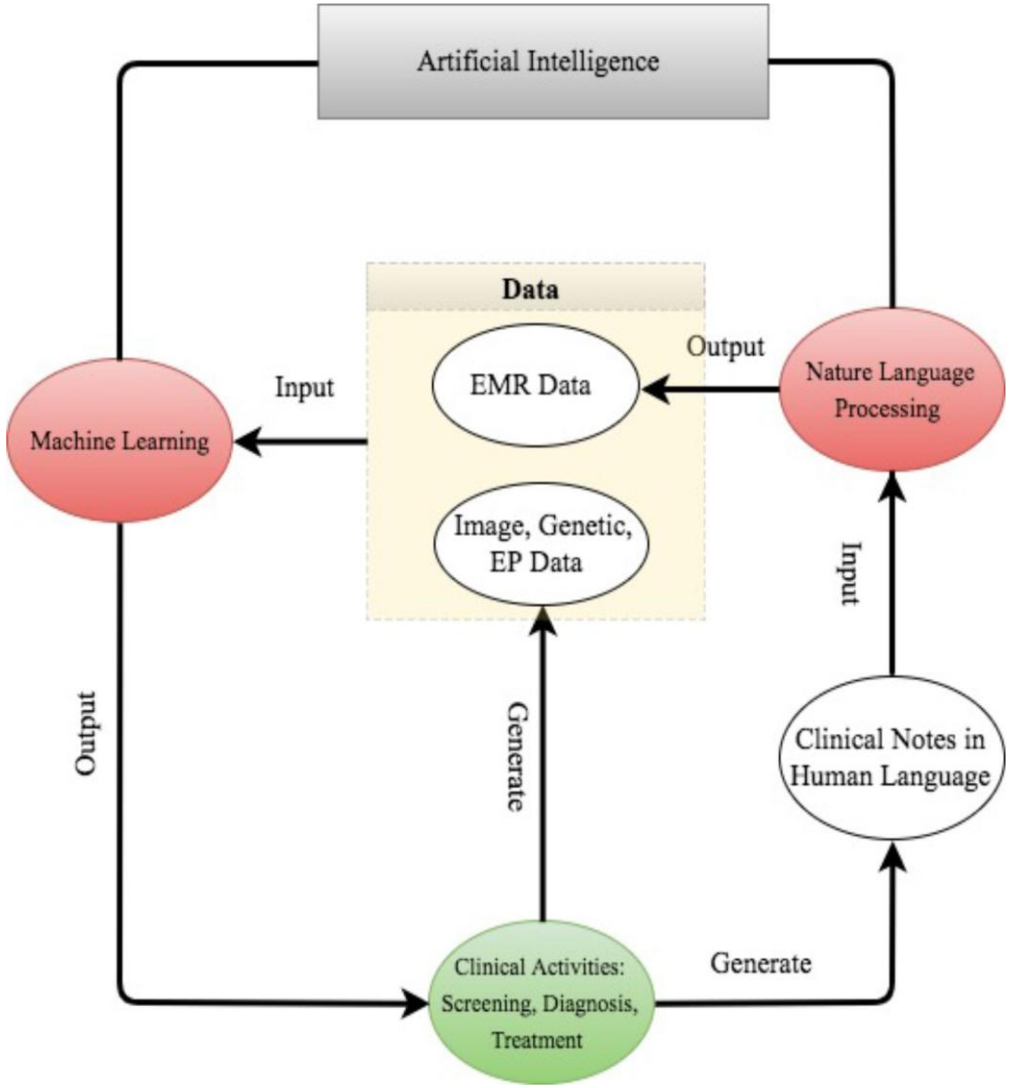
- *Up-to-date medical information* from journals textbooks and clinical practices to inform proper patient care
- *Reduce diagnostic and therapeutic errors* that are inevitable in the human clinical practice

*Extracts useful information* from a large patient population to assist making real-time inferences for health risk alert

# AI applications in healthcare

A substantial proportion of the AI literature analyses data from diagnosis imaging, genetic testing and electrodiagnosis




Artificial Intelligence
Data
EMR Data
Image, Genetic, EP Data
Machine Learning
Input
Output
Nature Language Processing
Input
Output
Generate
Clinical Notes in Human Language
Clinical Activities: Screening, Diagnosis, Treatment
Generate

# Classical ML

ML constructs data analytical algorithms to extract features from data

Inputs

- Patient '*traits*' and sometimes medical outcomes of interest
  - A patient's **traits** commonly include baseline data, such as age, gender, disease history
- Disease-specific data
  - Diagnostic imaging, gene expressions, electrophysiology test, physical examination results

Algorithms Categories

- Unsupervised learning
- Supervised learning

# Deep learning: a new era of ML

Rapid development of modern computing enables deep learning to build up neural networks with a large number of layers

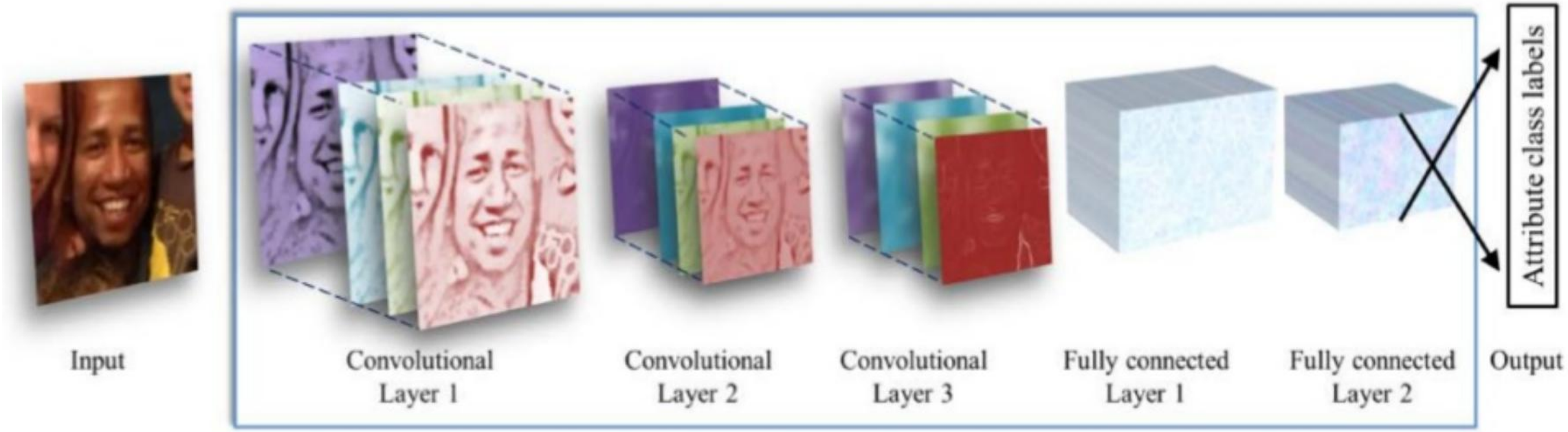
Deep learning can explore more complex *non-linear patterns* in the data

Different from the classical neural network, deep learning uses *more hidden layers* so that the algorithms can handle *complex data with various structures*

In the medical applications, the commonly used deep learning algorithms include

- Convolution neural network (CNN)
- Recurrent neural network (RNN)
- Deep belief network (DBN)
- Deep neural network (DNN)

# CNN Example




All 3 RGB channels First, resized to 256 x 256, then cropped to 227 x 227

96 filters size 3x7x7

256 filters size 96x5x5

384 filters size 256x3x3

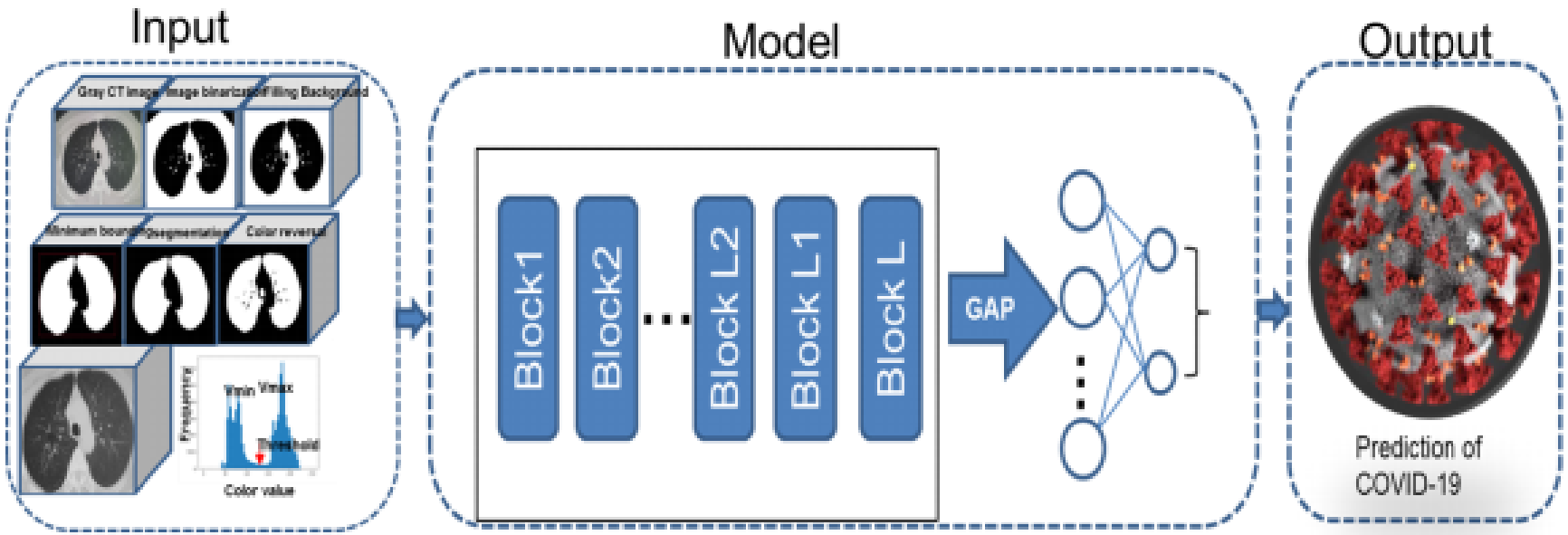
Both fully connected layers contain 512 neurons followed by ReLU and dropout layer

Output to class label (age / gender)

Each convolutional layer is followed by rectified linear operator (ReLU), max pooling layer of 3x3 regions with 2-pixel strides and a local

# Deep Learning and COVID-19

## ROI images extraction and DL algorithm framework



# DISCUSSION