Machine Learning Introduction

Lecture-1

Topics Covered

- **<u>Introduction to Machine Learning; Decision Trees</u>:**
- Overview of Supervised (regression and classification), unsupervised (clustering and dimensionality reduction), semi-supervised, and reinforcement learning with practical examples - Machine learning nomenclature: raw data, types of features and outputs, feature vector.
- Decision tree model of learning - Classification and regression using decision trees - Splitting criteria: entropy, information gain, Gini impurity – Overfitting & Pruning in decision trees.

Introduction

- Machine Learning (ML) is considered as the most dynamic and progressive form of human-like Artificial Intelligence.

- Today ML is being used extensively in various industries like automobiles, genetics, medicine, finance etc. to automate procedures, in reducing the processing time and to remove the possibility of human errors.

- ML helps in analyzing at a large scale, thus helping in making quicker and better decisions.

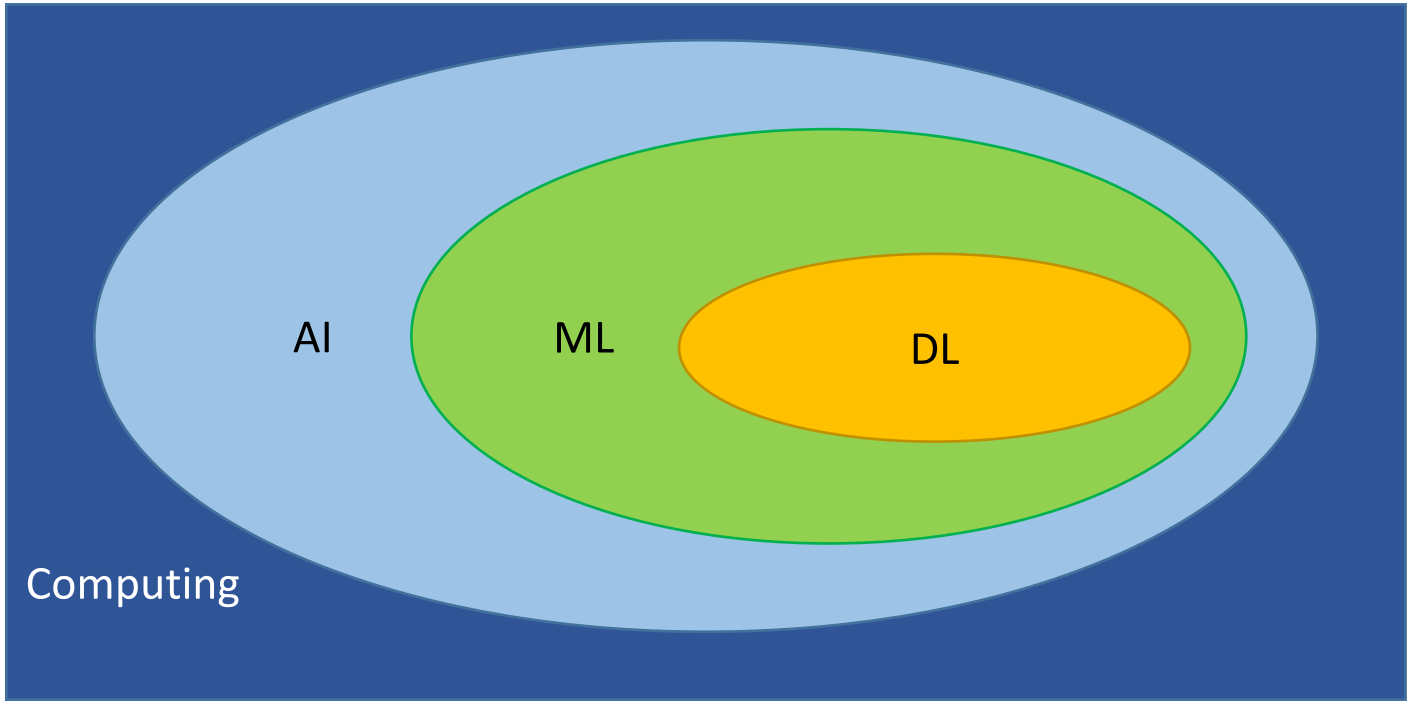
Artificial Intelligence (AI) & Machine Learning (ML)?

Artificial intelligence:

- Artificial intelligence is the name given to the process in which the computer makes decisions, **mimicking a human**.

Machine learning:

- Computer makes decisions based on experience.

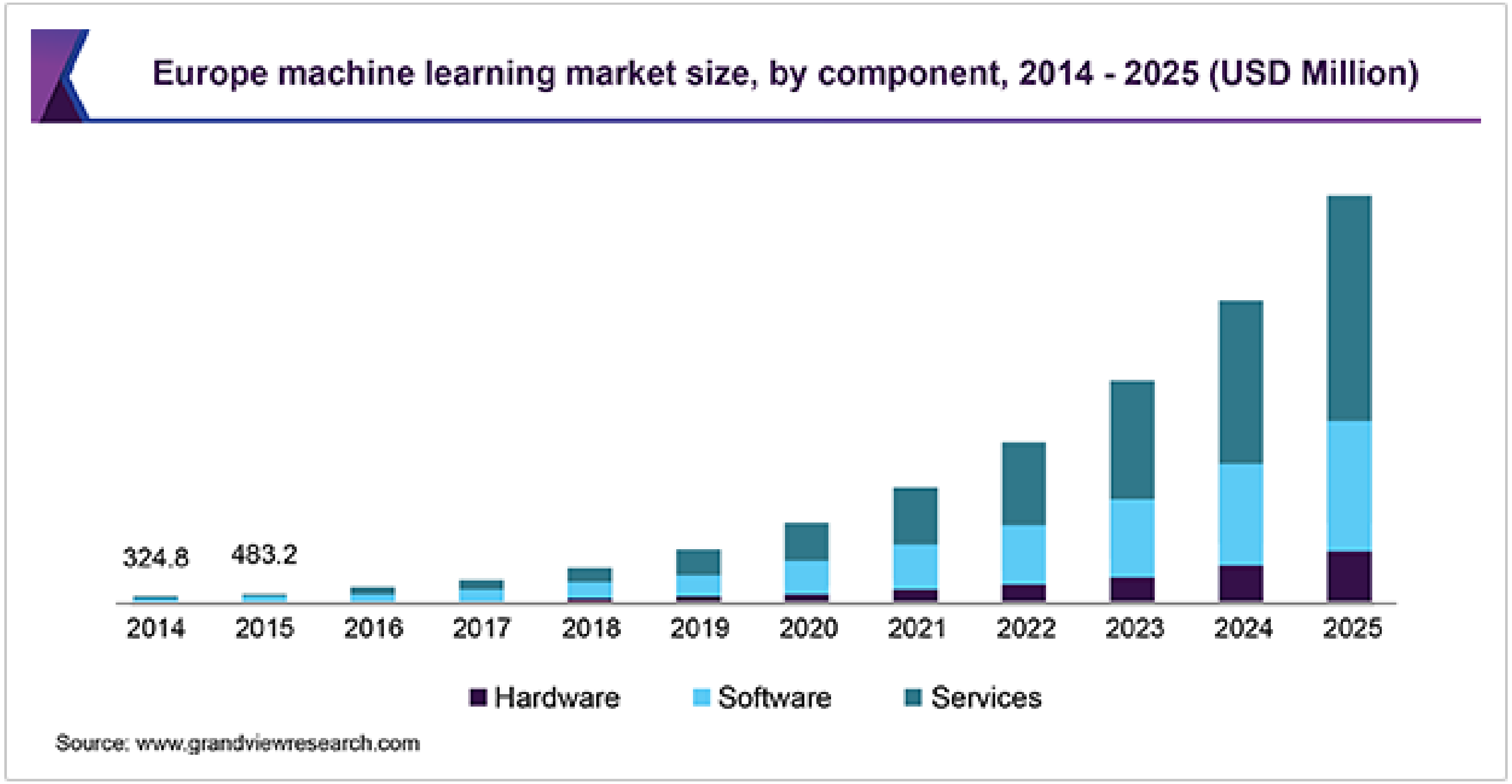


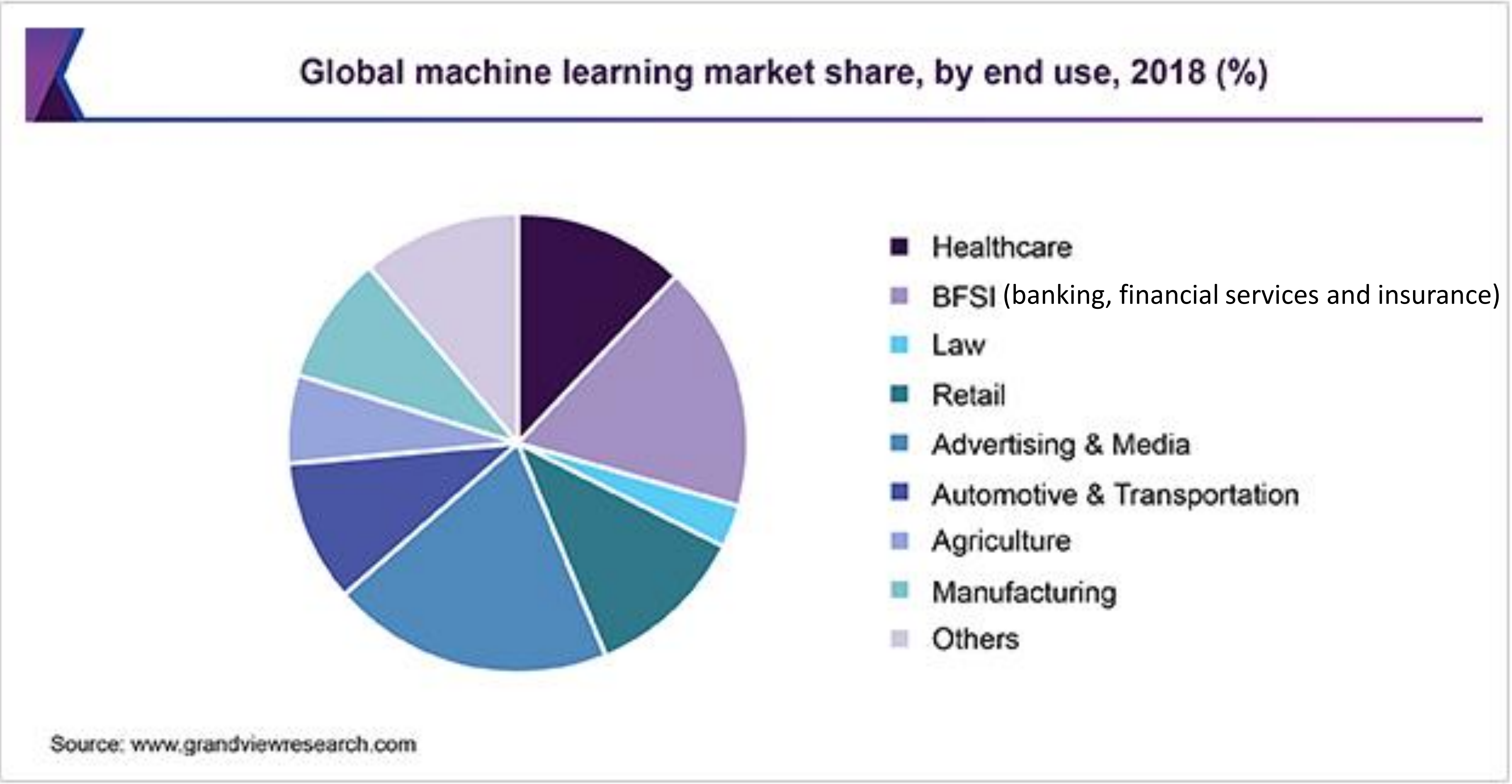
Growing popularity

- The word “Big Data” you keep hearing about is mainly made possible through ML.

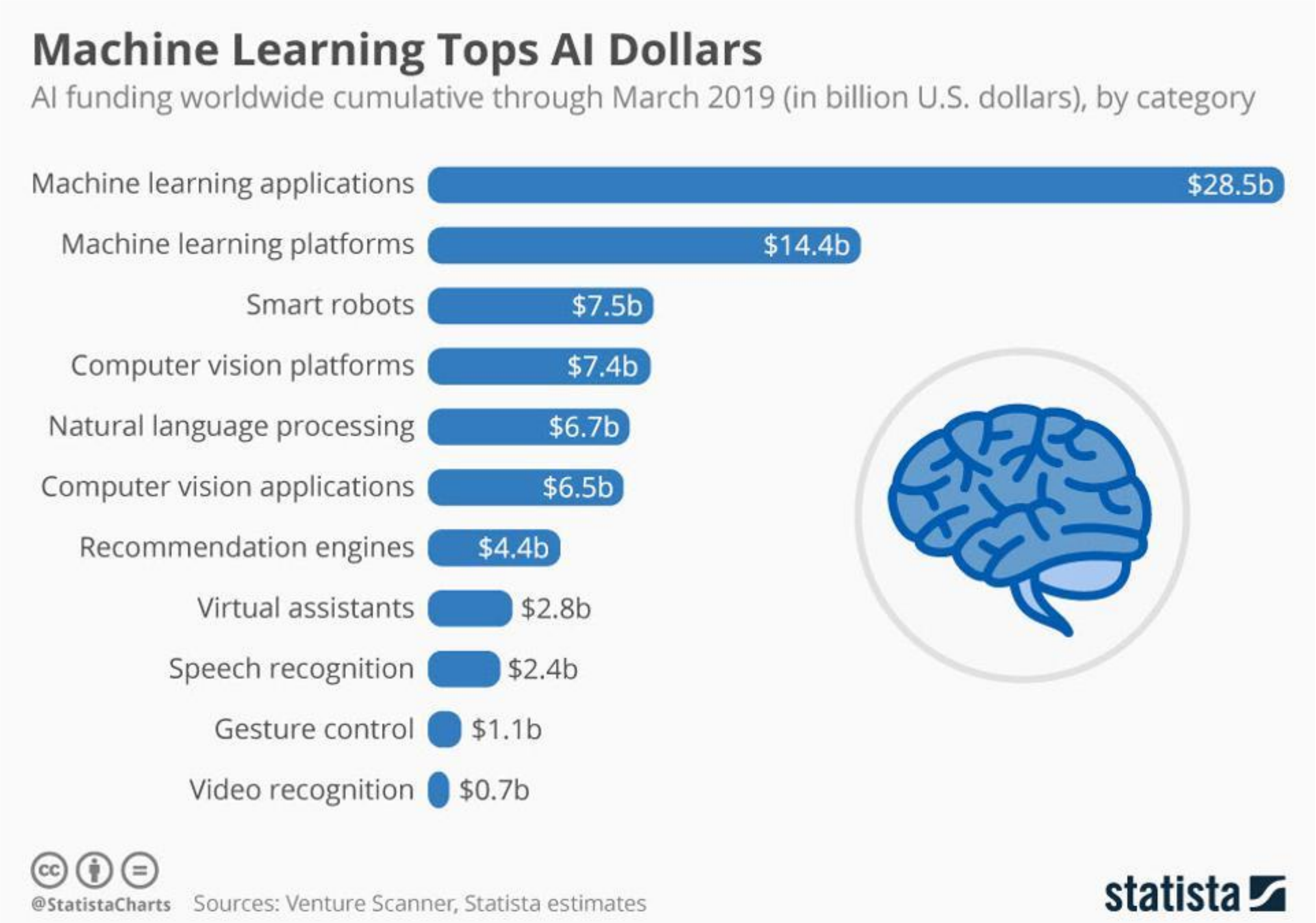
Growing popularity

Machine Learning Market Size, Share & Trends Analysis Report

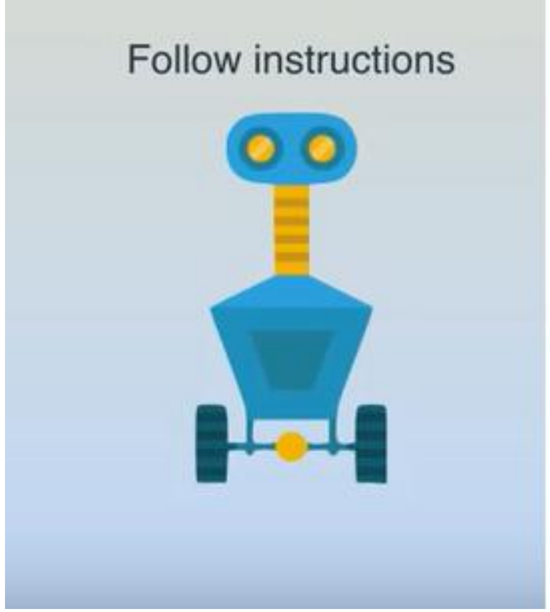




https://www.grandviewresearch.com/industry-analysis/machine-learning-market

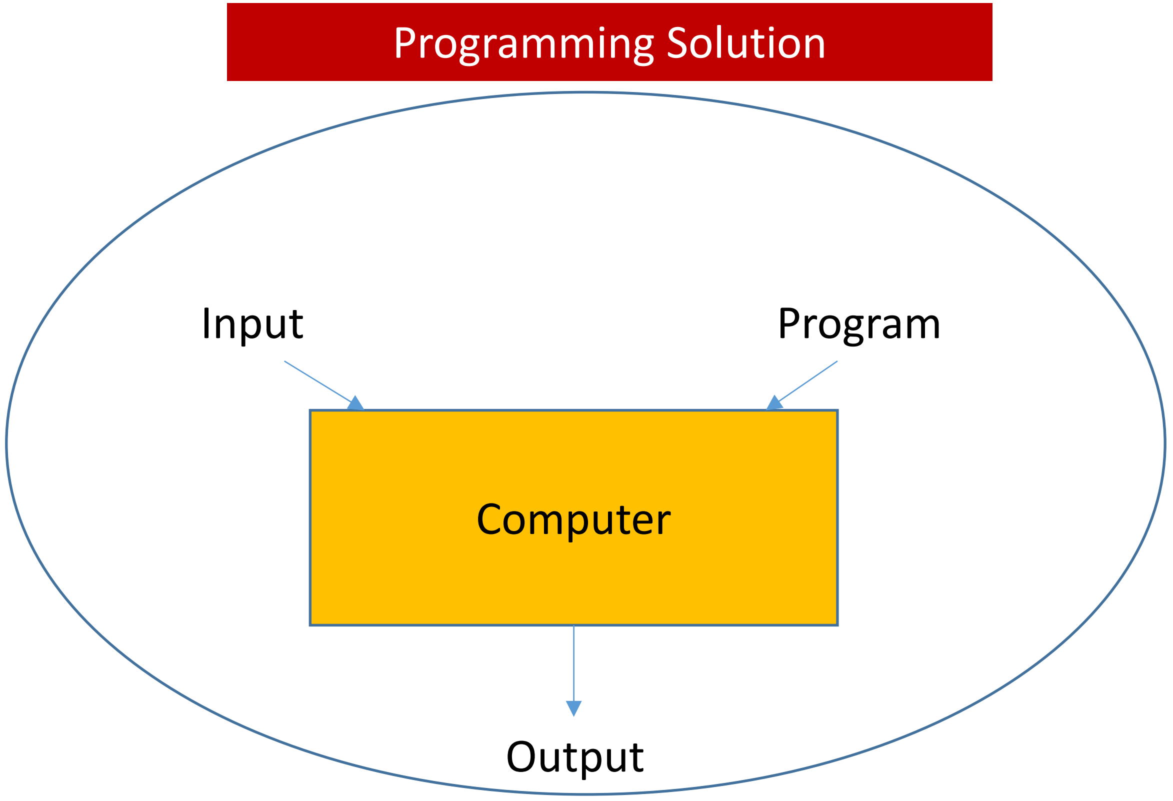

Machine Learning Tops AI Dollars
AI funding worldwide cumulative through March 2019 (in billion U.S. dollars), by category
Machine learning applications $28.5b
Machine learning platforms $14.4b
Smart robots $7.5b
Computer vision platforms $7.4b
Natural language processing $6.7b
Computer vision applications $6.5b
Recommendation engines $4.4b
Virtual assistants $2.8b
Speech recognition $2.4b
Gesture control $1.1b
Video recognition $0.7b
@StatistaCharts
Sources: Venture Scanner, Statista estimates
statista

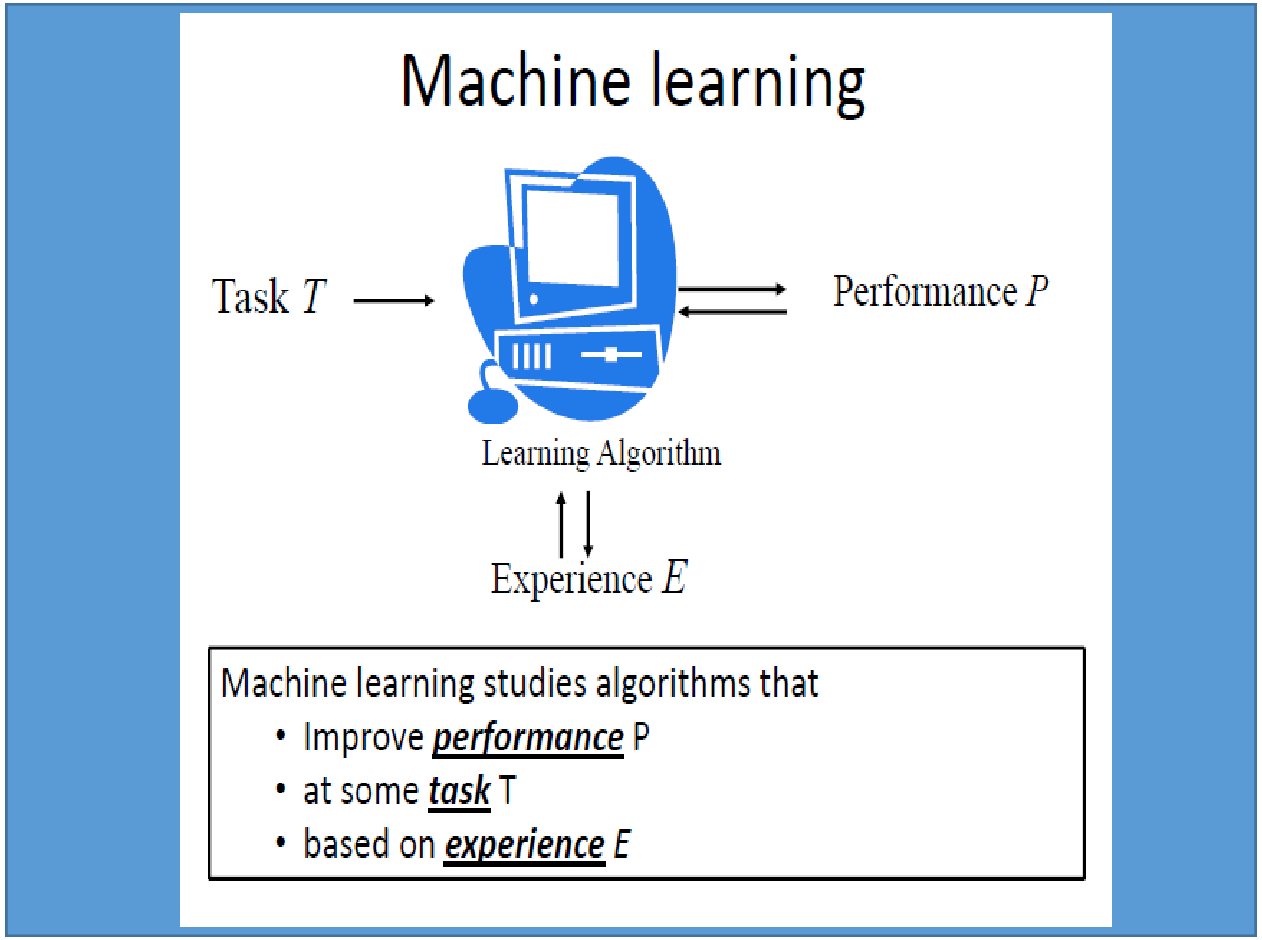
Machine Learning



We are going from programming computers to training computers.

Programming and Machine Learning





Machine Learning - Definition

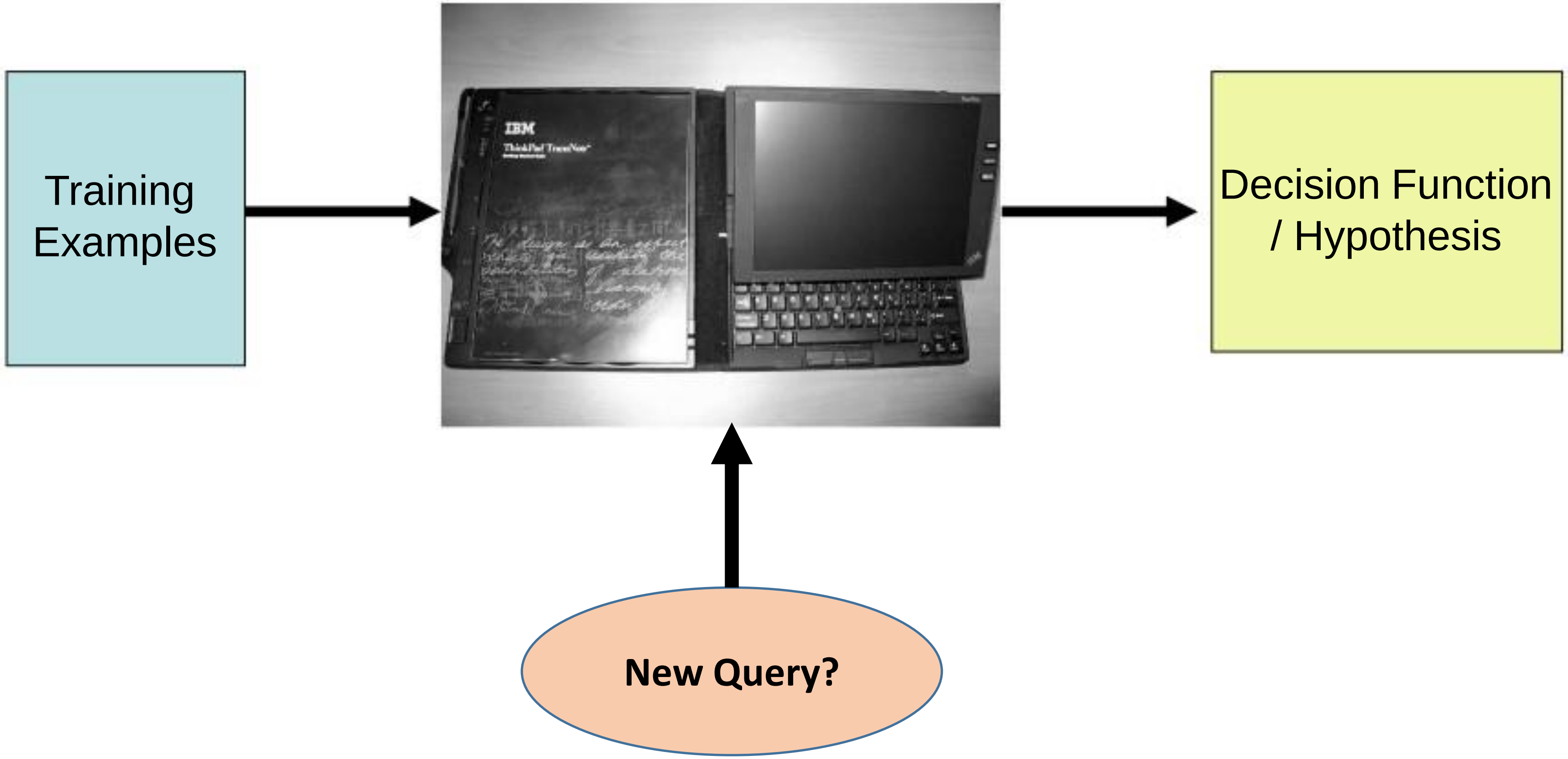
- 1997, Tom Mitchell gave a definition:

Machine Learning Scientist,

Carnegie Mellon University

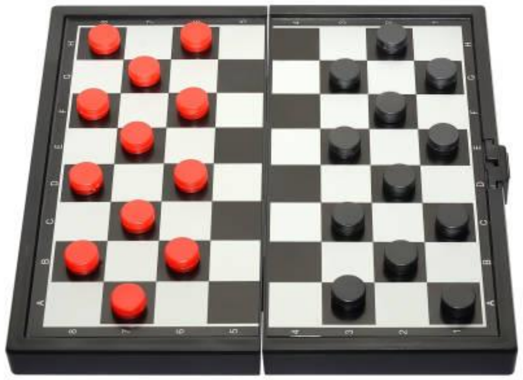
A computer program is said **to learn from experience E** with respect to some class of **tasks *T*** and **performance measure P**, if its performance at tasks in T, as measured by P, improves with experience E.

What Machine Learning does???

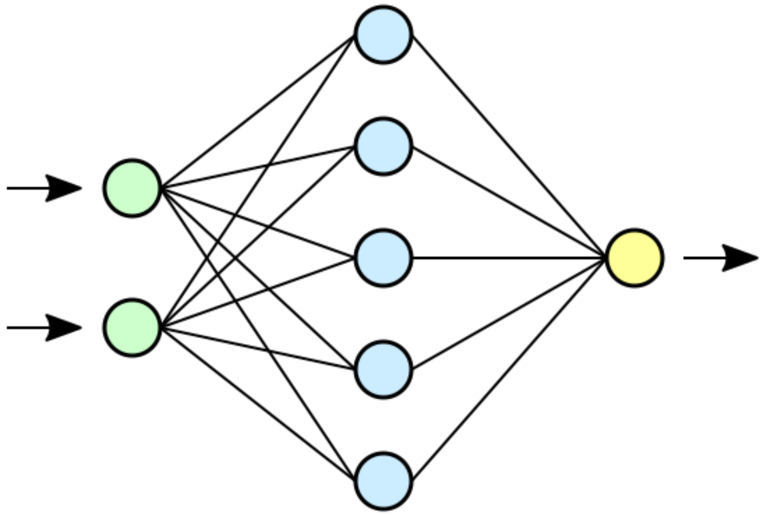


Introduction – History

- 1950s
  - Arthur Samuel (IBM)
  - Program – playing Checkers game



- 1960s
  - Rosenblatt
  - Perceptron - Neural Network Model
  - Pattern Recognition
  - Later *delta learning* rule
    - Rule for perceptron learning
    - Good classifier



Introduction – History

- 1950s
  - Arthur Samuel (IBM)
  - Program – playing Checkers game

- 1960s
  - Rosenblatt
  - Perceptron - Neural Network Model
  - Pattern Recognition
  - Later *delta learning* rule
    - Rule for perceptron learning
    - Good classifier

- 1969
  - Minsky and Papert
  - Limitation of perceptron model
  - Problem could not be represented
  - Inseparable data distribution

- 1970s
  - Symbolic concept (AI)

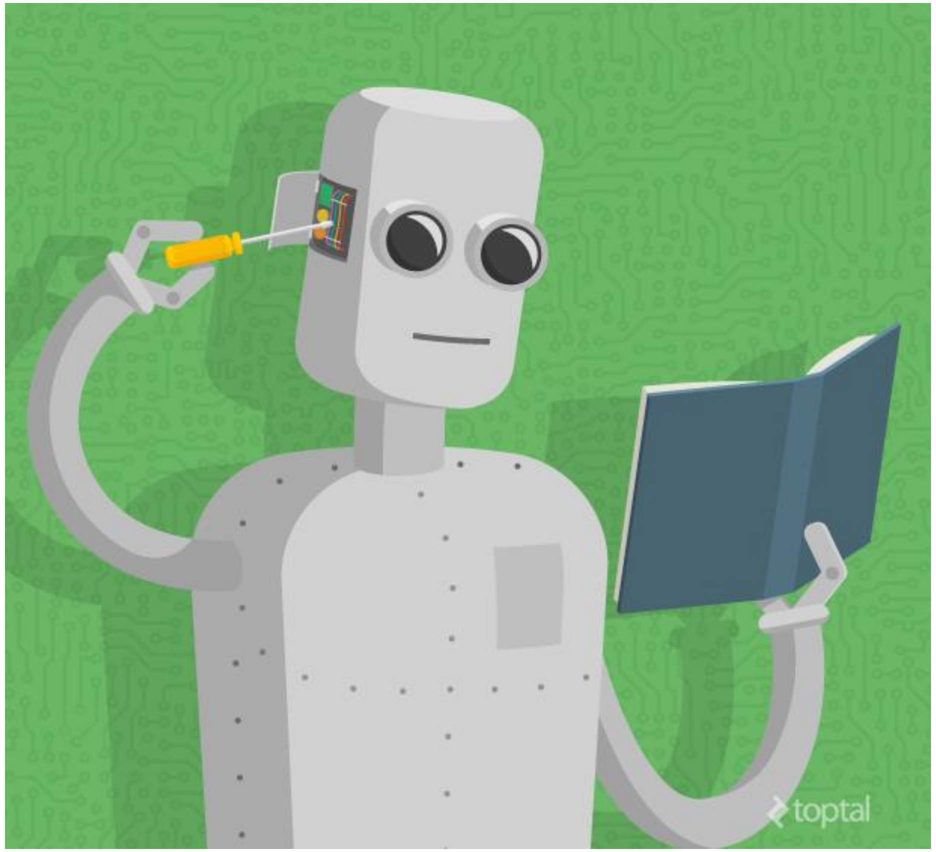
- 1986
  - Quinlan – Decision Tree
  - ID3 Algorithm
  - Improved: Regression
  - Still popular in ML

Introduction – History

- 1990s: Machine learning involved statistics to a large extent
- 1994 – Self driving car road test
- In 1995 –Support Vector Machines (SVMs)
- In 1997, ensembles or boosting - algorithm for classification
- 1997 – Deep Blue beats Garry Kasparov
- 2009 – Google builds self driving car
- 2011 – Watson wins Jeopardy
- 2015 - machine translation systems driven by NN - better than statistical machine translation systems

Current status

- Today,
  - Algorithms – developed for learning tasks
  - Theoretical understanding - emerged
  - Practical computer programs - developed
  - Commercial applications - appear.



Data and Features

What is data?

- Data is simply a table with information
- Each row is a data point
- Each row represented by certain features

What are features?

- Features are simply the columns of the table.
- Features may be size, name, type, weight, etc.
- Some features are special, and we call them *labels*.

NO.	SIZE	COLOR	SHAPE	FRUIT NAME
1	Big	Red	Rounded shape with a depression at the top	Apple
2	Small	Red	Heart-shaped to nearly globular	Cherry
3	Big	Green	Long curving cylinder	Banana
4	Small	Green	Round to oval, Bunch shape Cylindrical	Grape

What is the difference between labelled and unlabelled data?

Labels?

- If we are trying to predict a feature based on the others, that feature is the label.
- **Labeled data:** Data that comes with a label.
- **Unlabeled data:** Data that comes without a label

- The set of algorithms in which we use a **labeled dataset** is called *supervised learning*.

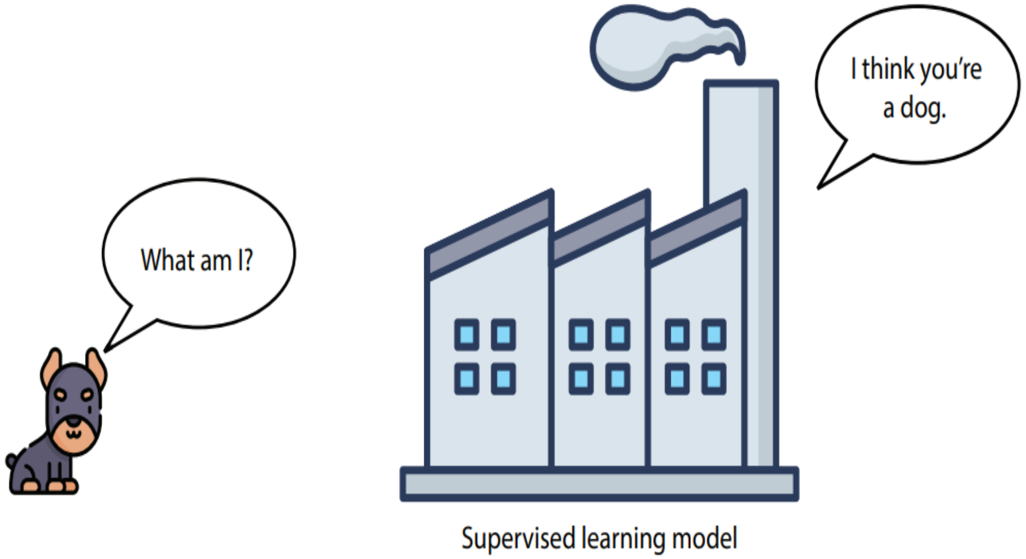
- The set of algorithms in which we use an **unlabeled dataset**, is called *unsupervised learning*.

Types of Machine Learning

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning
- Deep Learning

Supervised Learning

- A branch of machine learning that works with **labeled data**.
- Some of the most common **applications**:
  - image recognition
  - various forms of text processing
  - recommendation systems.
- Goal of a supervised learning model: **predict the labels**.



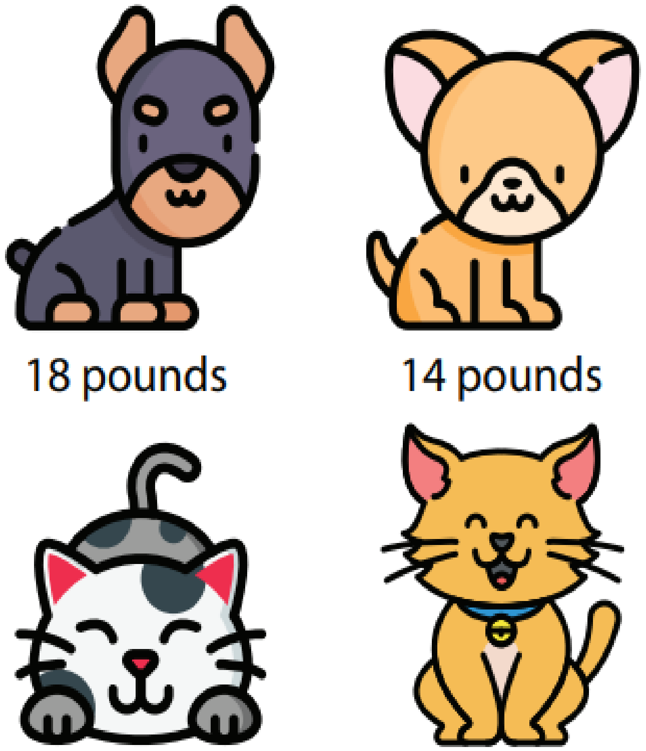
Types of labeled datasets

- Numbers and states are the two types of data used in supervised learning models.

- In this dataset, the labels are numbers.
- We call this type - numerical data

- numerical data - is any type of data that uses numbers such as 4, 2.35, or –199.
- Example: prices, sizes, or weights.

Labeled data



18 pounds
14 pounds
12 pounds
9 pounds

In this example, each data point in the dataset is labeled with the weight of the animal.

Types of labeled datasets

- In this dataset, the labels are states.
- We call this type - categorical data.
- categorical data - is any type of data that uses categories, or states, such as male/female or cat/dog/bird.
- For this type of data, we have a finite set of categories to associate to each of the data points.

each data point in the dataset is labeled with the type of animal (dog or cat)

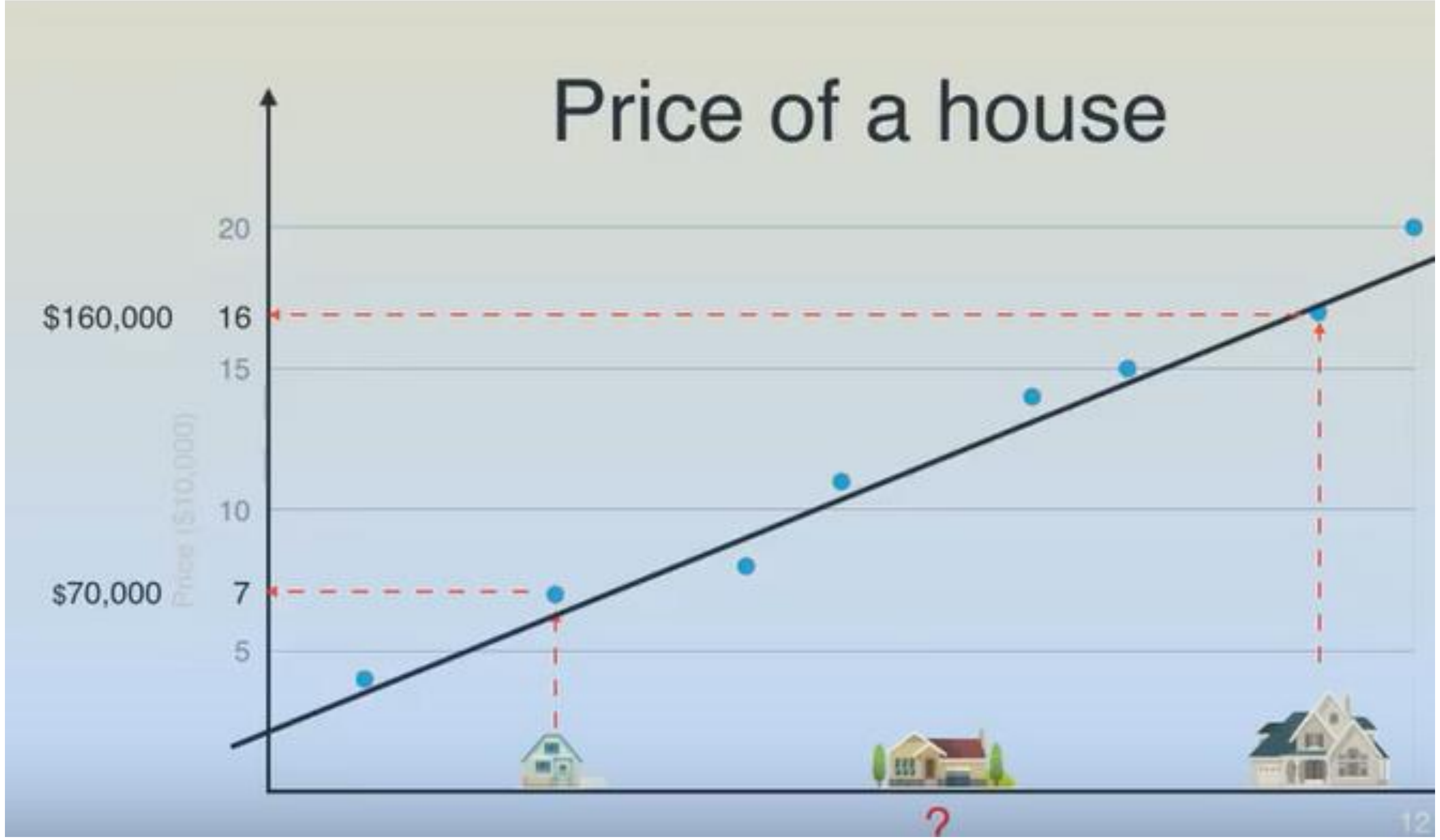
Types of supervised learning models

- **regression models** are the types of models that **predict numerical data**.
- The **output** of a regression model is a **number**, such as the **weight of the animal**.

- **classification models** are the types of models that predict **categorical data**.
- The **output** of a classification model is a **category**, or a state, such as the type of animal (**cat or dog**).

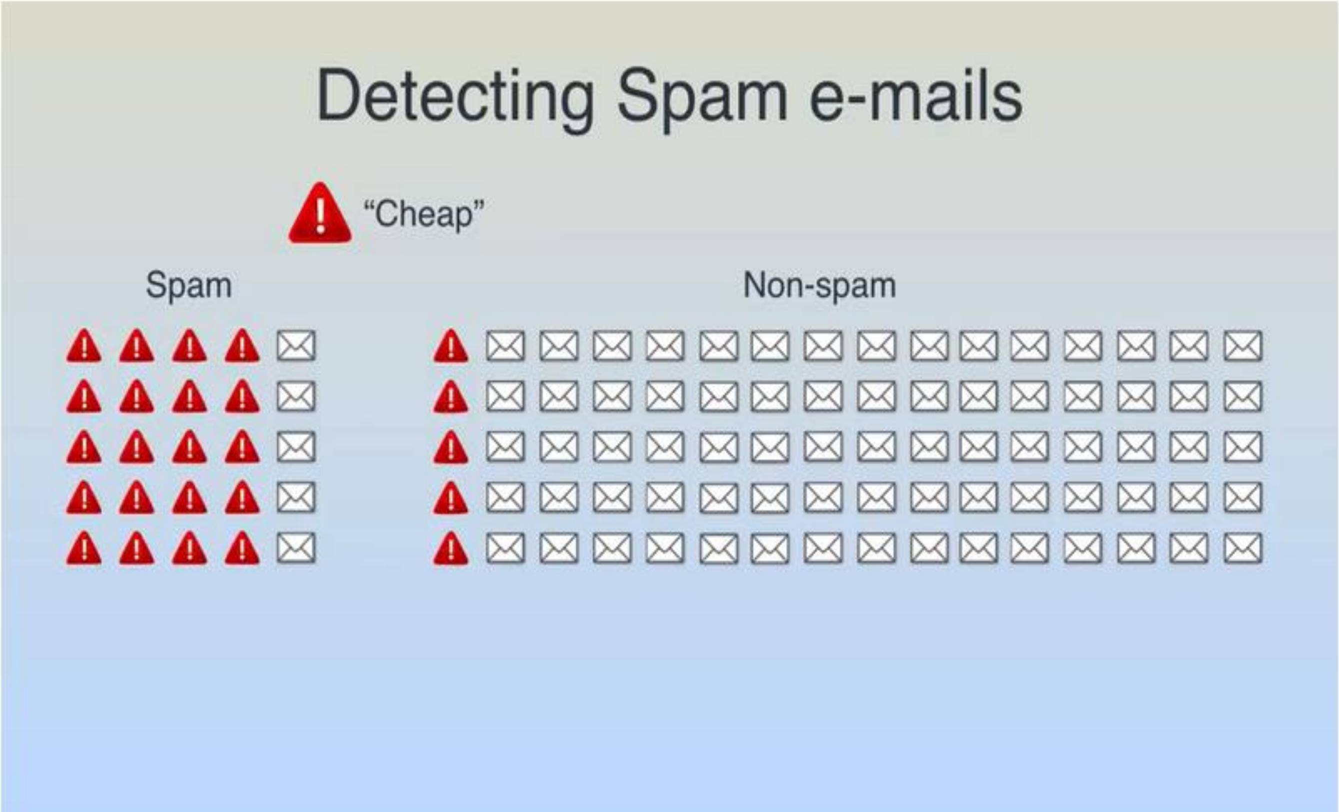
Example of regression model

- **Model 1**: housing prices model.
- Each **data point** is a **house**.
- The **label** of each house is its **price**.
- Goal: when a new house (data point) comes on the market, we would like to predict its label (price).



Example of classification model

- **Model 2**: email spam–detection model.
- Each **data point** is an **email**.
- The label of each email is either **spam or ham**.
- Goal: when a **new** email (data point) comes into our inbox, we would like to predict its label (whether it is **spam** or **ham)**.



Examples of supervised learning models

- Difference between models 1 and 2.

- **Housing prices model**, can return a number from many possibilities, such as \$100, \$250,000, or \$3,125,672.33.
- Thus, it is a regression model.

- The **spam detection model**, can return only two things: spam or ham. Thus, it is a classification model.

Regression models

- Predict numbers based on the features
- In the housing example, the features can be anything that describes a house, such as the **size, the number of rooms**, the **distance to the closest school**, or the **crime rate in the neighborhood**.

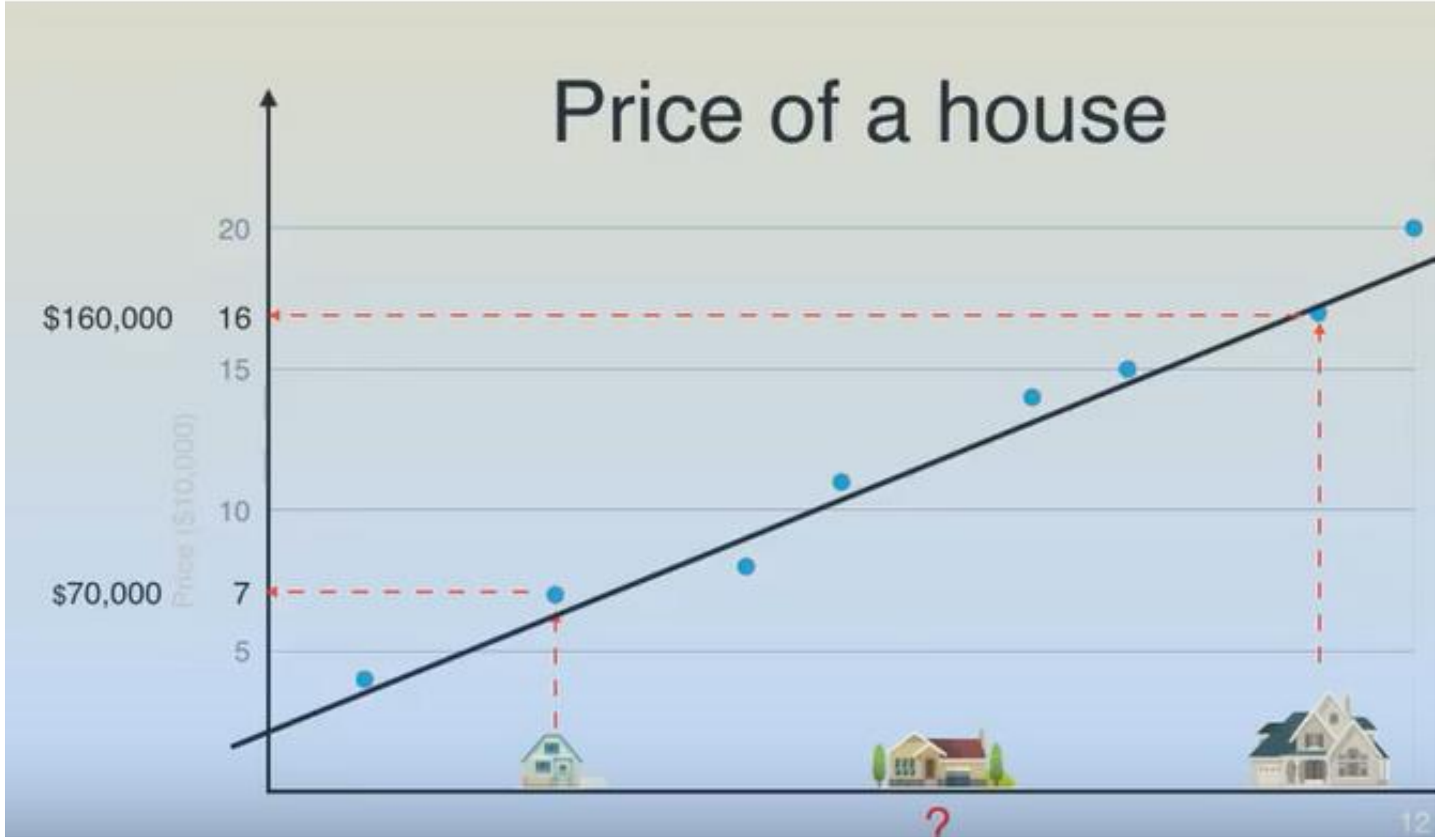
House size	No. of rooms	Distance to school	Crime rate in the neighborhood	Price
IDV	IDV	IDV	IDV	DV

Other applications of regression model:

- **Stock market**: predicting the price of a certain stock based on other stock prices and other market signals
- **Medicine**: predicting the expected life span of a patient or the expected recovery time, based on symptoms and the medical history of the patient
- **Sales**: predicting the expected amount of money a customer will spend, based on the client's demographics and past purchase behavior
- **Video recommendations**: predicting the expected amount of time a user will watch a video, based on the user's demographics and other videos they have watched

Linear regression

- The most common method used for regression is **linear regression**, which uses linear functions (lines or similar objects) to make our predictions based on the features.



Supervised Learning

Learning the physical characters of fruits through **training**.

Apple:

- Size: Big
- Color: Red
- Shape: Rounded shape with a depression at the top

<Apple> <Big, Red, Rounded shape with a depression at the top>

Supervised Learning

Learning the physical characters of fruits through **training**.

Cherry:

- Size: Small
- Color: Red
- Shape: Heart-shaped to nearly globular

<Cherry> <Small, Red, Heart-shaped to nearly globular>

Supervised Learning

Learning the physical characters of fruits through **training**.

Banana:

- Size: Big
- Color: Green
- Shape: Long curving cylinder

Supervised Learning

Learning the physical characters of fruits through **training**.

Grape:

- Size: Small
- Color: Green
- Shape: Round to oval, Bunch shape Cylindrical

Supervised Learning

Machine already **learned** about the fruits through **training**.

Input: <Big, Red, Rounded shape>

Response: <Apple>

Input: <Small, Red, Heart-shaped>

Response: <Cherry>

Input: <Big, Green, Long curving cylinder>

Response: <Banana>

Input: <Small, Green, Round to oval shape>

Response: <Grape>

Apply that knowledge to the test data

Supervised Learning

- Already **learned** about the physical characters of fruits through **training**.

NO.	SIZE	COLOR	SHAPE	FRUIT NAME
1	Big	Red	Rounded shape with a depression at the top	Apple
2	Small	Red	Heart-shaped to nearly globular	Cherry
3	Big	Green	Long curving cylinder	Banana
4	Small	Green	Round to oval, Bunch shape Cylindrical	Grape

Supervised Learning

Input attributes + output

+ Apple

+ Orange

Decision Function / Hypothesis

Supervised Classification

For a training input, output is known

Unsupervised Learning

Decision Function / Hypothesis

Unsupervised Classification

Unsupervised Learning

- Consider physical character of that particular fruit.
  - Suppose you have considered **color**.
  - Arrange them on considering base condition as **color**
  - Then the groups will be some thing like this.

RED COLOR GROUP: apples & cherry fruits

GREEN COLOR GROUP: bananas & grapes

Unsupervised Learning

RED COLOR GROUP: apples & cherry fruits

GREEN COLOR GROUP: bananas & grapes

BIG SIZE

apples

SMALL SIZE

cherry

BIG SIZE

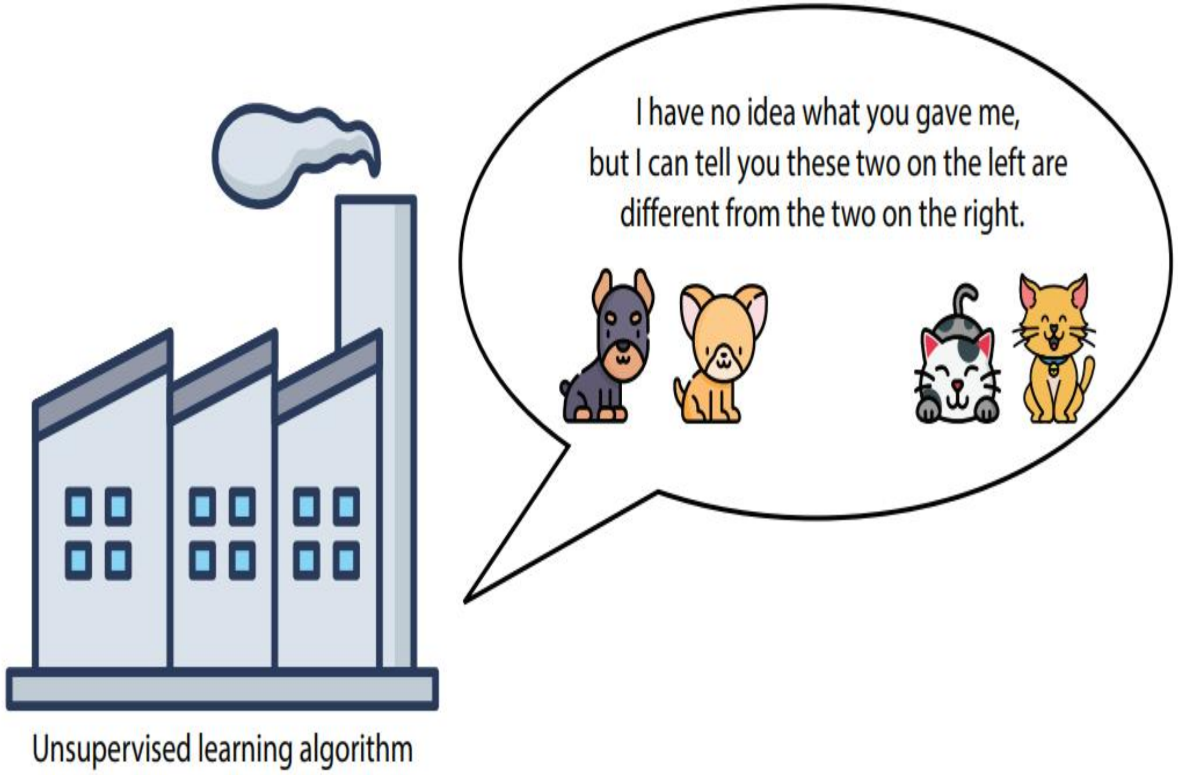
bananas

SMALL SIZE

grapes

Unsupervised learning

- Machine learning algorithms that works with unlabeled data.
- MLA must extract as much information as possible from a dataset (has no labels, or targets) to predict.
- Determine - two pictures are **similar or different**



Unsupervised learning

- Even if the labels are there, we can still use unsupervised learning techniques on our data **to preprocess** it and apply supervised learning methods more effectively.

- **clustering algorithms** The algorithms that group data into clusters based on similarity.

- **dimensionality reduction algorithms** The algorithms that simplify our data and describe it with fewer features

- **generative algorithms** The algorithms that can generate new data points that resemble the existing data

Clustering

- Consider the two datasets used in “Supervised learning”—the housing dataset and the email dataset.

- Imagine that they have no labels
  - House price prediction – **price is not available**
  - Email classification – **spam or ham is not available**.

- **housing dataset -** What can we do with this dataset?
  - Here is an idea: we could somehow group the houses by similarity.
  - For example, we could group them by **location, size, or a combination** of these factors.
  - This process is called **clustering**.

Clustering is an unsupervised machine learning - group the elements in our dataset into clusters where all the data points are similar.

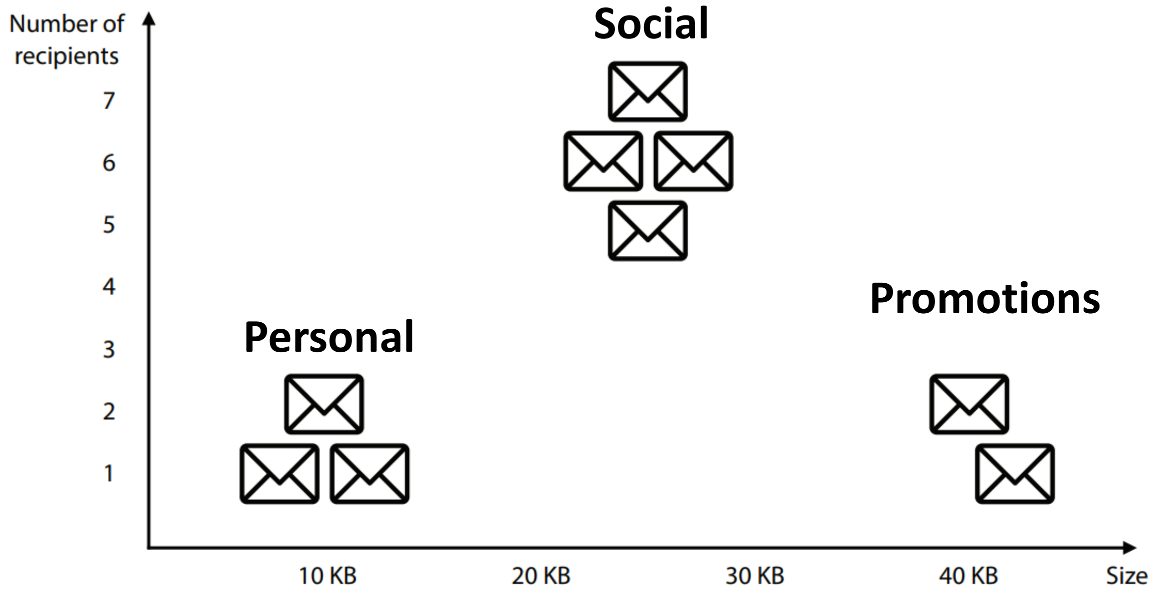
Clustering

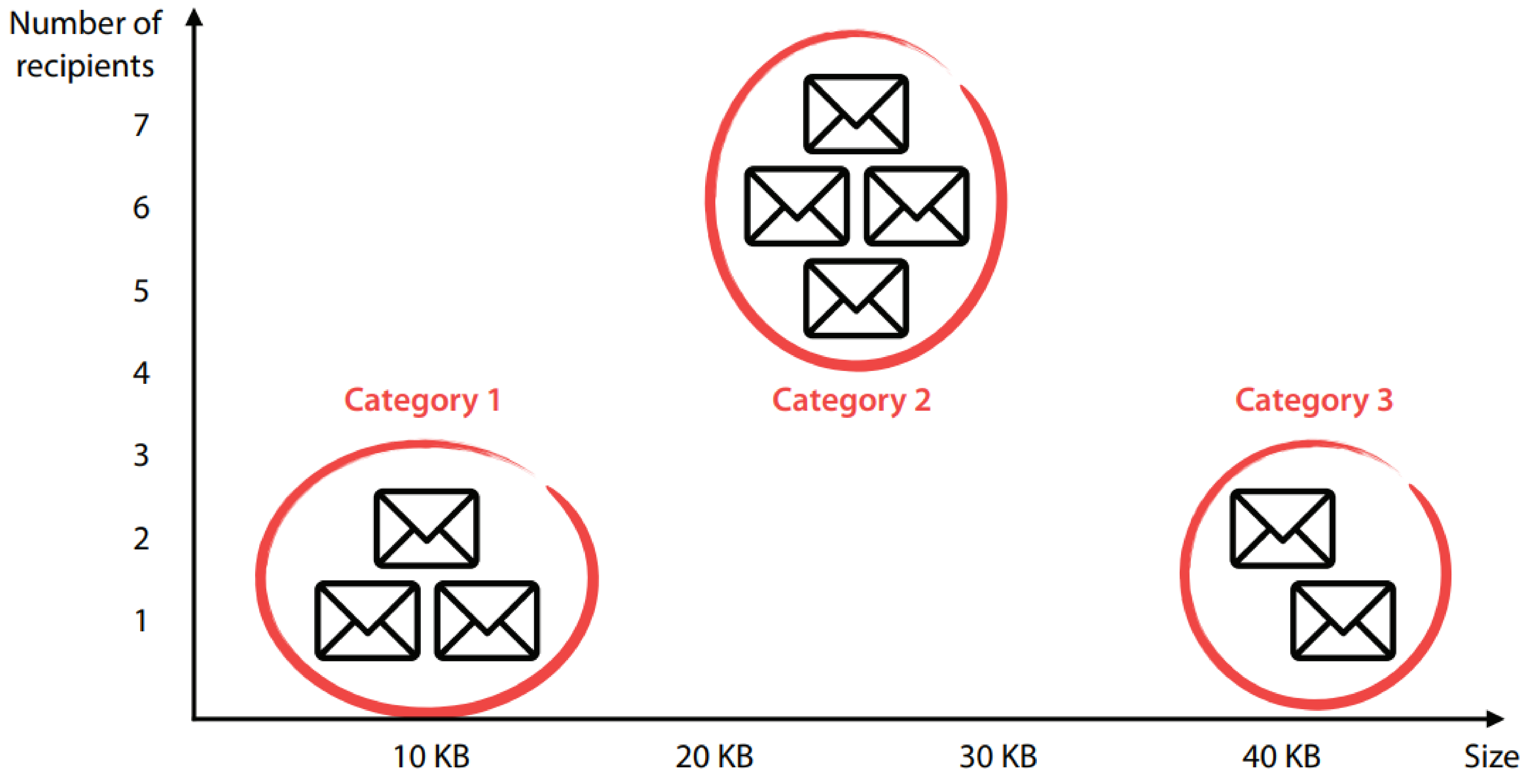
- Example – email dataset
  - The dataset is unlabeled, we don't know whether each email is spam or ham.

  - We can apply some clustering to this dataset.
  - Group the emails – based on the number of words in the message, the sender, the number and size of the attachments, or the types of links inside the email.
  - After clustering the dataset, a human (or a combination of a human and a supervised learning algorithm) could label these clusters by categories such as "**Personal**," "**Social**," and "**Promotions**."

cluster the emails into three categories based on size and number of recipients

Email	Size	Recipients
1	8	1
2	12	1
3	43	1
4	10	2
5	40	2
6	25	5
7	23	6
8	28	6
9	26	7





Number of recipients
7
6
5
4
3
2
1
Category 1
Category 2
Category 3
10 KB
20 KB
30 KB
40 KB
Size

Other applications of clustering

- **Market segmentation**: dividing customers into groups based on demographics and previous purchasing behavior to create different marketing strategies for the groups
- **Genetics**: clustering species into groups based on gene similarity
- **Medical imaging**: splitting an image into different parts to study different types of tissue
- **Video recommendations**: dividing users into groups based on demographics and previous videos watched and using this to recommend to a user the videos that other users in their group have watched

Popular clustering algorithms

- **K-means clustering**: this algorithm groups points by picking some **random centers** of mass and moving them closer and closer to the points until they are at the right spots.
- **Hierarchical clustering**: this algorithm starts by **grouping the closest points together** and continuing in this fashion, until we have some well-defined groups.
- **Density-based spatial clustering (DBSCAN):** this algorithm starts **grouping points together in places with high density**, while labeling the isolated points as noise.
- **Gaussian mixture models**: this algorithm does not assign a point to one cluster but instead assigns fractions of the point to each of the existing clusters.
- For example, if there are three clusters, A, B, and C, then the algorithm could determine that 60% of a particular point belongs to group A, 25% to group B, and 15% to group C.

Dimensionality reduction

- Simplifies data without losing too much information

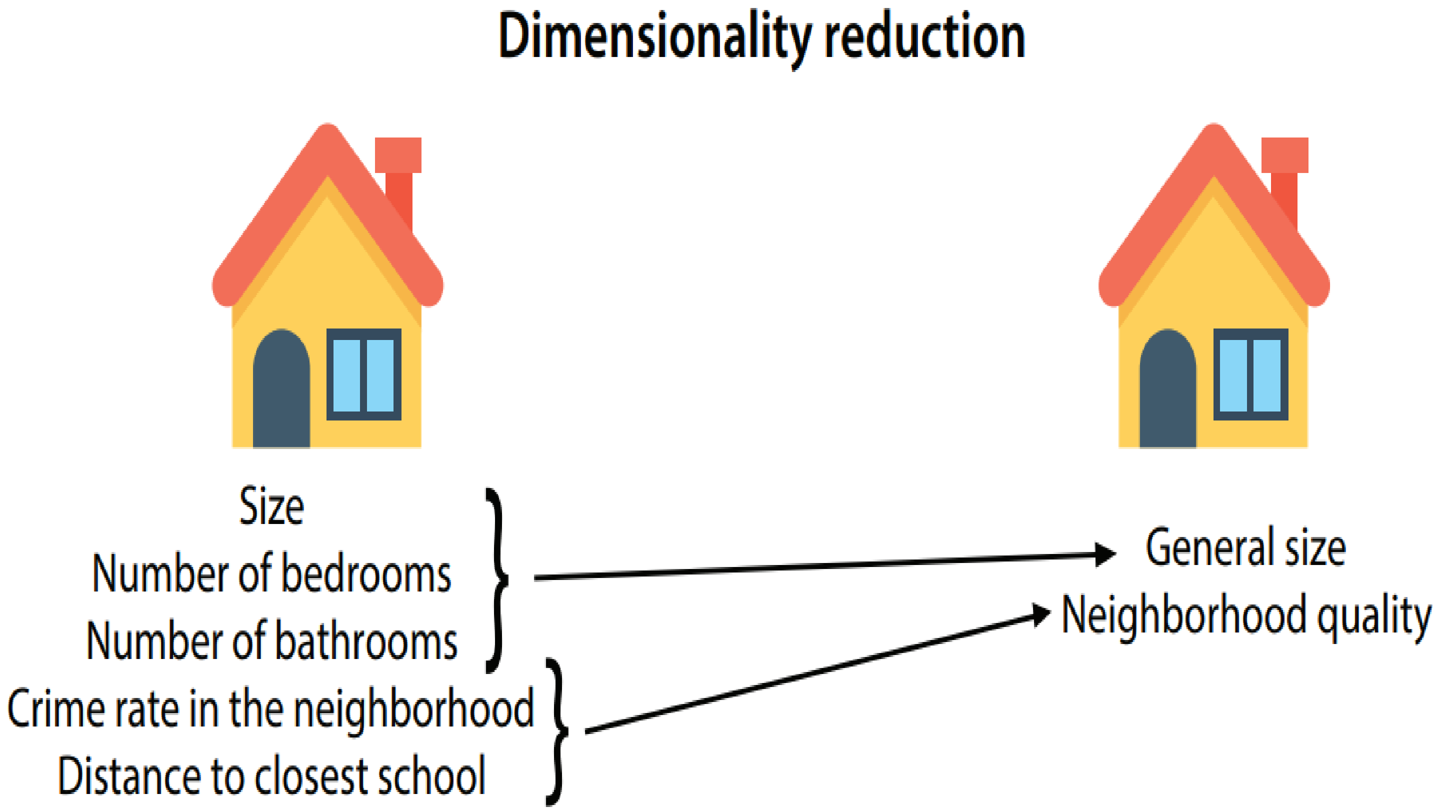
- Example: housing dataset

  Imagine the features are the following:
  - C1: Size
  - C2: Number of bedrooms
  - C3: Number of bathrooms
  - C4: Crime rate in the neighborhood
  - C5: Distance to the closest school

  This dataset has five columns of data.
  What if we wanted to turn the dataset into a simpler one with **fewer columns**, without losing a lot of information?

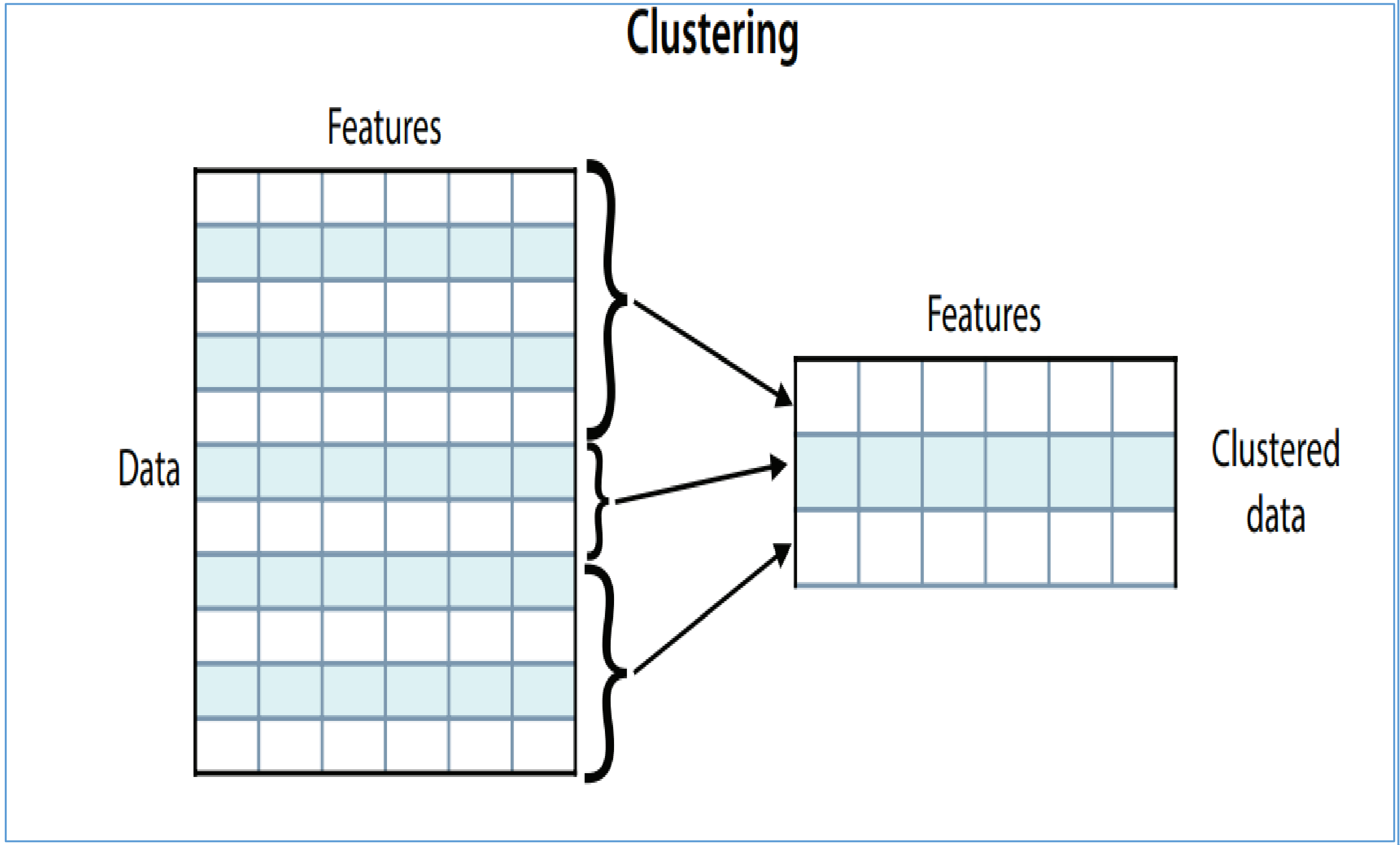
Dimensionality reduction

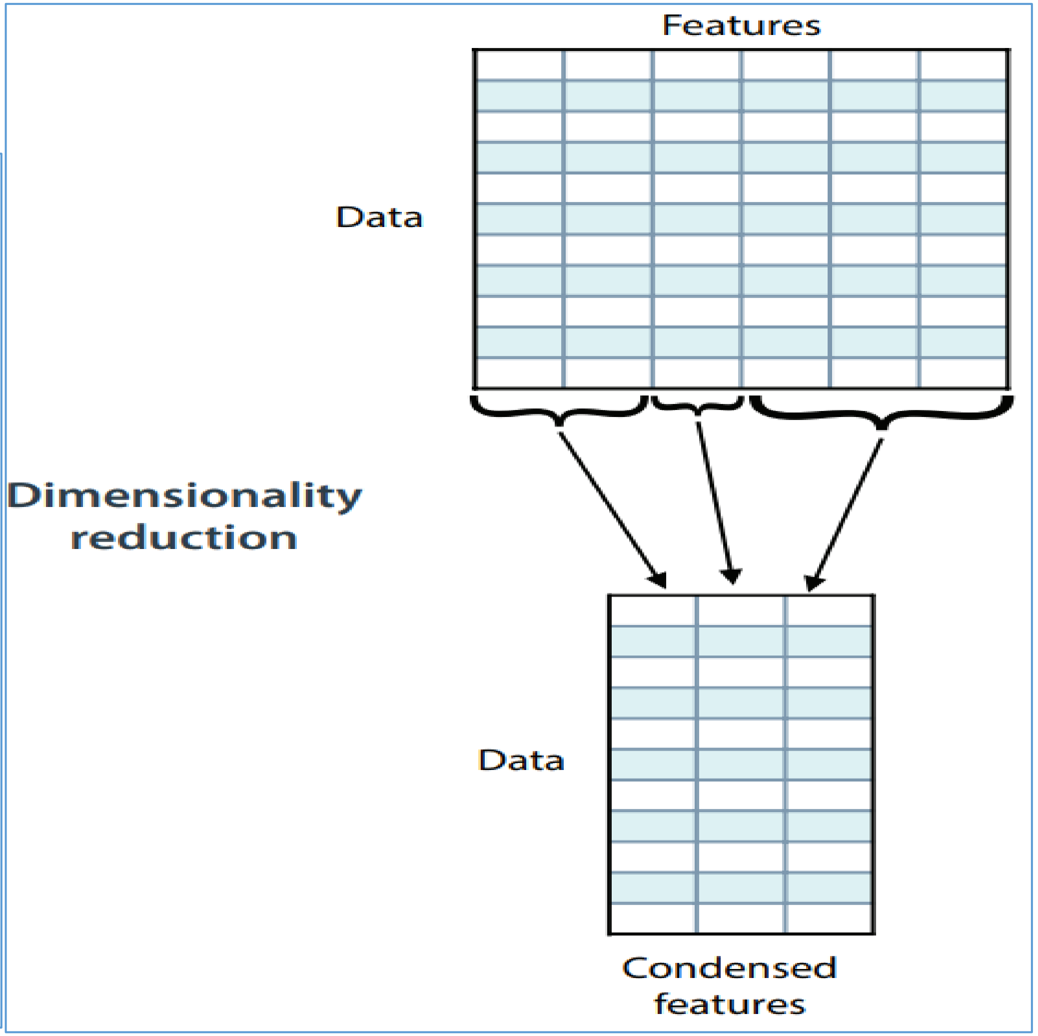
- first three features are **similar**, because they are all related to the **size** of the house.
- fourth and fifth features are similar to each other, because they are **related to the quality of the neighborhood**.



Clustering & dimensionality reduction

- If we have a table full of data, each row corresponds to a data point, and each column corresponds to a feature.
- we can use **clustering** to reduce the number of rows in our dataset and **dimensionality reduction** to reduce the number of columns





Other ways of simplifying our data: Matrix factorization and singular value decomposition

- How can we reduce both the rows and the columns at the same time?
  - **matrix factorization** and **singular value decomposition (SVD)**.
  - These two algorithms express a **big matrix of data** into a product of **smaller matrices**
  - Netflix use matrix factorization extensively to generate recommendations.
    - a large table where each row corresponds to a **user**, each column to a **movie**, and each **entry in the matrix is the rating** that the user gave the movie.
    - With matrix factorization, one can extract certain features, such as type of movie, actors appearing in the movie, and others, and be able to predict the rating that a user gives a movie, based on these features.

- Two common types of unsupervised learning algorithms are clustering and dimensionality reduction.

  - **Clustering** is used to group data into **similar clusters** to extract information or make it easier to handle.

  - **Dimensionality reduction** is a way to **simplify our data**, by joining certain similar features and losing as little information as possible.

  - **Matrix factorization** and **singular value decomposition** are other algorithms that can simplify our data by reducing both the number of rows and columns.

- **Generative machine learning** is an innovative type of unsupervised learning, consisting of **generating data** that is similar to our dataset.
  - Generative models can paint realistic faces, compose music, and write poetry.

Goal:
To solve problems that cannot be solved by numerical means alone

ML / DL

Machine Learning - Examples

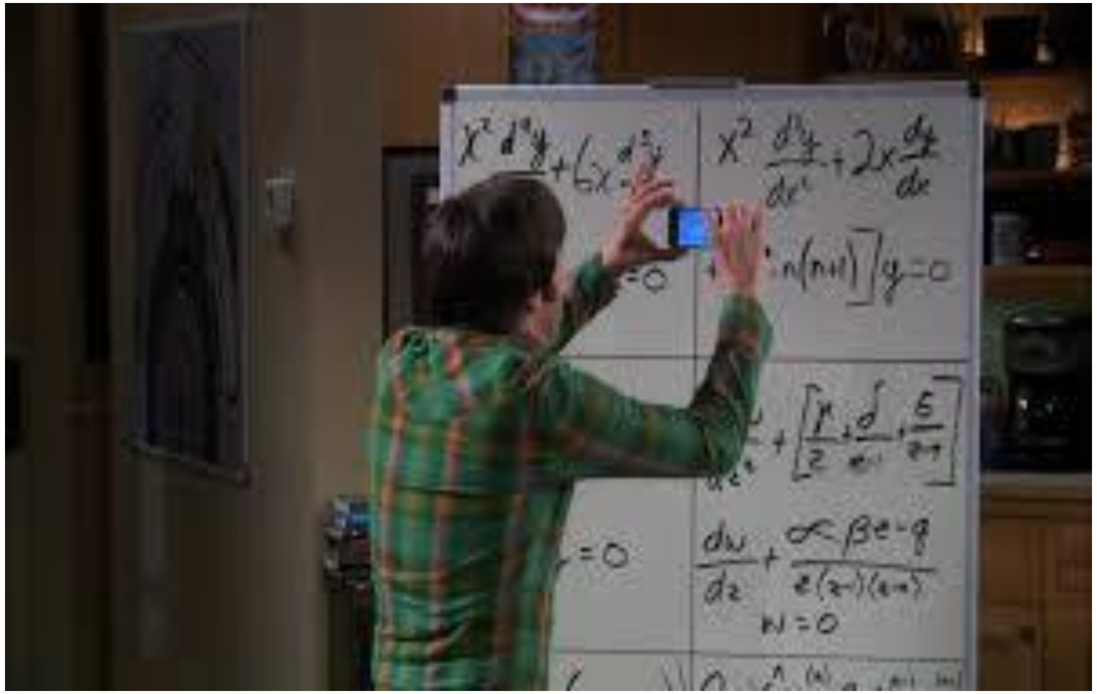
- In general, to have a well-defined learning problem, we must identity these three features:
  - The class of **tasks (T)**
  - The measure of **performance (P)**to be improved and
  - The source of **experience (E)**

A chess learning problem:

- **Task T**: playing chess
- **Performance measure P**: % of games won against opponents (70%)
- **Training experience E**: playing practice games against itself

Machine Learning - Examples

- **A handwriting recognition learning problem:**
  - **Task T**: recognizing and classifying handwritten words within images
  - **Performance measure *P:*** percent of words correctly classified
  - **Training experience E**: database of handwritten words



Machine Learning - Examples

- **A robot driving learning problem:**
  - **Task T**: driving on public four-lane highways
  - **Performance measure P**: average distance traveled before an error
  - **Training experience E**: sequence of images and steering commands recorded while observing a human driver

Applications of Machine Learning

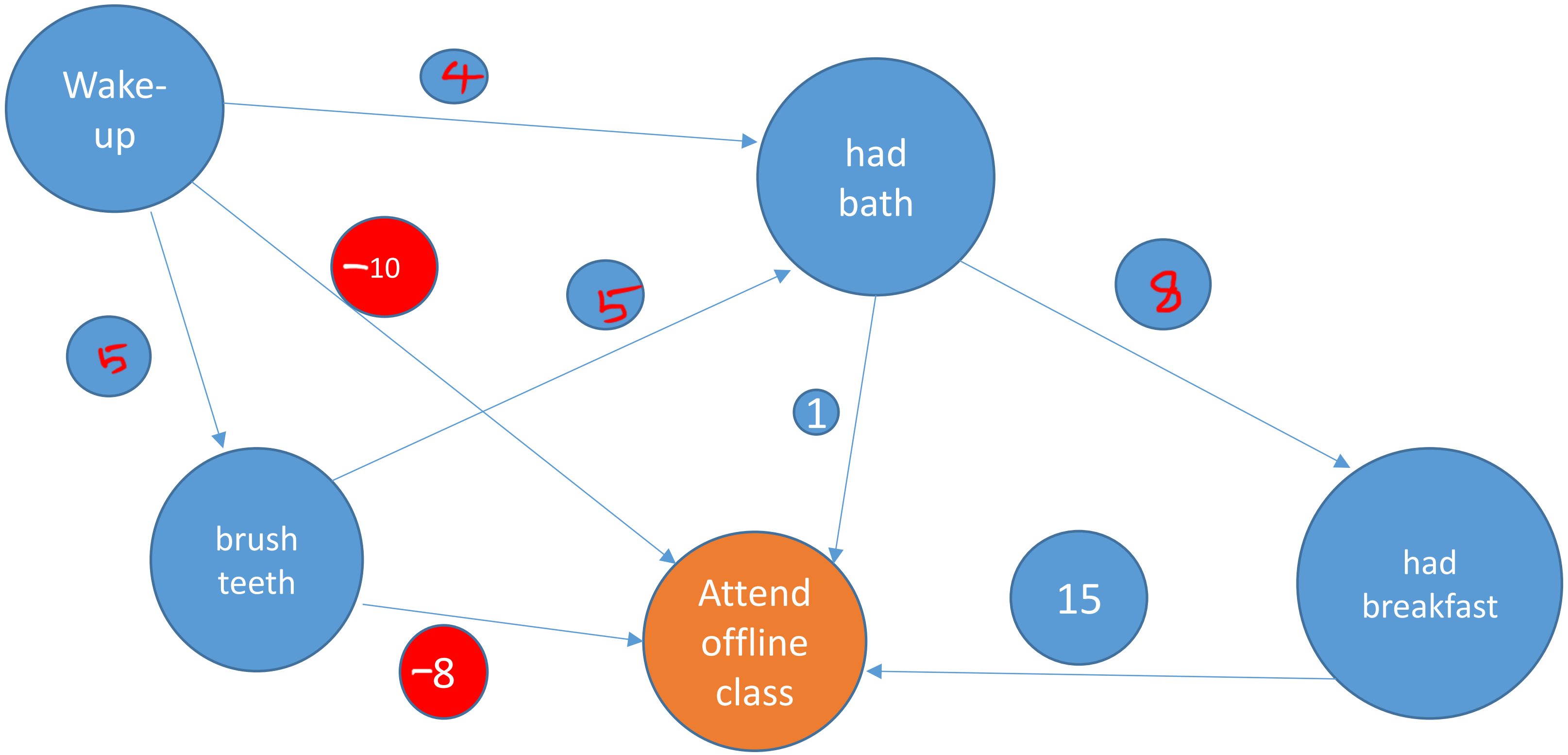
- Speech and Hand Writing Recognition
- Robotics (Robot locomotion)
- Search Engines (Information Retrieval)
- Learning to Classify new astronomical structures
- Medical Diagnosis
- Learning to drive an autonomous vehicle
- Computational Biology/Bioinformatics
- Computer Vision (Object Detection algorithms)
- Detecting credit card fraud
- Stock Market analysis
- Game playing
- ………………..
- …………………

ML solves problems that cannot be solved by numerical means alone

Reinforcement Learning

Reward based learning

- Reinforcement Learning
  - Rewards (+ve or –ve)
  - Wake-up, brushed teeth, had bath, had breakfast, reach college



Types of Machine Learning

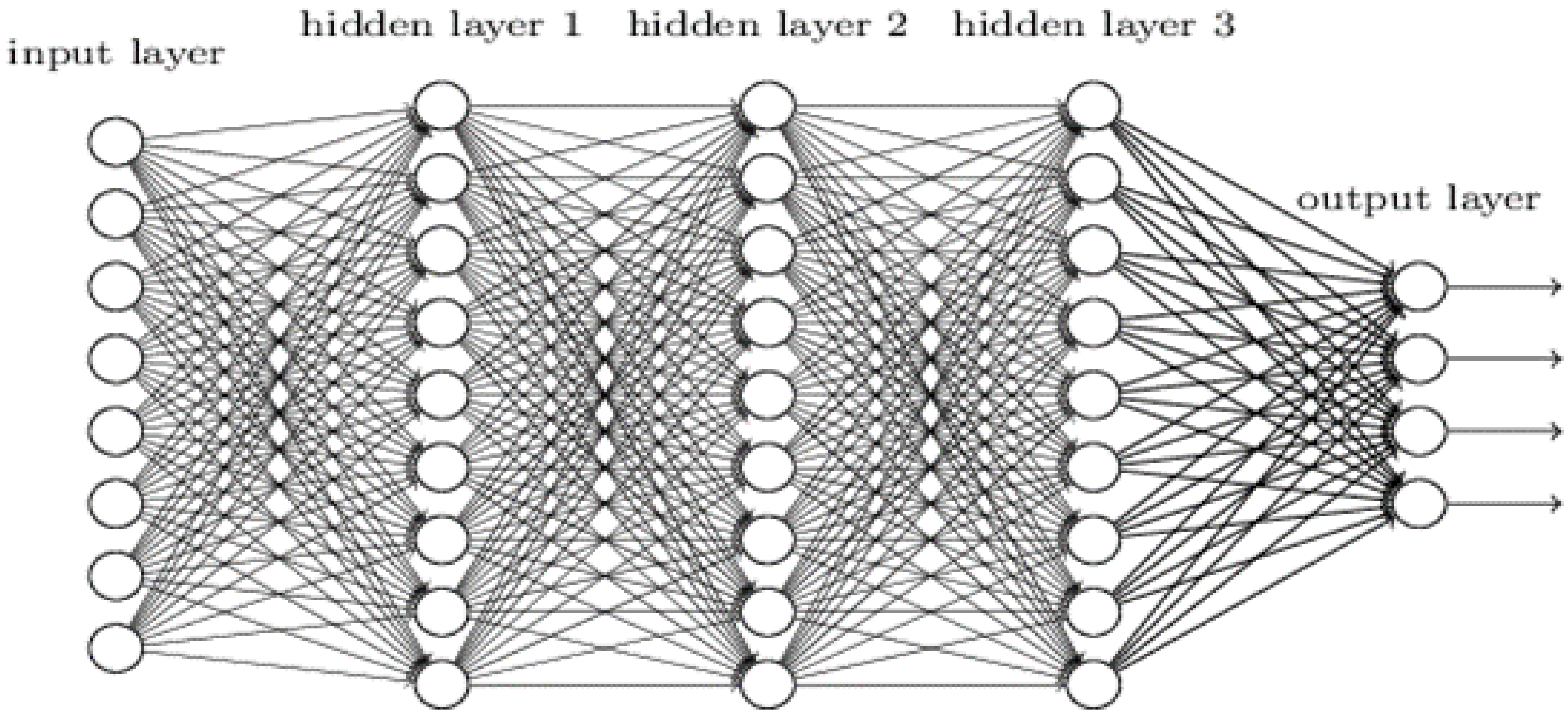
- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning
- Deep Learning

Deep Learning

In the past few years, *Deep Learning* has generated much excitement in Machine Learning

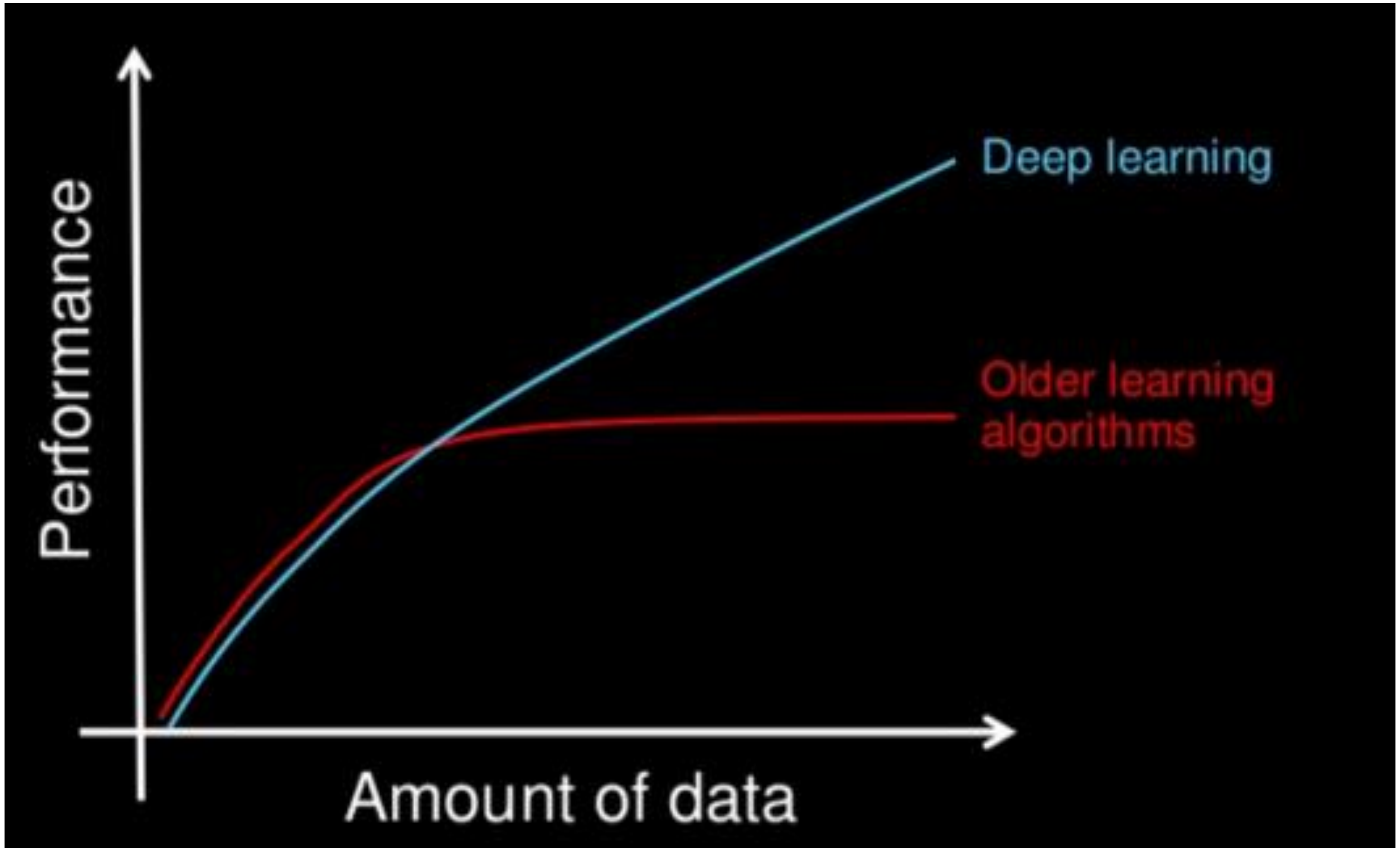
Many breakthrough results in *speech recognition, computer vision and text processing*.

Deep Learning

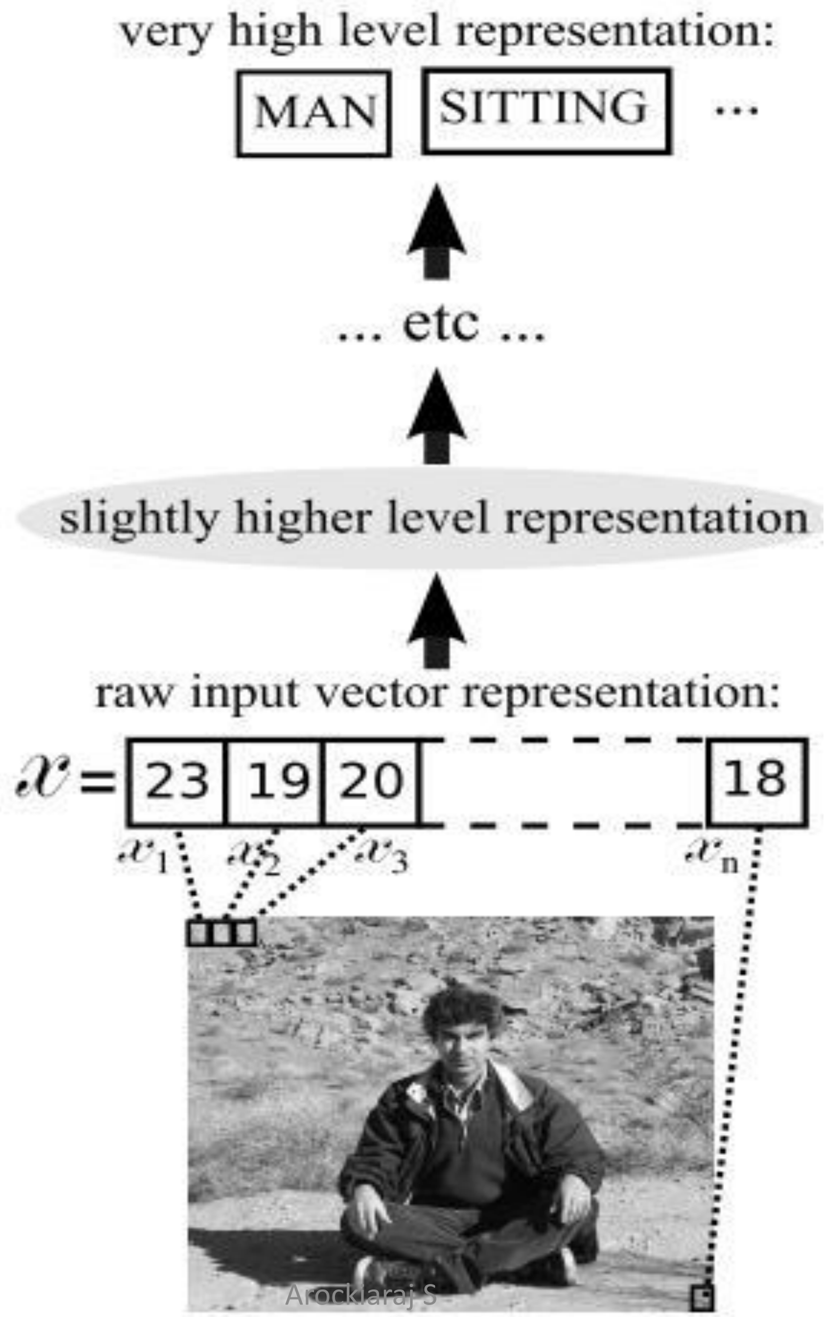


"very <u>large</u> neural networks and <u>huge</u> amounts of data that we have access to"

Why Deep Learning?
- Performance…



Low level representation



Thank you