NUS| SCHOOL OF Computing

AI and ML Fundamentals

Dr. Ai Xin

NUS National University of Singapore | Advanced Computing for Executives
School of Computing

Agenda

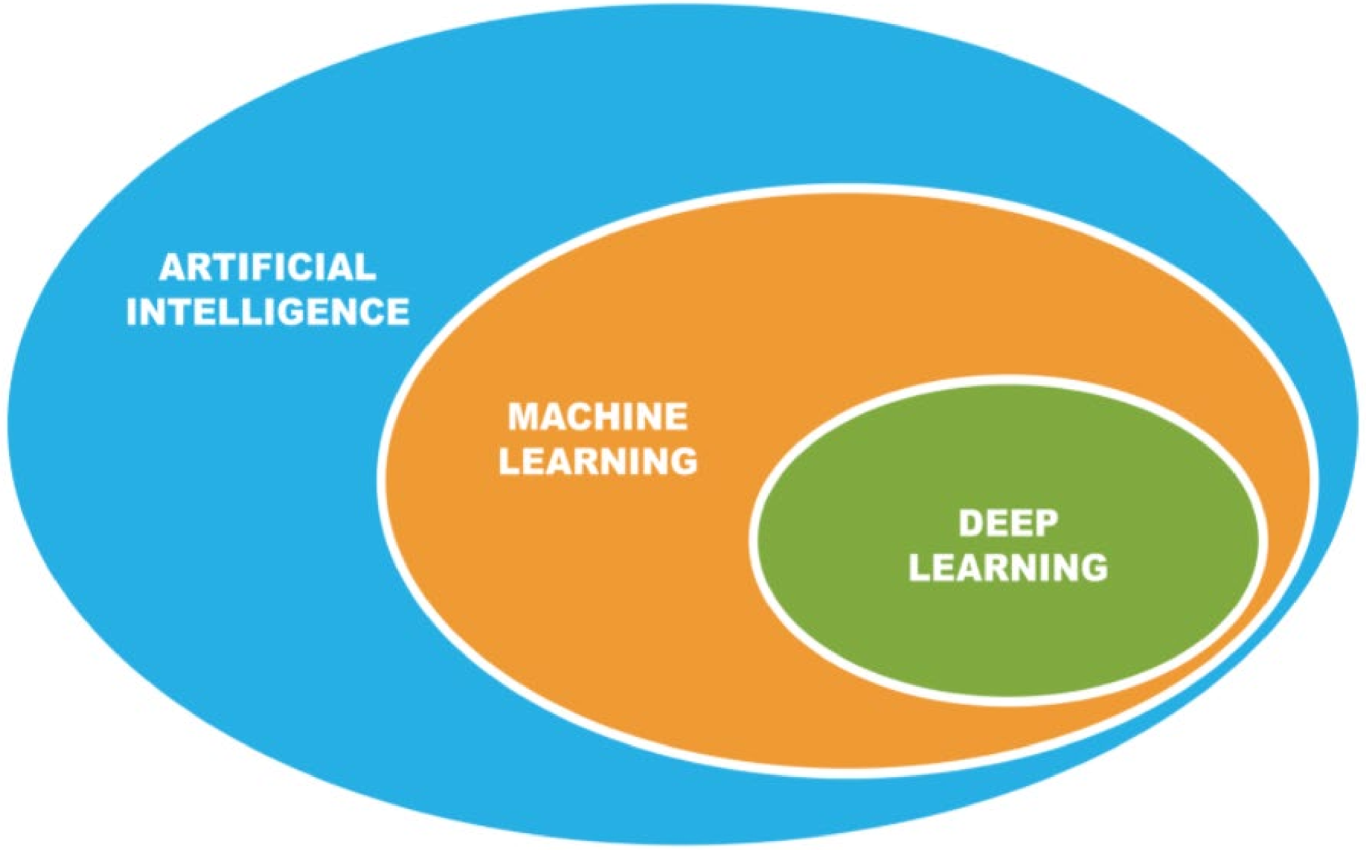
- Day 1: Introduction to AI and ML
  - Introduction and AI applications case study
- Day 2: Machine Learning I
  - Intro, K-Means, KNN and Decision Tree
- Day 3: Machine Learning II
  - Linear Regression, Logistic Regression and Neural Networks
- Day 4: Deep Learning
  - Convolutional Neural Networks for Computer Vision
  - Transformer for Natural Language Processing
- Day 5: Project Presentation

Day 1: Introduction to AI and ML

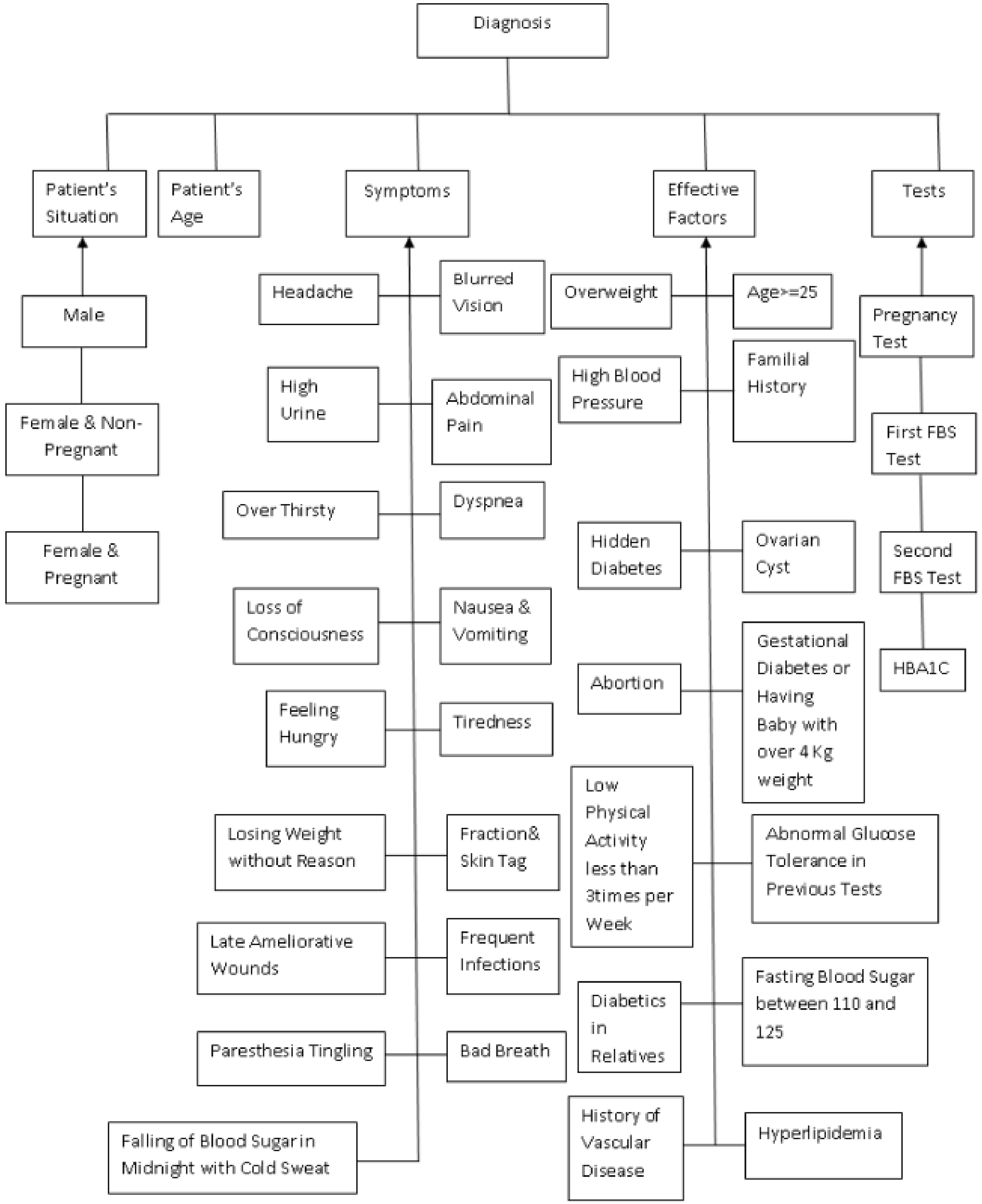
WHAT IS ARTIFICAL INTELLIGENCE

AI definition: the effort to automate intellectual tasks normally performed by humans

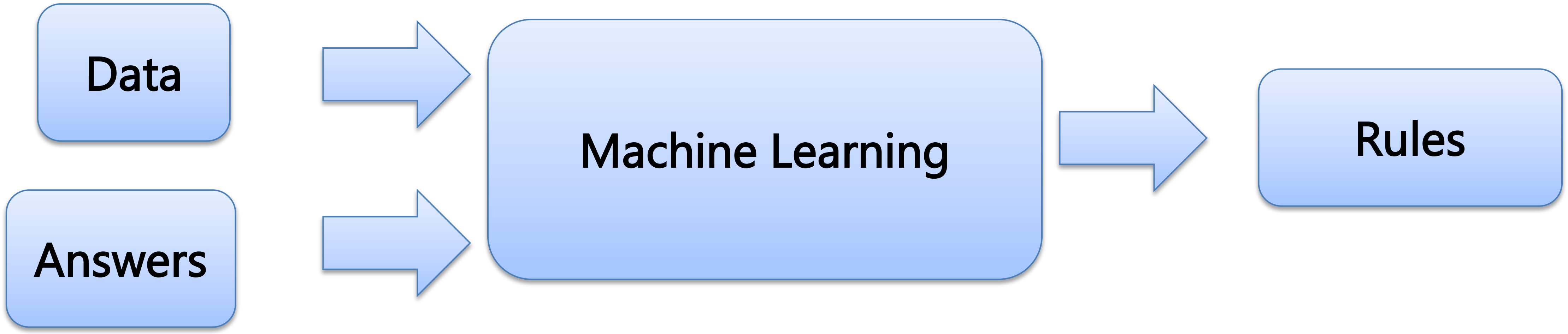
AI is a general field that encompasses machine learning and deep learning, and many other approaches which don’t involve any learning.



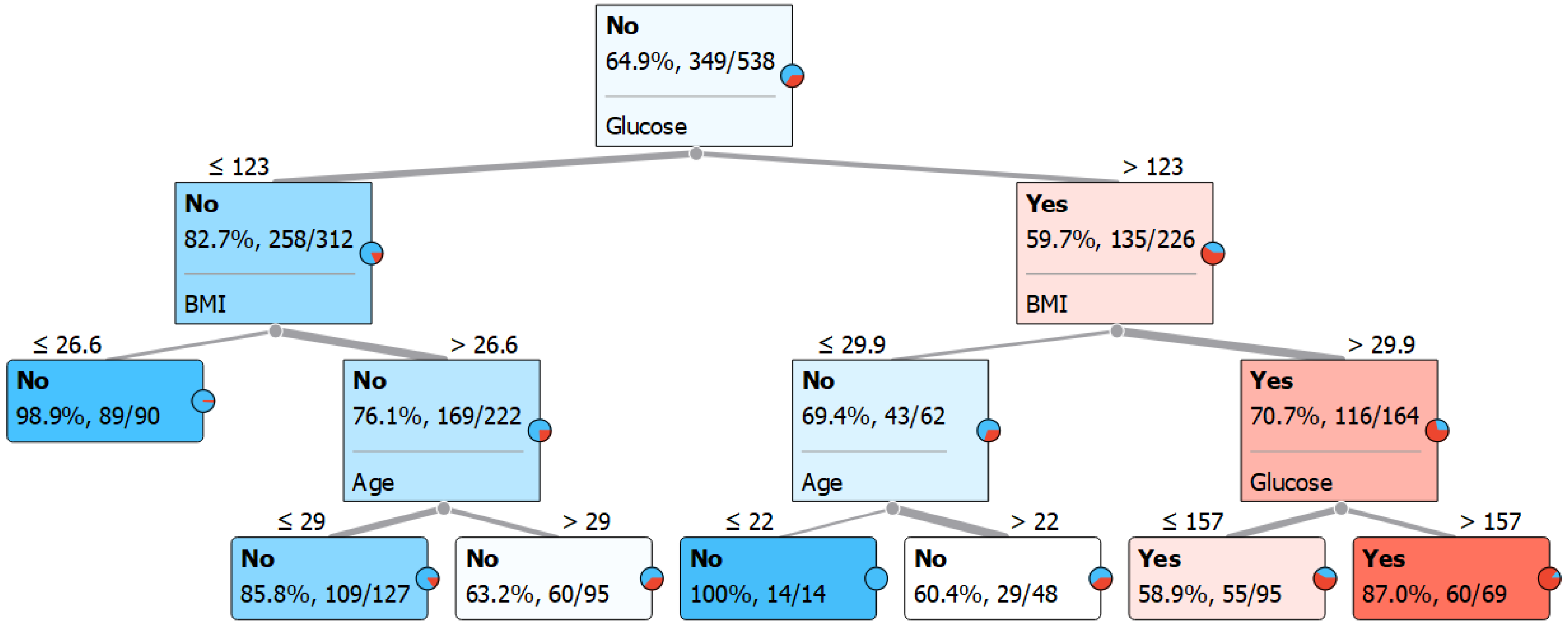
Symbolic AI (1950s- 1980s)



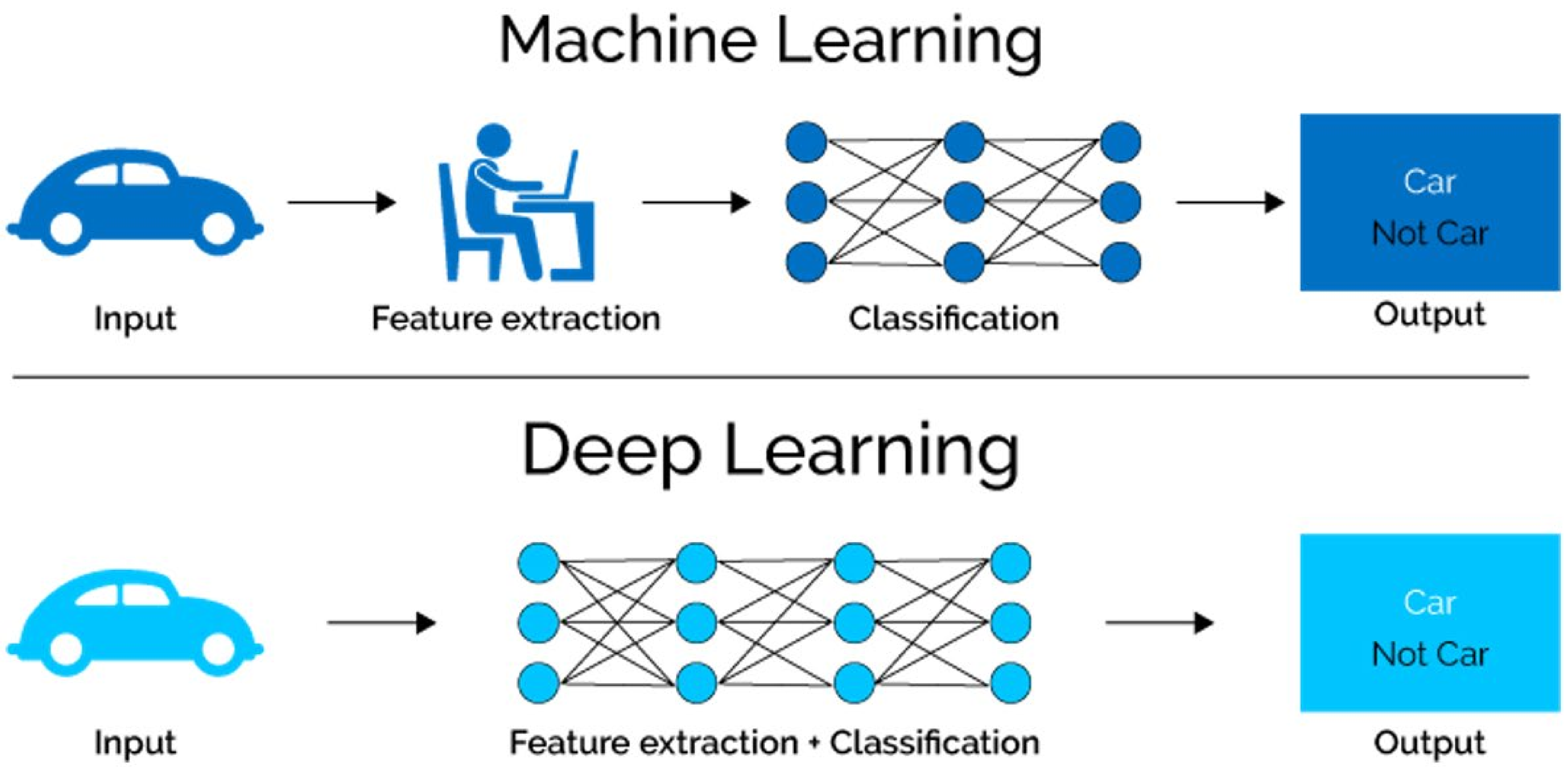
Machine Learning (1990s- now)



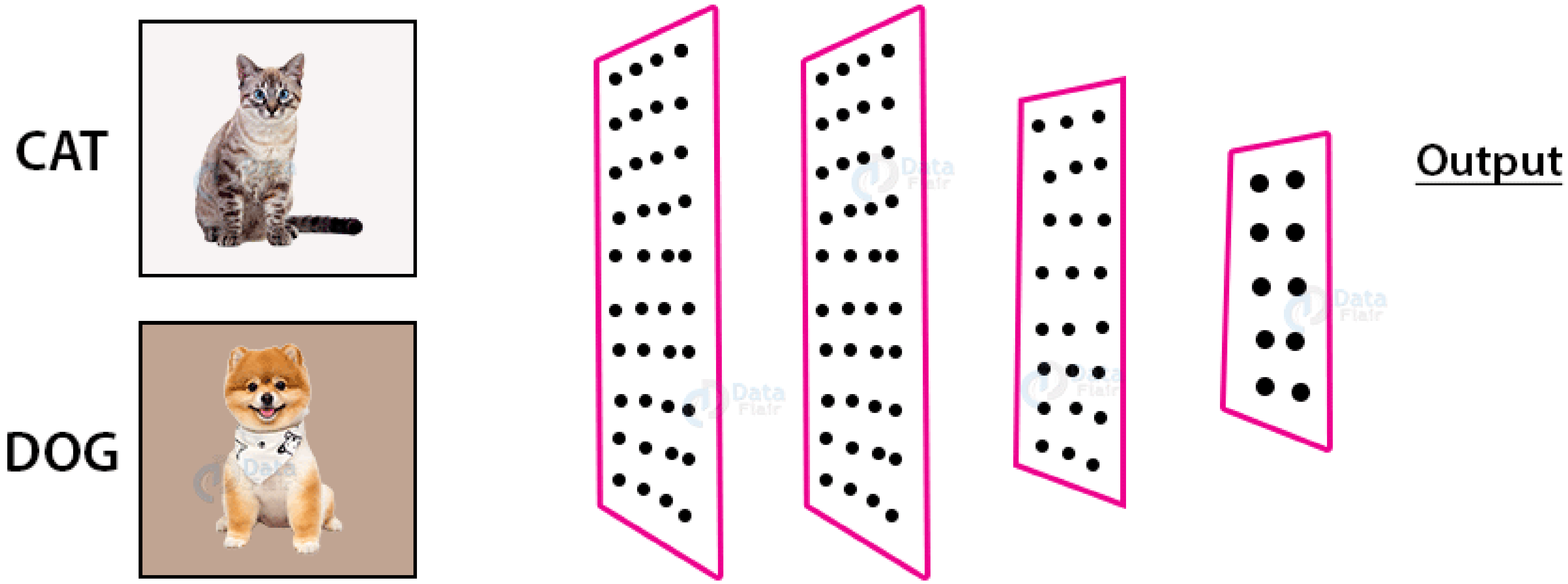
	Outcome	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age
1	Yes	6	148	72	35	0	33.6	0.627	50
2	No	1	85	66	29	0	26.6	0.351	31
3	Yes	8	183	64	0	0	23.3	0.672	32
4	No	1	89	66	23	94	28.1	0.167	21
5	Yes	0	137	40	35	168	43.1	2.288	33
6	No	5	116	74	0	0	25.6	0.201	30
7	Yes	3	78	50	32	88	31.0		
8	No	10	115	0	0	0	35.3		
9	Yes	2	197	70	45	543	30.5		
10	Yes	8	125	96	0	0	0.0		
11	No	4	110	92	0	0	37.6		
12	Yes	10	168	74	0	0	38.0		
13	No	10	139	80	0	0	27.1		
14	Yes	1	189	60	23	846	30.1		
15	Yes	5	166	72	19	175	25.8		
16	Yes	7	100	0	0	0	30.0		
17	Yes	0	118	84	47	230	45.8		
18	Yes	7	107	74	0	0	29.6		
19	No	1	103	30	38	83	43.3		

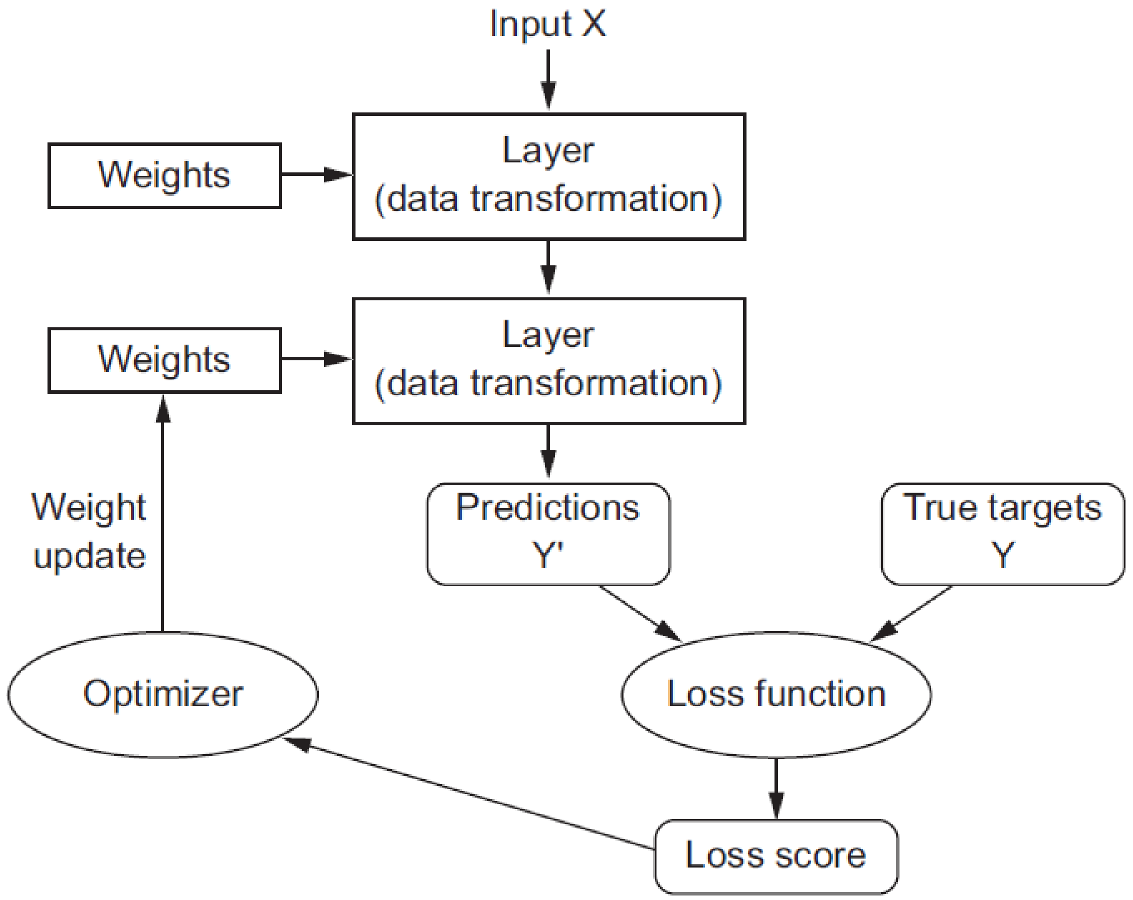


Deep Learning (2010s- now)



Deep Learning (2010s- now)

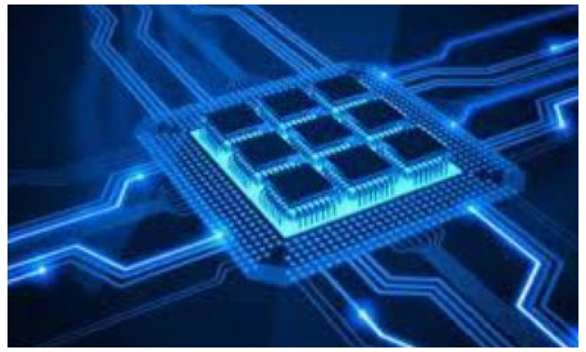




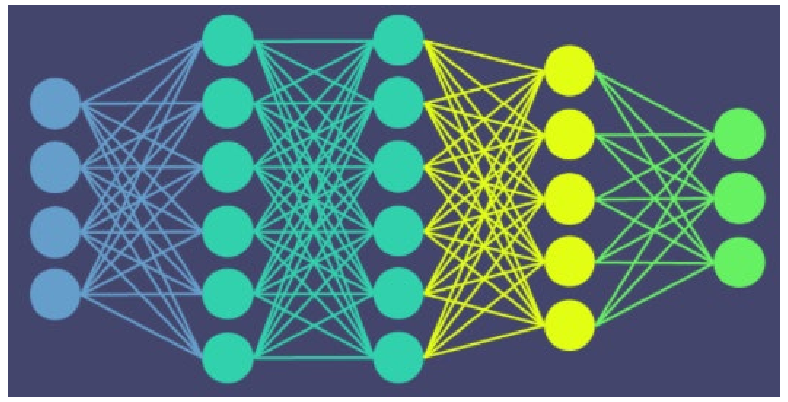
WHY NOW?

Big Data

Computing Power



Algorithms

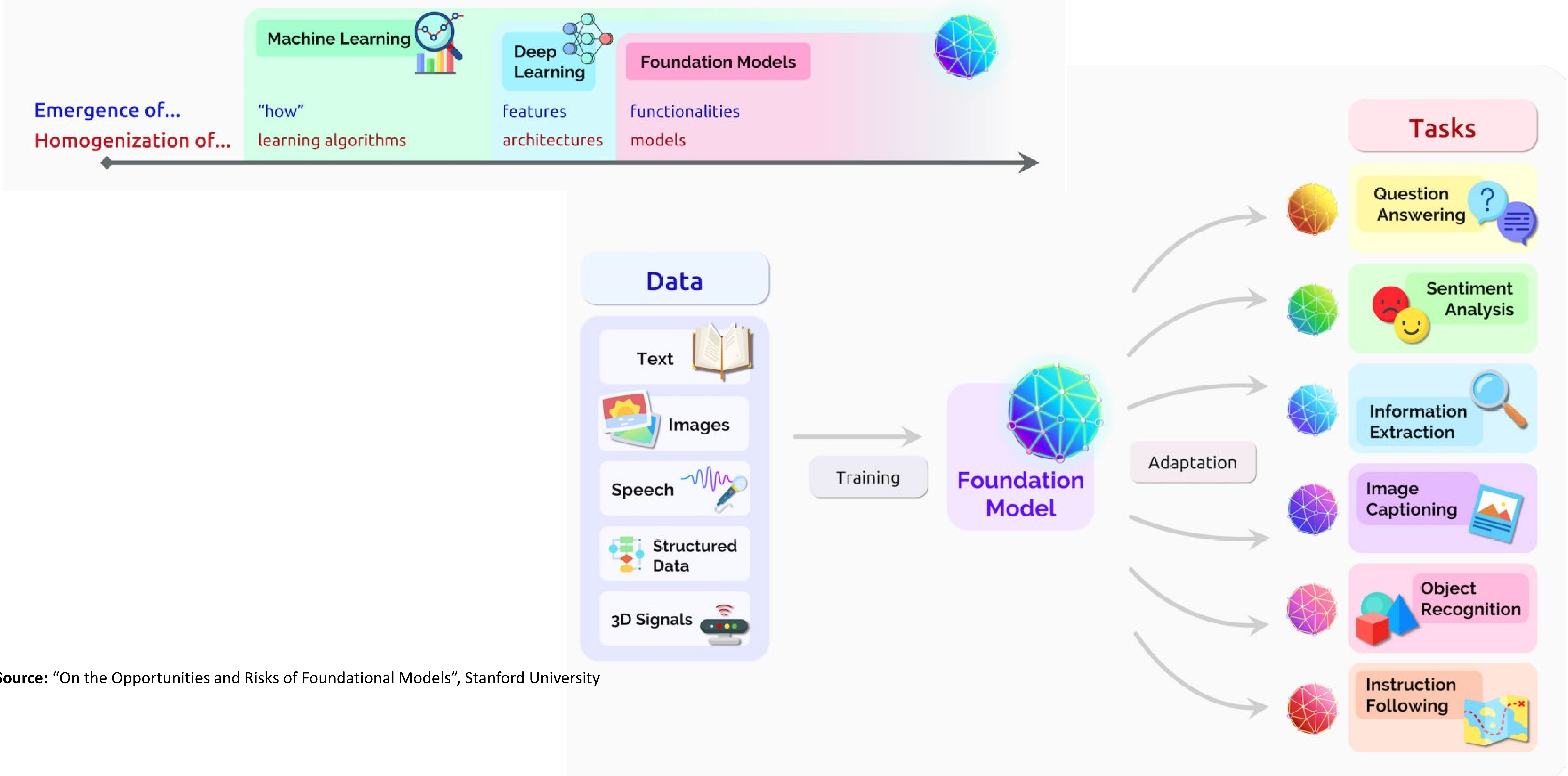


Data is Critical To Breakthroughs in AI

Year	Breakthroughs in AI	Datasets (First Available)	Algorithms (First Proposed)
1994	Human-level read-speech recognition	Spoken Wall Street Journal articles and other texts (1991)	Hidden Markov Model (1984)
1997	IBM Deep Blue defeated Garry Kasparov	700,000 Grandmaster chess games, aka "The Extended Book" (1991)	Negascout planning algorithm (1983)
2005	Google's Arabic- and Chinese-to-English translation	1.8 trillion tokens from Google Web and News pages (collected in 2005)	Statistical machine translation algorithm (1988)
2011	IBM Watson became the world Jeopardy! champion	8.6 million documents from Wikipedia, Wiktionary, Wikiquote, and Project Gutenberg (updated in 2010)	Mixture-of-Experts algorithm (1991)
2014	Google's GoogleNet object classification at near-human performance	ImageNet corpus of 1.5 million labeled images and 1,000 object categories (2010)	Convolutional neural network algorithm (1989)
2015	Google's Deepmind achieved human parity in playing 29 Atari games by learning general control from video	Arcade Learning Environment dataset of over 50 Atari games (2013)	Q-learning algorithm (1992)
Average No. of Years to Breakthrough:		**3 years**	**18 years**

Source: Train AI 2017, https://www.crowdflower.com/train-ai/

Generative AI (2020s- now)



Source: "On the Opportunities and Risks of Foundational Models", Stanford University

Case Study:
How can AI help business in the real world?

Case Study 1: Smart Farming

PROBLEM

The efficient use of farmland is critical, especially in view of the increased use of pesticide chemicals, which brings environmental risks and direct hazards for human health.

SOLUTION

Use computer vision techniques to sense where crops are threatened by pests, and control robotic equipment to fire accurate blasts of pesticide chemicals at the affected crops, while leaving others untouched.

Case Study 2: Smart Lift

KONE

PROBLEM

When a lift breaks down, it affects a lot of people significantly and it takes time for the workers to find the problem and fix it.

SOLUTION

KONE has connected more than 1 million of its escalators and elevators to the cloud. They are fitted with sensors to collect the data and machine learning algorithms are trained on this data to "understand" when faults or breakdowns are likely to occur.

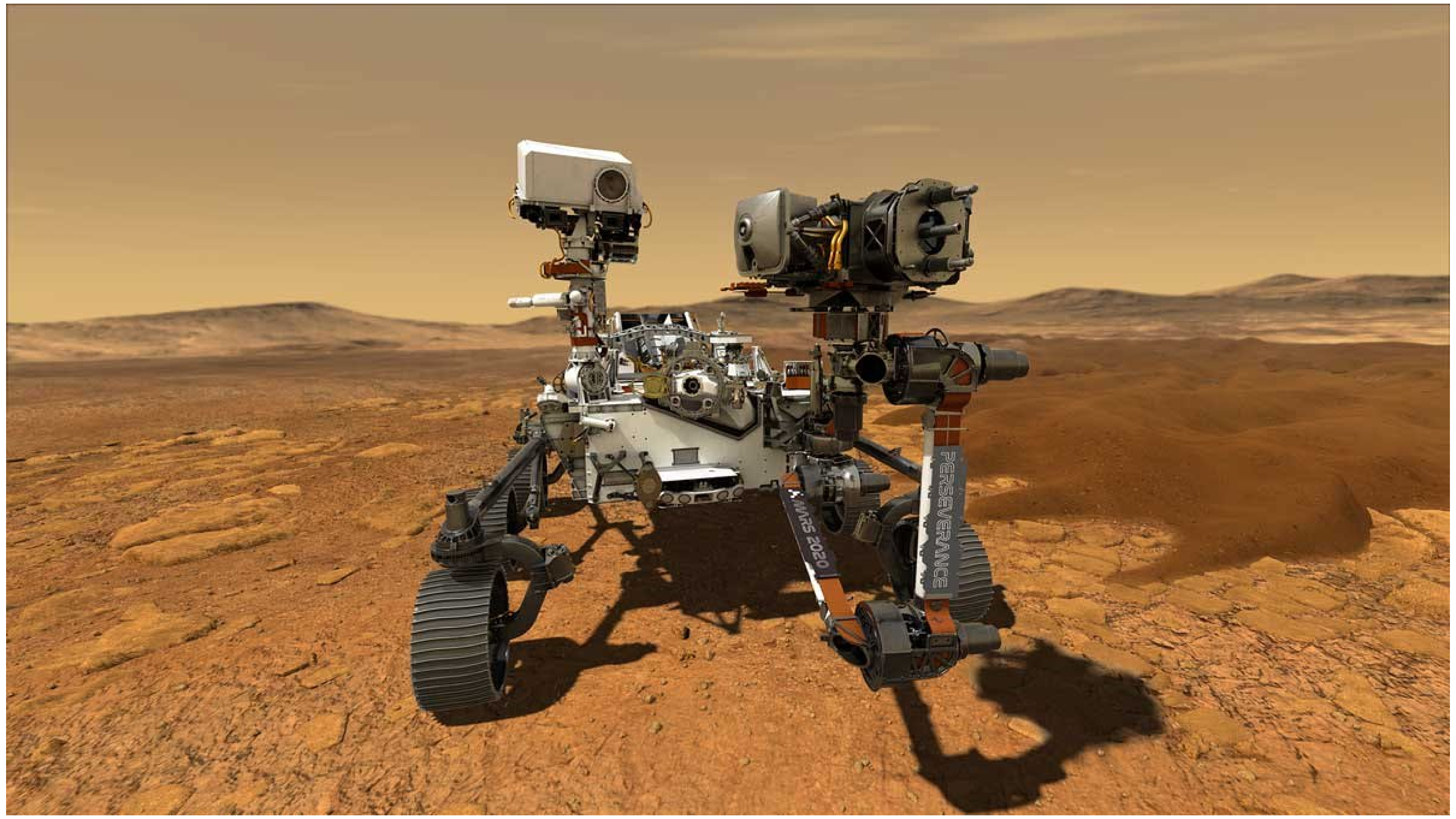
Case Study 3: Autonomous Machine

PROBLEM

One of the biggest obstacles in space exploration is the limited amount of bandwidth available for sending information back to Earth.

SOLUTION

Space exploration generates huge volumes of data, and it is far more efficient to use autonomous machines to work out what is worth sending home and what can be discarded.



Case Study 4: Pizza Checker

PROBLEM

If pizzas are cooked or delivered that do not meet their expectations in terms of consistency and quality, customers will be dissatisfied.

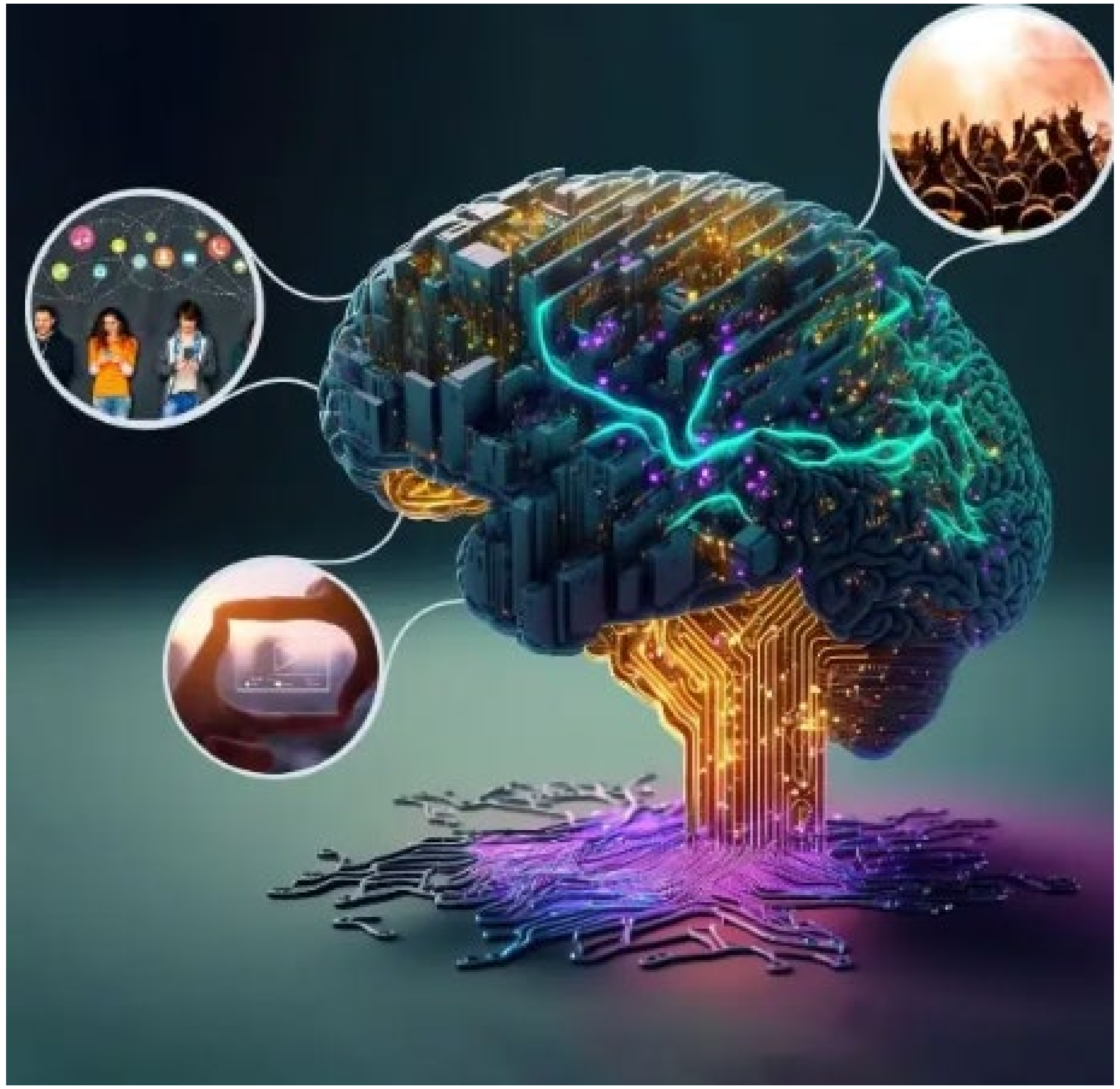
SOLUTION

Photographs every pizza when it leaves the oven, and then uses deep learning algorithms to inspect it for quality before it reaches the customer.

Case Study 5: GenAI in Media

SUMMARY

- Write Stories for the newsrooms
- Help journalists digest information, create summaries, create video content and more
- In the worlds of sports broadcasting, generate real-time commentary in multiple languages and create interactive, personalized features for viewers
- Post-production processes for film: generate realistic visual effects and streamlining editing
- Revolutionize music creation
- For artists, generate new images, manipulate existing images and even complete unfinished art works



Case Study 6: GenAI in Healthcare

SUMMARY

- Deliver personalized advice to patients
- Interpreting medical images and generating reports based on the images
- Personalized treatment
- Streamlining medical notetaking, dealing with routing calls and enquires at clinics
- In general, make healthcare more accessible and efficient

Each group to study one case and share with the others

Project

- Each Group to propose a real-life AI and ML solution in the below industries
  - Healthcare Industry (Group 1)
    - E.g. diagnosing diseases, analyzing medical images, predicting patient outcomes, drug discovery, personalized medicine, and improving healthcare operations.
  - Finance Industry (Group 2)
    - E.g. fraud detection, algorithmic trading, risk assessment, credit scoring, customer service automation, and financial market analysis.
  - Retail and E-commerce Industry (Group 3)
    - E.g. personalized shopping experiences, recommendation systems, demand forecasting, inventory management, pricing optimization, and supply chain optimization.
  - Manufacturing (Group 4)
    - E.g. predictive maintenance, quality control, process optimization, supply chain management, robotics, and autonomous vehicles in manufacturing and industrial settings

Project

- Each Group to propose a real-life AI and ML solution in the below industries
  - Transportation and Logistics (Group 5)
    - E.g. route optimization, demand forecasting, fleet management, autonomous vehicles, traffic pattern analysis, and supply chain optimization
  - Education (Group 6)
    - E.g. adaptive learning platforms, intelligent tutoring systems, personalized learning experiences, automated grading, and plagiarism detection.
  - Entertainment and Media (Group 7)
    - content recommendation, personalized advertising, sentiment analysis, speech recognition, and natural language processing in the entertainment and media industry.
- Let me know if you want to work on problems in other areas

Project

- Each group will need to present with
  - Presentation Slides:
    - Problem
    - Challenges
    - Proposed Solution
  - Jupyter Notebook Demo: a simple implementation to show the feasibility of your proposal / solution
    - Collect simple data/image/text from internet or create by yourselves
    - Build or Load AI and ML model, and apply the model on the data/image/text using
      - Sklearn
      - Tensorflow
      - Transformer (Hugging Face)
      - and etc.

Day 2: Machine Learning I
(Intro, K-Means, KNN, Decision Tree)

What is Machine Learning?

What is Machine Learning?

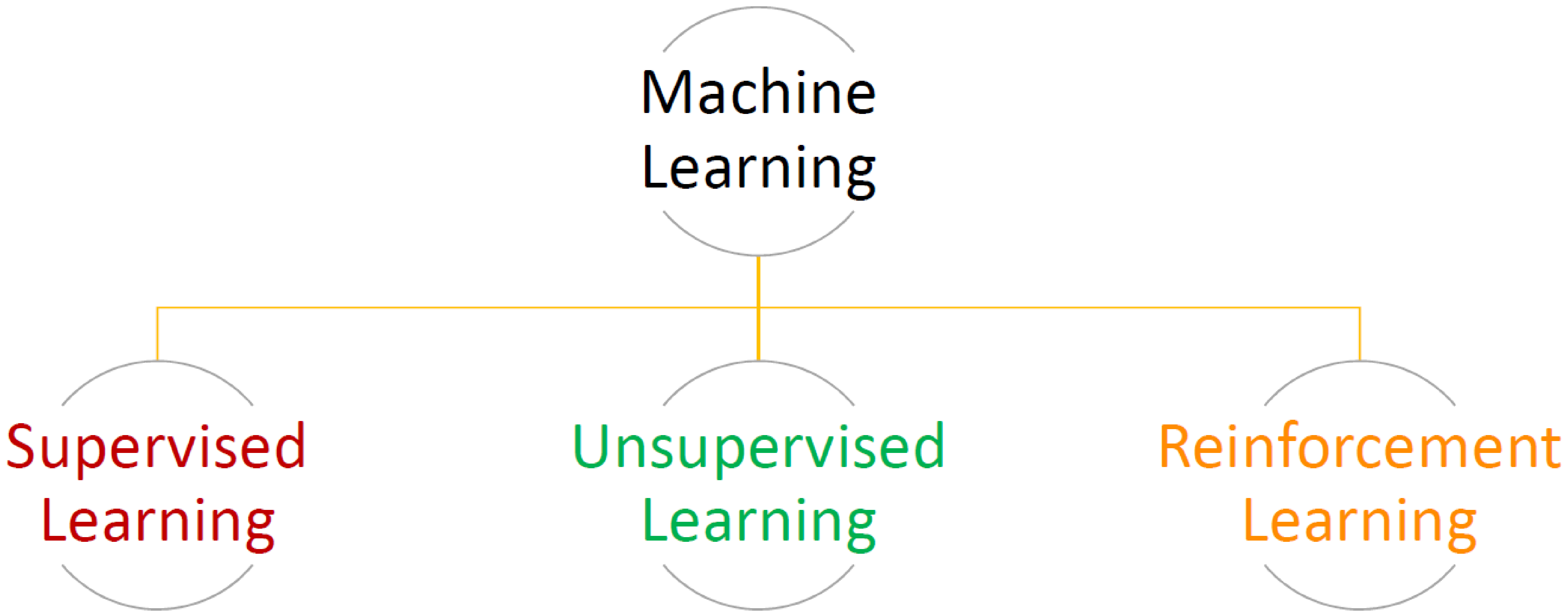
- Machine Learning is the science (and art) of programming computers so they can learn from data.

- A slightly more general definition:

  [Machine Learning is the] field of study that gives computers the ability to learn without being explicitly programmed.

  —Arthur Samuel, 1959

Types of Machine Learning



Supervised vs. unsupervised

DATA DEFINITIONS: LABELED AND UNLABELED DATA

No Target (labels) **Variables or Features**

We do not know what type of iris flowers they are.

There is no labels.

?

	sepal length	sepal width	petal length	petal width
45	5.100	3.800	1.900	0.400
46	4.800	3.000	1.400	0.300
47	5.100	3.800	1.600	0.200
48	4.600	3.200	1.400	0.200
49	5.300	3.700	1.500	0.200
50	5.000	3.300	1.400	0.200
51	7.000	3.200	4.700	1.400
52	6.400	3.200	4.500	1.500
53	6.900	3.100	4.900	1.500
54	5.500	2.300	4.000	1.300

Target (labels) **Variables or Features**

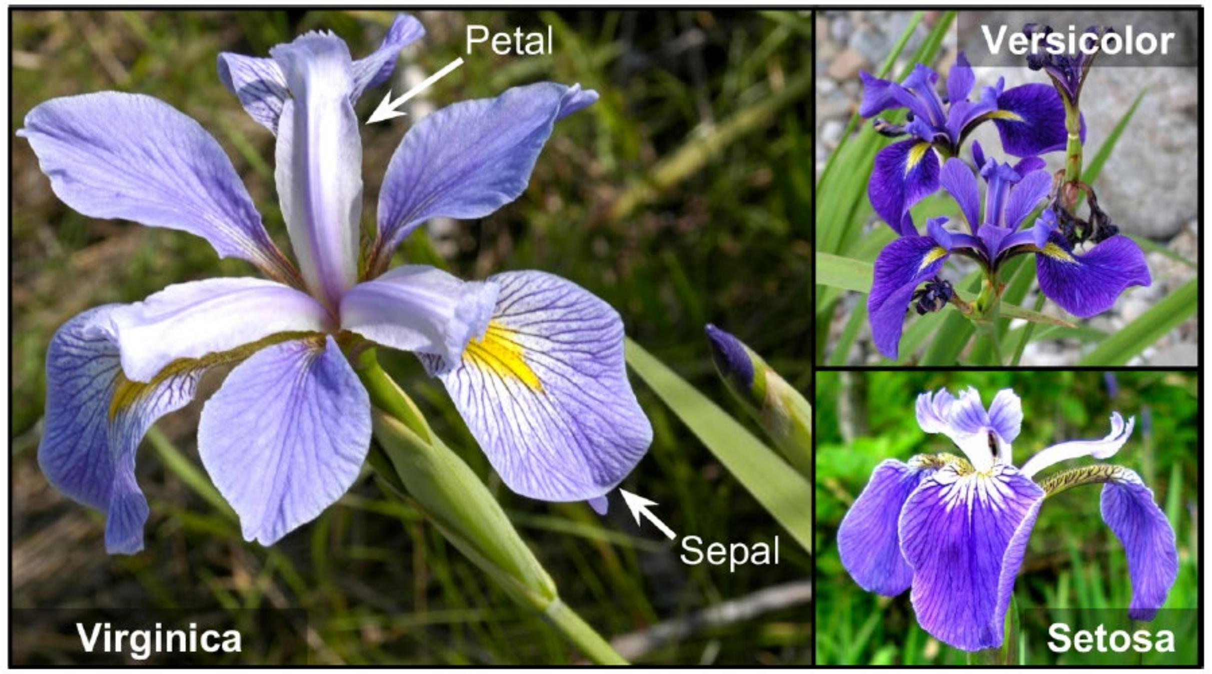
We collected the data from a known group of iris flower types.

The iris column is the target or labels.

Hence labeled data.

Classification or Prediction techniques are used in these cases.

	iris	sepal length	sepal width	petal length	petal width
43	Iris-setosa	4.400	3.200	1.300	0.200
44	Iris-setosa	5.000	3.500	1.600	0.600
45	Iris-setosa	5.100	3.800	1.900	0.400
46	Iris-setosa	4.800	3.000	1.400	0.300
47	Iris-setosa	5.100	3.800	1.600	0.200
48	Iris-setosa	4.600	3.200	1.400	0.200
49	Iris-setosa	5.300	3.700	1.500	0.200
50	Iris-setosa	5.000	3.300	1.400	0.200
51	Iris-versicolor	7.000	3.200	4.700	1.400
52	Iris-versicolor	6.400	3.200	4.500	1.500
53	Iris-versicolor	6.900	3.100	4.900	1.500
54	Iris-versicolor	5.500	2.300	4.000	1.300
55	Iris-versicolor	6.500	2.800	4.600	1.500



Classification vs. Regression in Supervised Learning

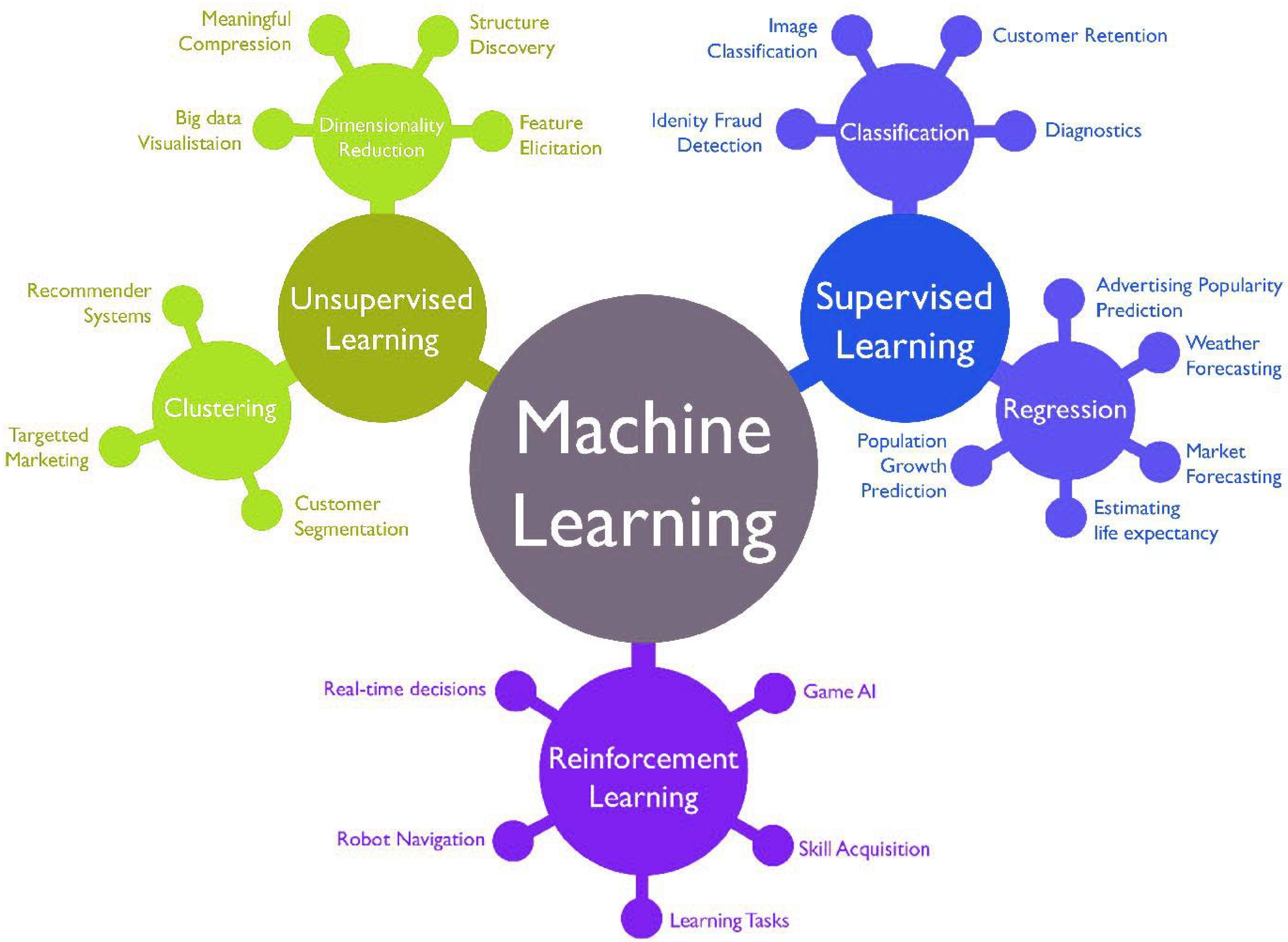
- Regression

month	town	flat_type	block	street_name	storey_range	floorAreaSqm	resale_price
2012-09	CHOA CHU KANG	4 ROOM	119	TECK WHYE LANE	04 TO 06	104	**400000**
2013-10	JURONG WEST	3 ROOM	510	JURONG WEST ST 52	13 TO 15	74	**375000**
2013-03	JURONG EAST	5 ROOM	284	TOH GUAN RD	07 TO 09	120	**655000**

- Classification

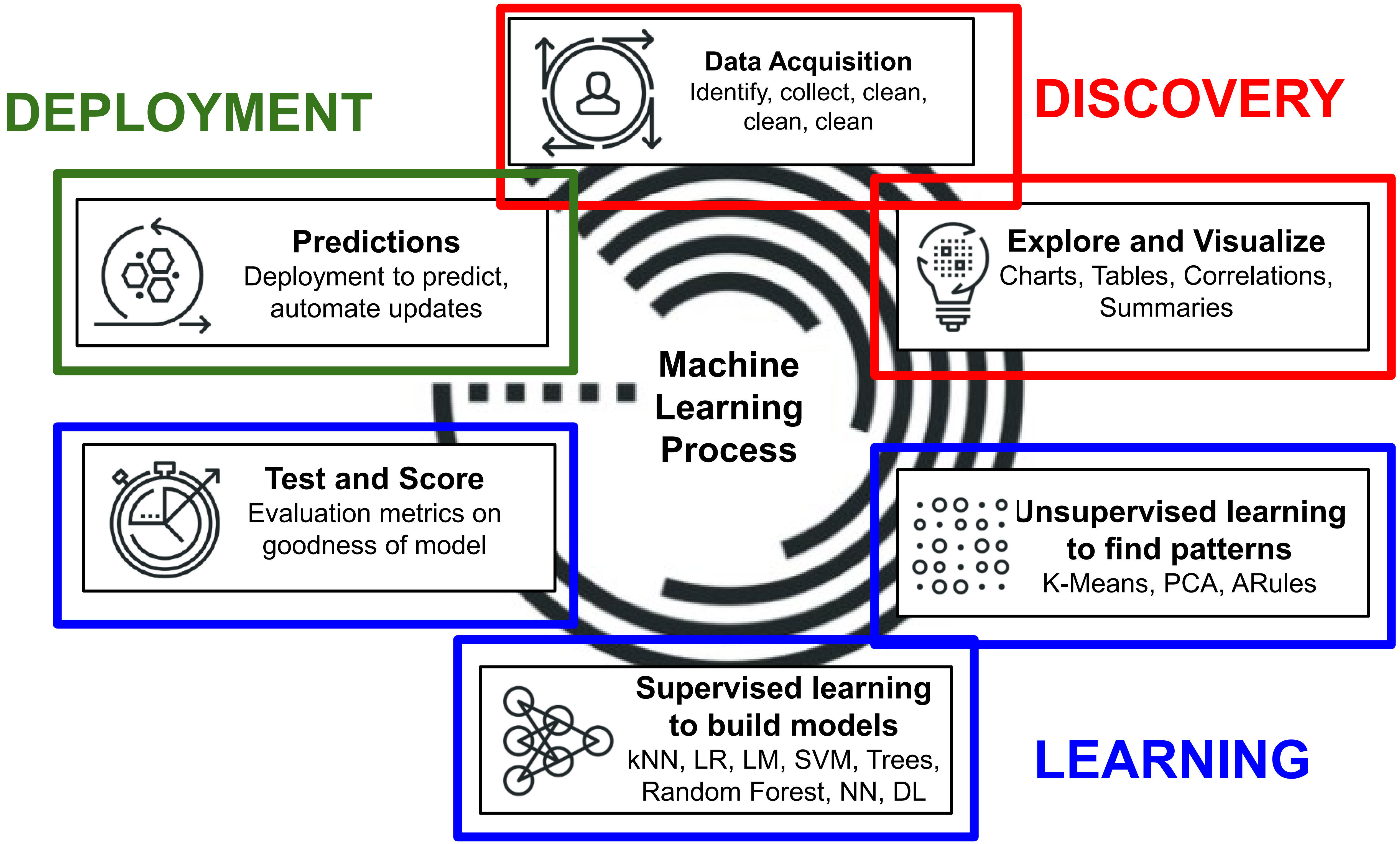
month	town	flat_type	block	street_name	storey_range	floorAreaSqm	resale_price
2012-09	CHOA CHU KANG	4 ROOM	119	TECK WHYE LANE	04 TO 06	104	**Medium**
2013-10	JURONG WEST	3 ROOM	510	JURONG WEST ST 52	13 TO 15	74	**Low**
2013-03	JURONG EAST	5 ROOM	284	TOH GUAN RD	07 TO 09	120	**High**

Machine Learning Approaches



[HTTPS://WWW.WORDSTREAM.COM/BLOG/WS/2017/07/28/MACHINE-LEARNING-APPLICATIONS](https://www.wordstream.com/blog/ws/2017/07/28/machine-learning-applications), ABDUL WAHID

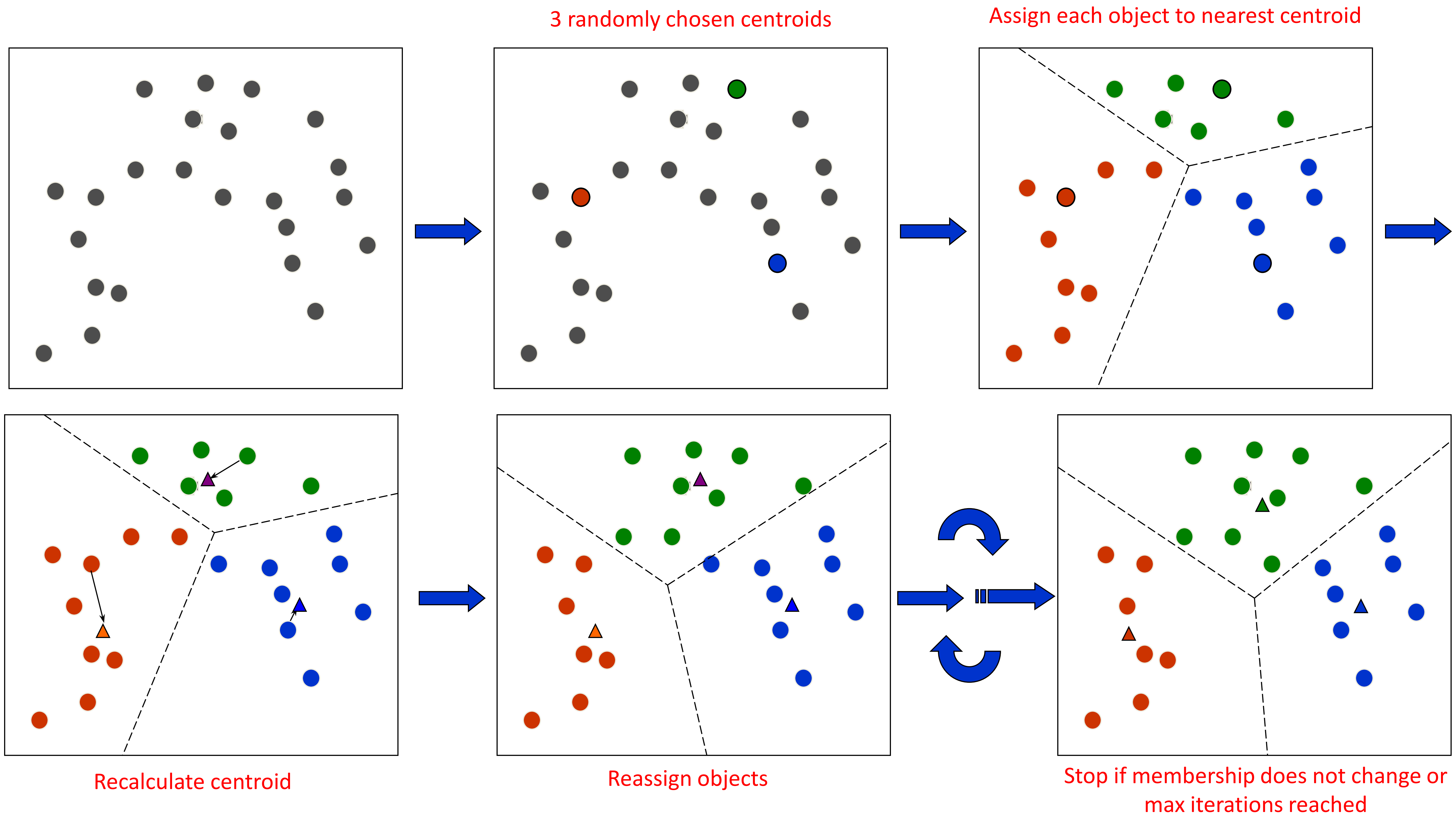
Machine Learning Process



k-means Clustering

- Steps to perform k-means clustering:
  1. Define the number of clusters (k).
  2. Choose k data points randomly to serve as the initial centroids for the k clusters.
  3. Assign each data point to the cluster represented by its nearest centroid.
  4. Find a new centroid for each cluster by calculating the mean vector of its members.
  5. Undo the memberships of all data objects. Repeat steps 3 to 5 until cluster membership no longer changes.

k-means Clustering - Illustration



k-means Clustering

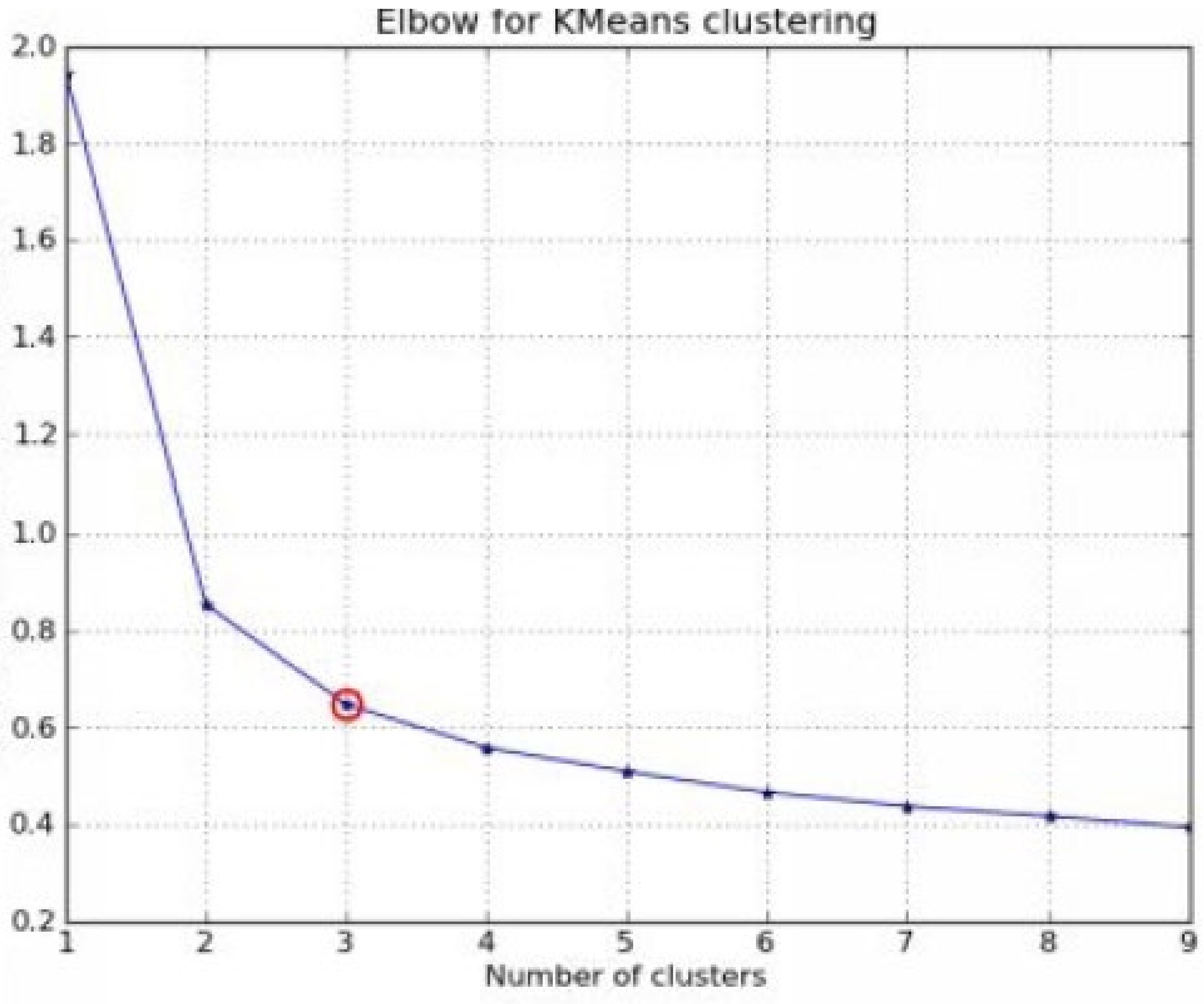
- Strengths:
  - Algorithm is efficient and easy to implement
- Weaknesses:
  - May not know what the value of k should be beforehand
  - Sensitive to the choice of initial k centroids: clusters tend to converge to a local optimum solution
  - Sensitive to noise

k-means Clustering

- Determine the value of k using Elbow Method
  - Sum of Squared Errors (SSE) at each number of clusters is calculated and graphed
  - Look for a change of slope from steep to shallow (an elbow) to determine the optimal number of clusters
  - This method is inexact, but still potentially helpful

k-means Clustering

- Elbow Method
  - SSE measures the distances from each data object to the cluster centroid, take a squared value on the distances and then sum them up together.
  - We can see the larger K the lower SSE, but beyond certain point (e.g. K=3) adding more clusters will not reduce SSE significant, therefore K=3 is the desired K.



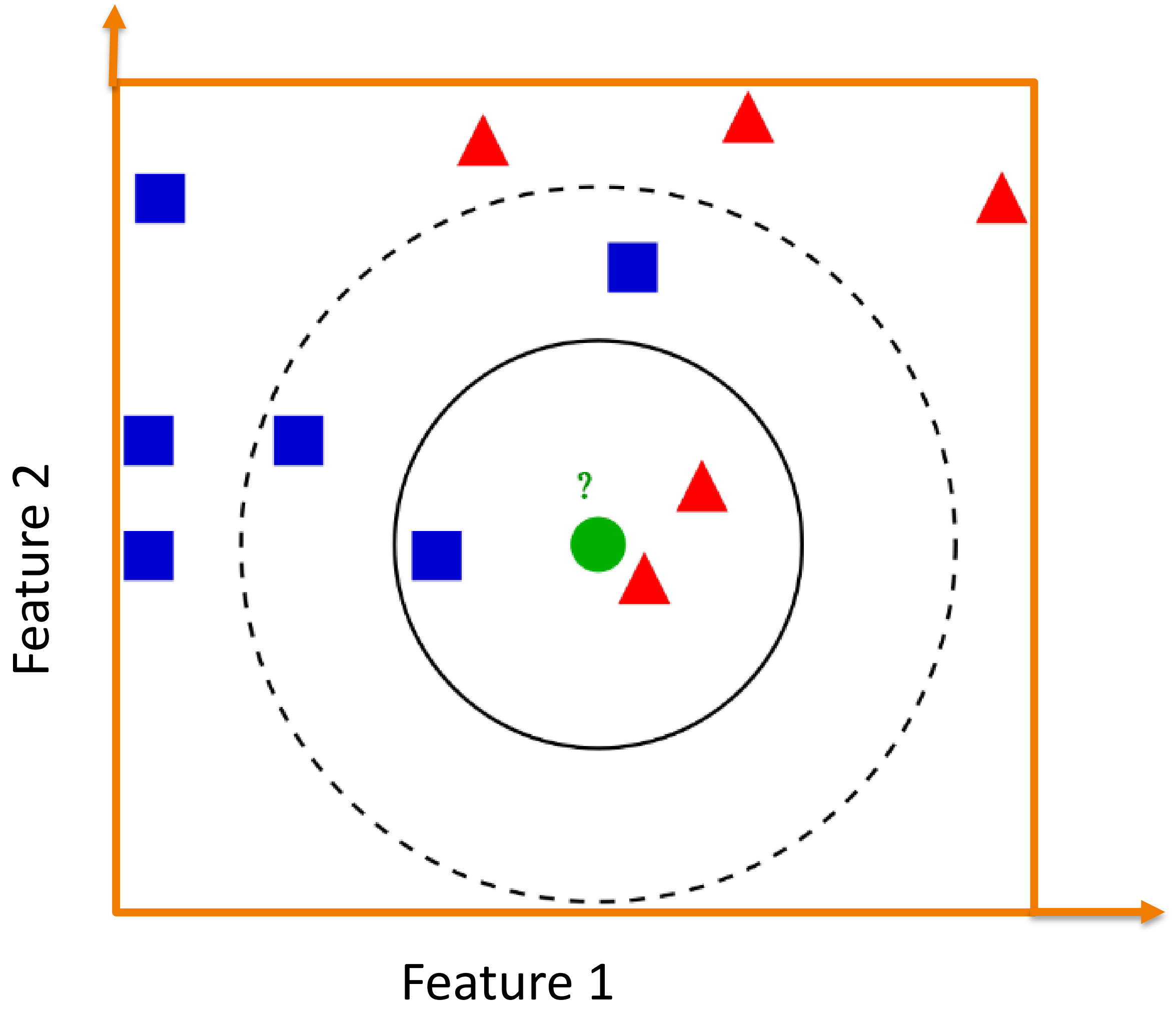
Lab 1: K-Means

K-Nearest Neighbors (KNN)

- The k-nearest neighbors algorithm (k-NN) is a non-parametric method used for classification and regression. In both cases, each input sample has k closest training samples (“neighbors”) in the feature space.
- In k-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors.
- In k-NN regression, the output is the estimated value for the object. This value is the average of the values of k nearest neighbors (function estimation using KNN).

K-Nearest Neighbors (KNN)

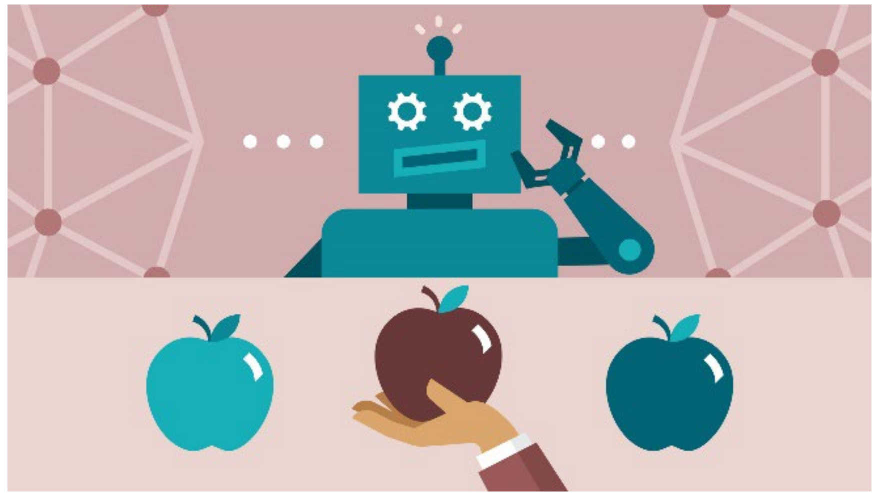
- Example of k-NN classification. The test sample (green dot) should be classified either to blue squares or to red triangles.
- If k = 3 (solid line circle) it is assigned to the red triangles because there are 2 triangles and only 1 square inside the inner circle.
- If k = 5 (dashed line circle) it is assigned to the blue squares (3 squares vs. 2 triangles inside the outer circle).



k=3 vs k=5

WWW.WIKIPEDIA.ORG

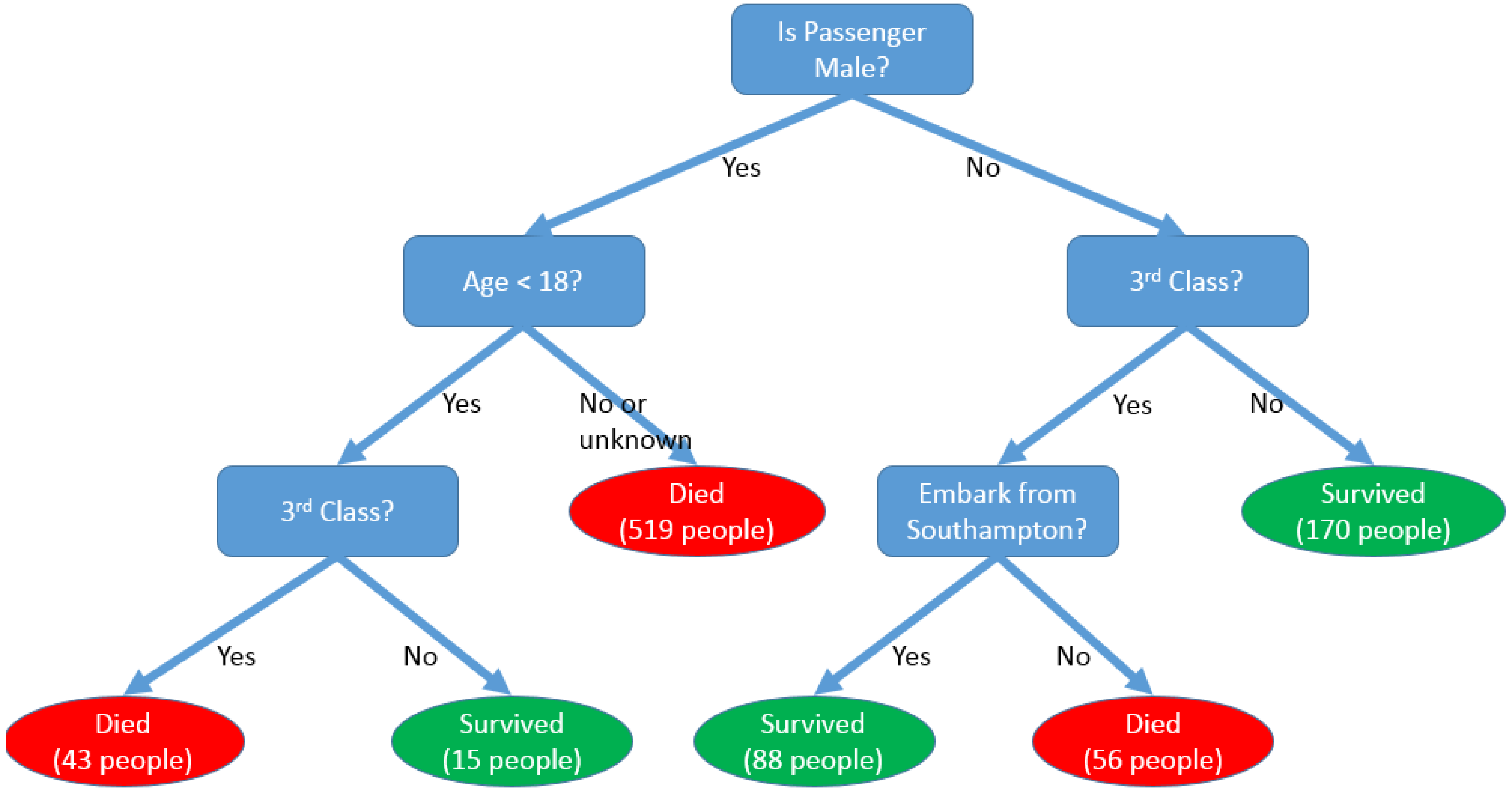
K-Nearest Neighbors (KNN)



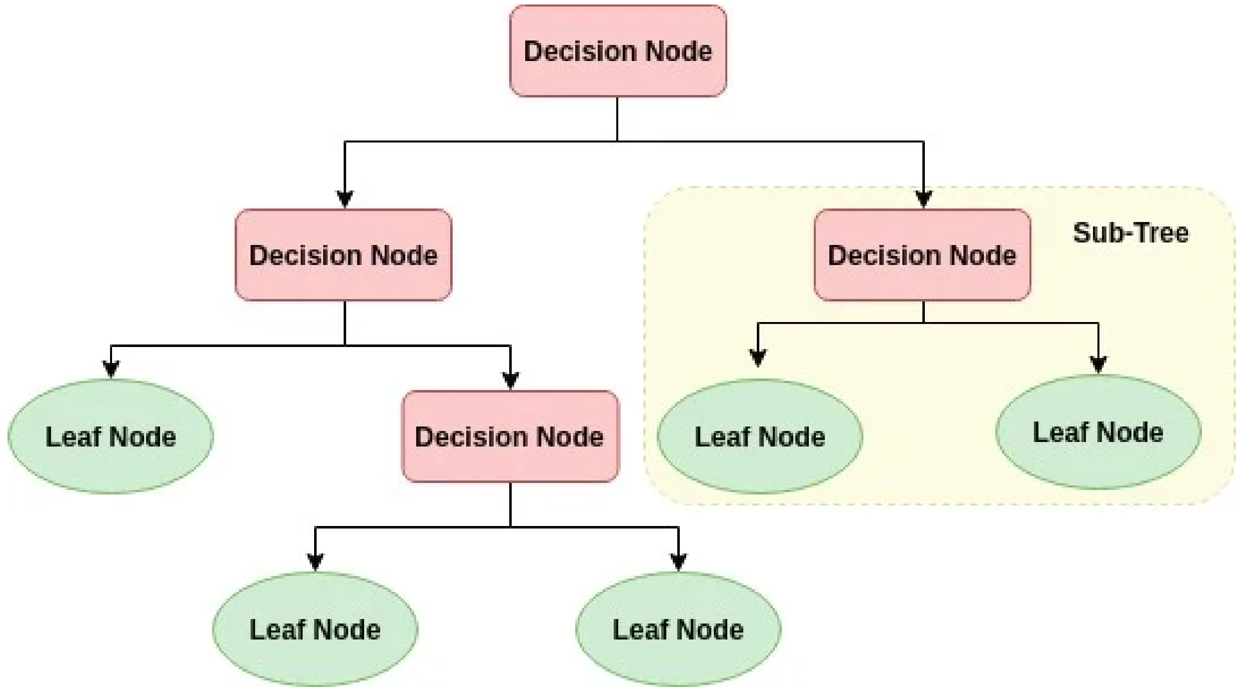
- Advantages:
  - Easy to implement
  - Simple and understandable
  - Completely data-driven
- Disadvantages:
  - Slow in testing
  - Sensitiveness to noisy or irrelevant data
  - Sensitive to K

Lab 2: KNN

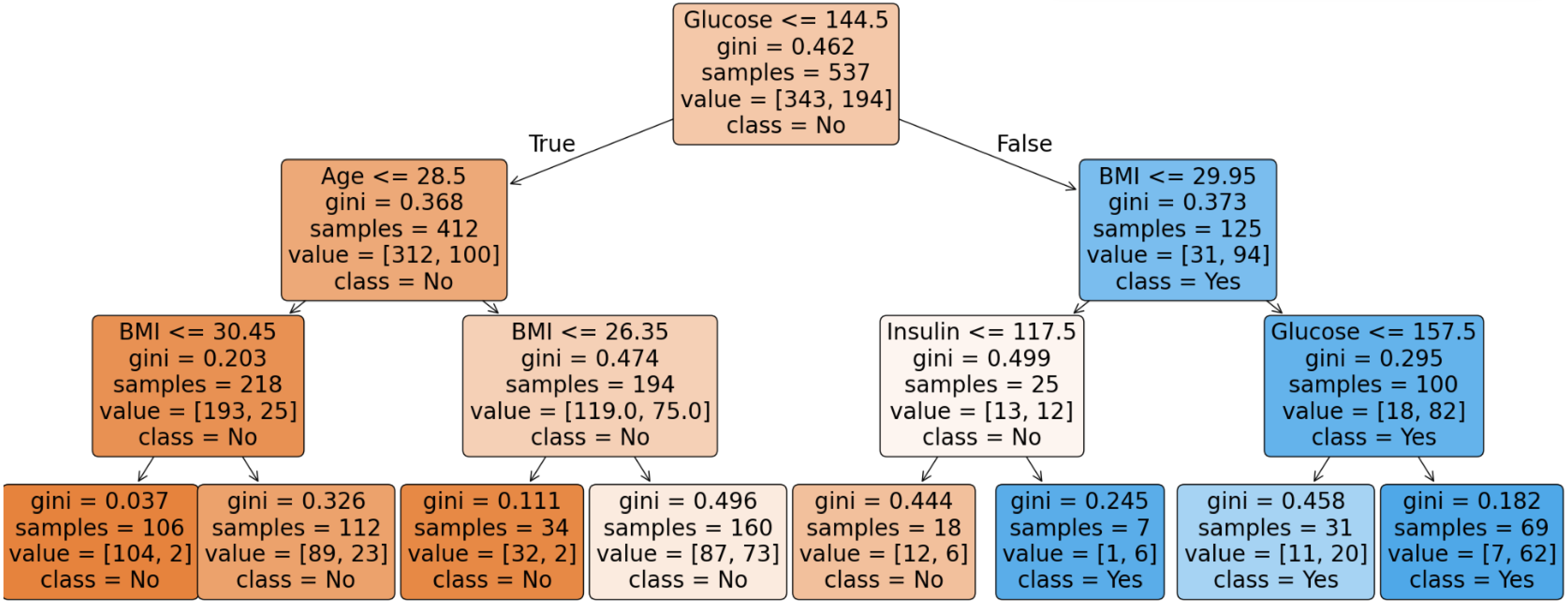
Decision Tree



- Intuitive and easy to interpret
- Require very little data preparation
- Easily deployed in rule-based system
- Build-in variable selection



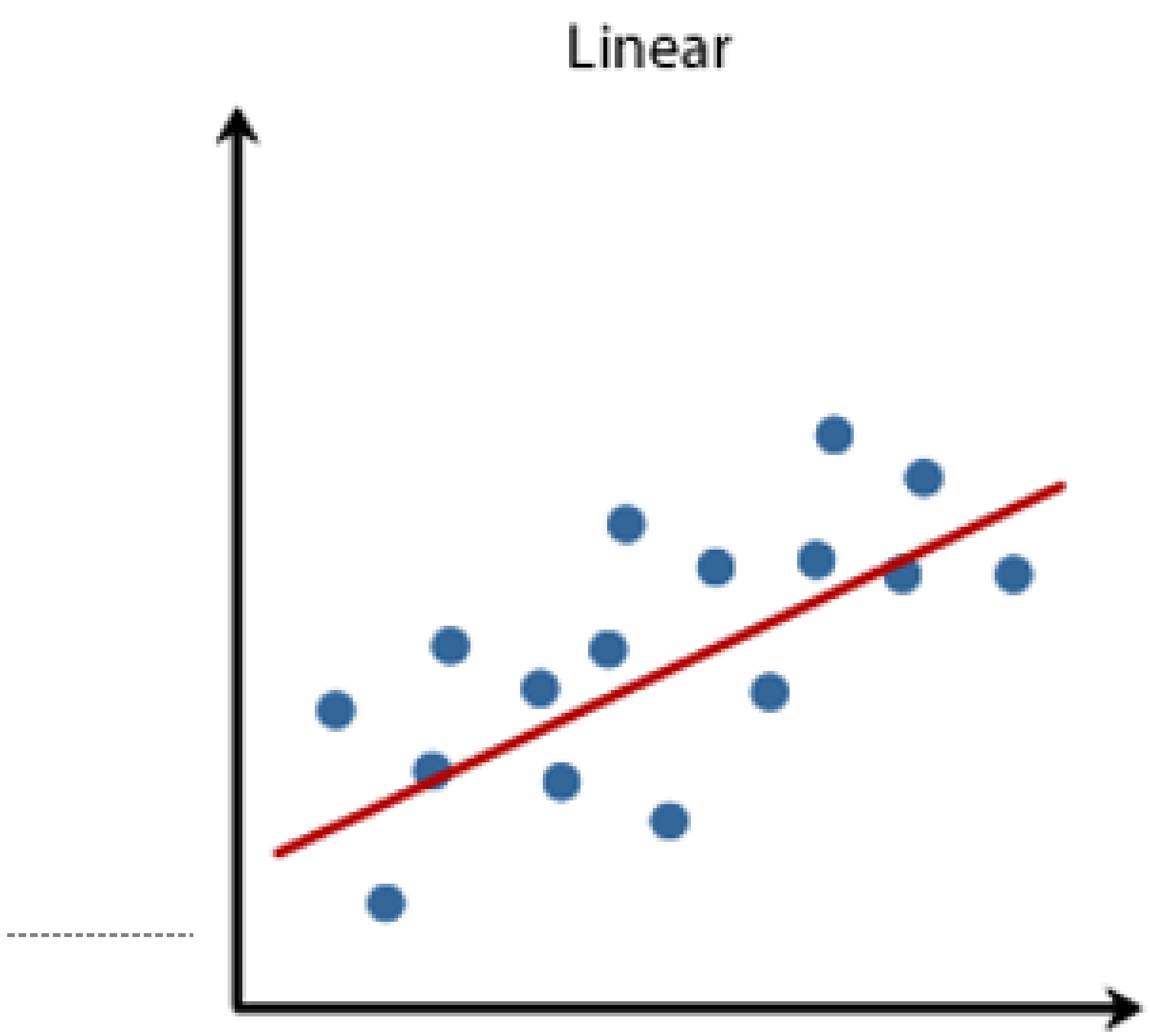
Source:
https://bigwhalelearning.files.wordpress.com/2014/11/titanic_heuristic.png
https://www.datacamp.com/community/tutorials/decision-tree-classification-python


Glucose <= 144.5
gini = 0.462
samples = 537
value = [343, 194]
class = No
True
False
Age <= 28.5
gini = 0.368
samples = 412
value = [312, 100]
class = No
BMI <= 29.95
gini = 0.373
samples = 125
value = [31, 94]
class = Yes
BMI <= 30.45
gini = 0.203
samples = 218
value = [193, 25]
class = No
BMI <= 26.35
gini = 0.474
samples = 194
value = [119.0, 75.0]
class = No
Insulin <= 117.5
gini = 0.499
samples = 25
value = [13, 12]
class = No
Glucose <= 157.5
gini = 0.295
samples = 100
value = [18, 82]
class = Yes
gini = 0.037
samples = 106
value = [104, 2]
class = No
gini = 0.326
samples = 112
value = [89, 23]
class = No
gini = 0.111
samples = 34
value = [32, 2]
class = No
gini = 0.496
samples = 160
value = [87, 73]
class = No
gini = 0.444
samples = 18
value = [12, 6]
class = No
gini = 0.245
samples = 7
value = [1, 6]
class = Yes
gini = 0.458
samples = 31
value = [11, 20]
class = Yes
gini = 0.182
samples = 69
value = [7, 62]
class = Yes

Lab 3: Decision Tree

Day 3: Machine Learning I
(Linear and Logistic Regression, Neural Networks)

Linear Regression



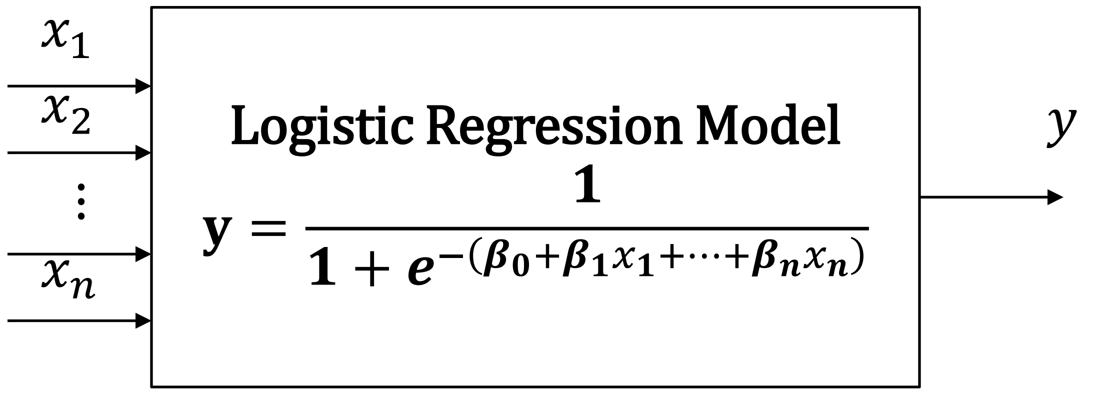
Simple Linear Regression	$y = b_0 + b_1 x_1$
Multiple Linear Regression	$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n$

Lab 4: Linear Regression

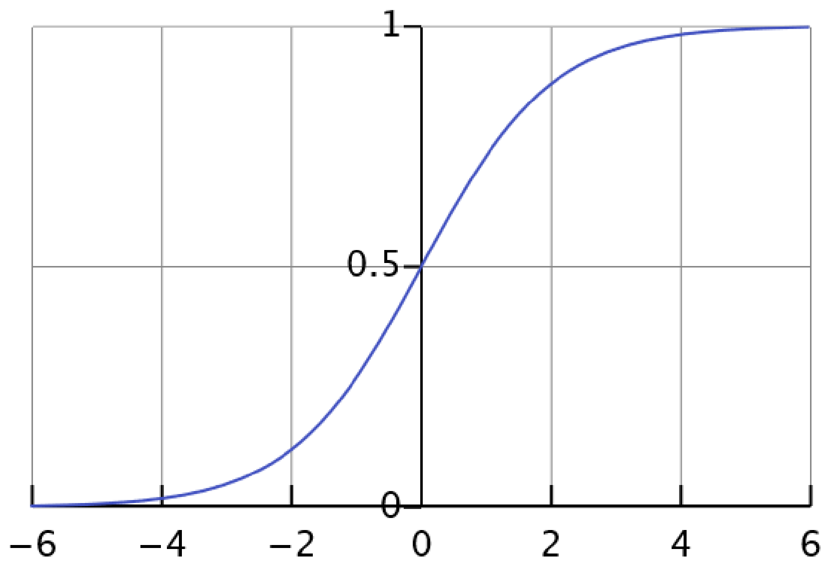
Logistic Regression

- Logistic regression model is represented in terms of logistic function as

$$\mathbf{y} = \frac{\mathbf{1}}{\mathbf{1} + e^{-(\beta_0+\beta_1 x_1+\cdots+\beta_n x_n)}}$$



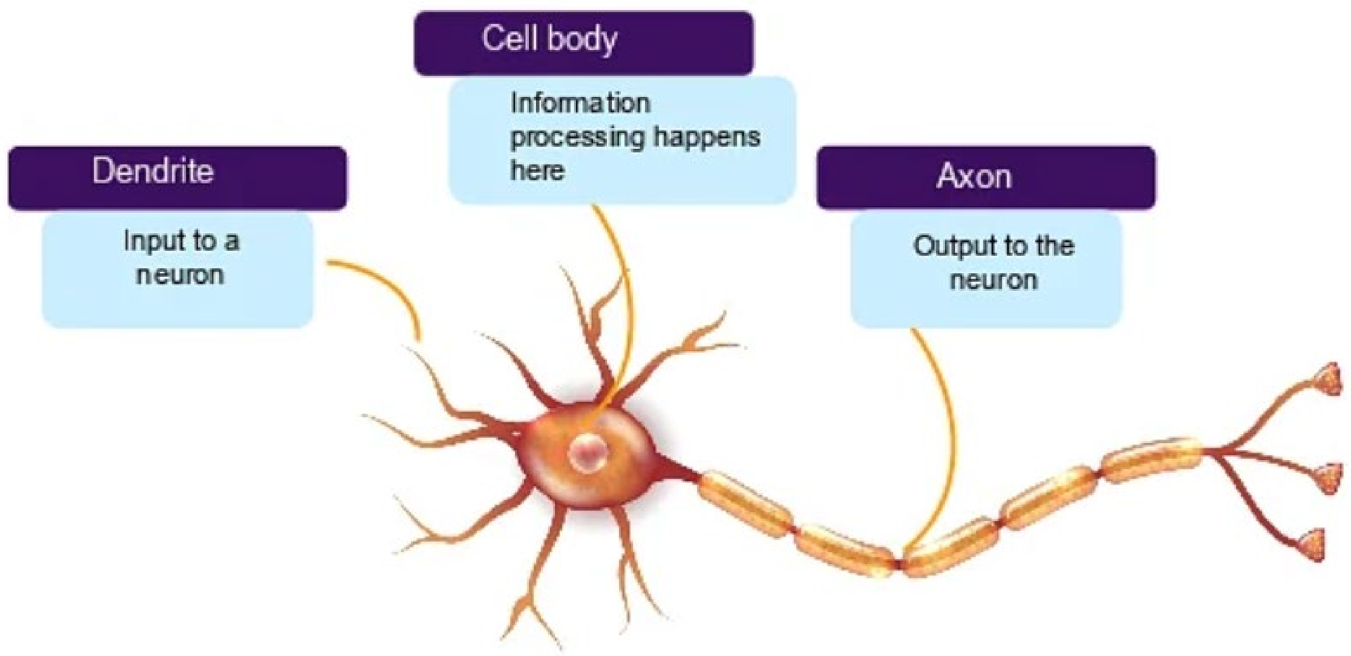
- Where x_i is input to the model and β_i is coefficients estimated by the model
- The probability formula is the function known as logistic curve
- Decision boundary is often set at 0.5



Lab 5: Logistic Regression

Neural Networks: A Single Neuron

Deep Learning: a vision approach, Andrew Glassner, 2021



https://www.simplilearn.com/tutorials/deep-learning-tutorial/what-is-neural-network

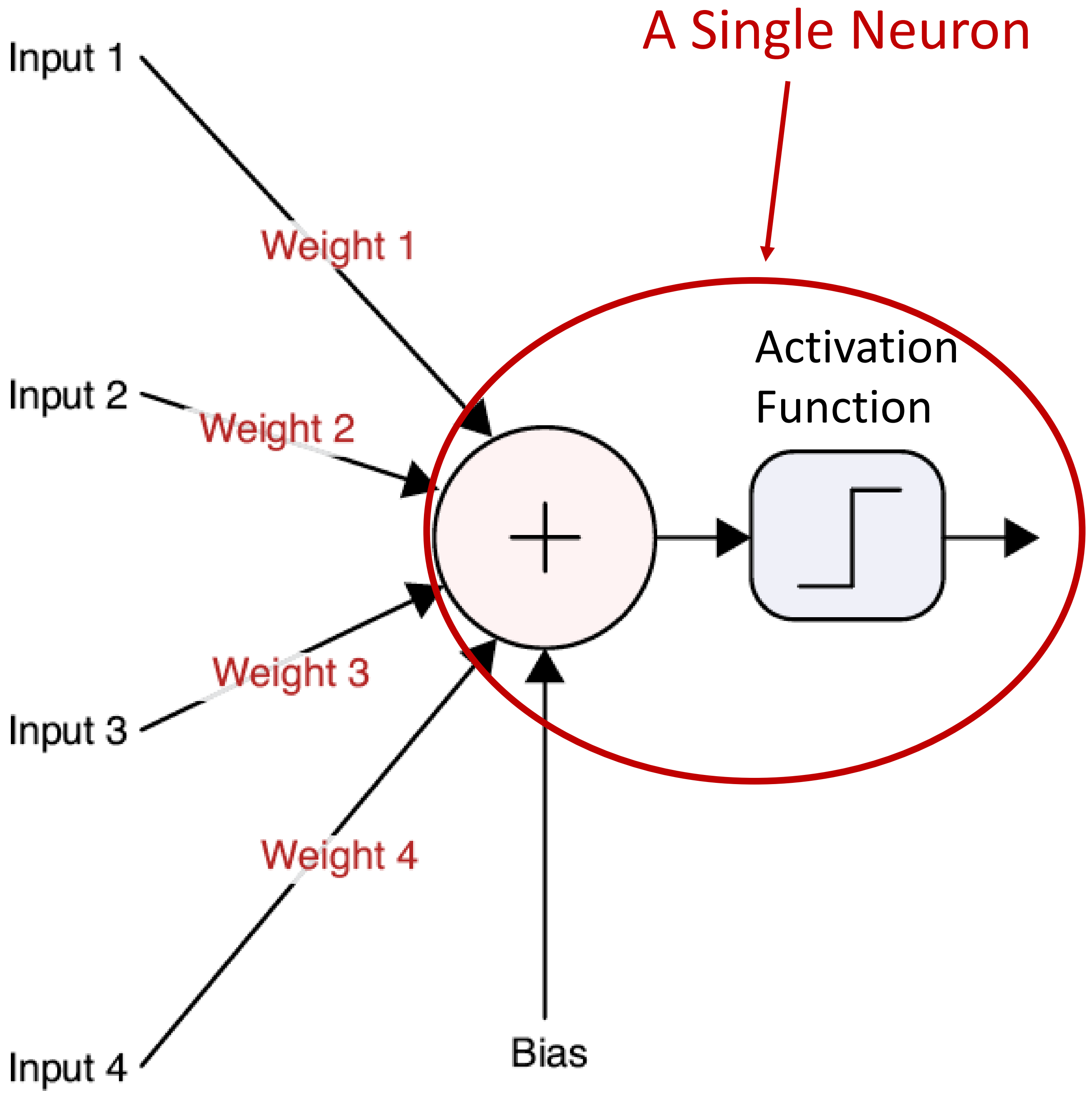


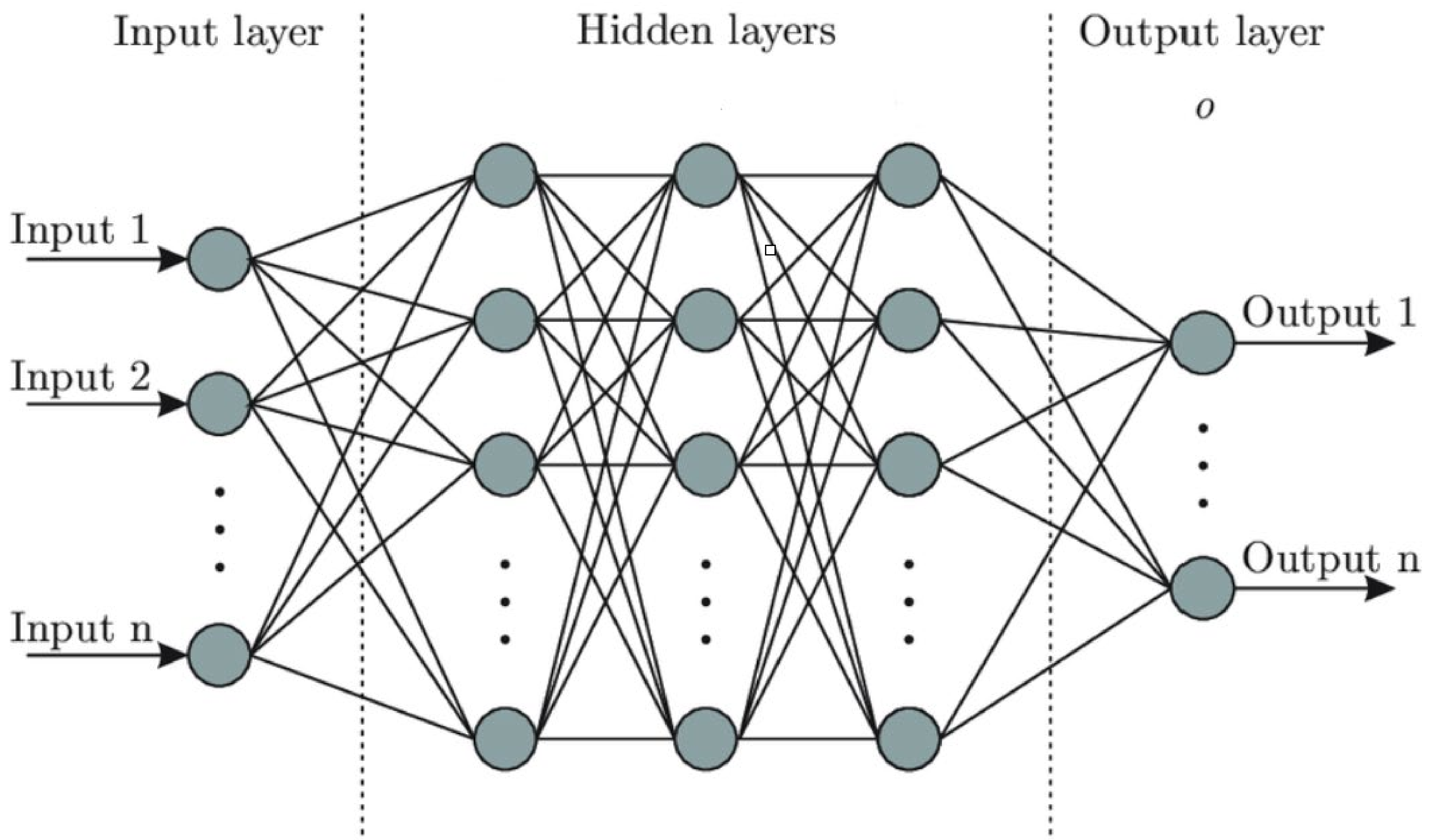
Figure 13-4: A neuron is often drawn with the weights on the arrows.

Neural Networks: A Simple Neural Network

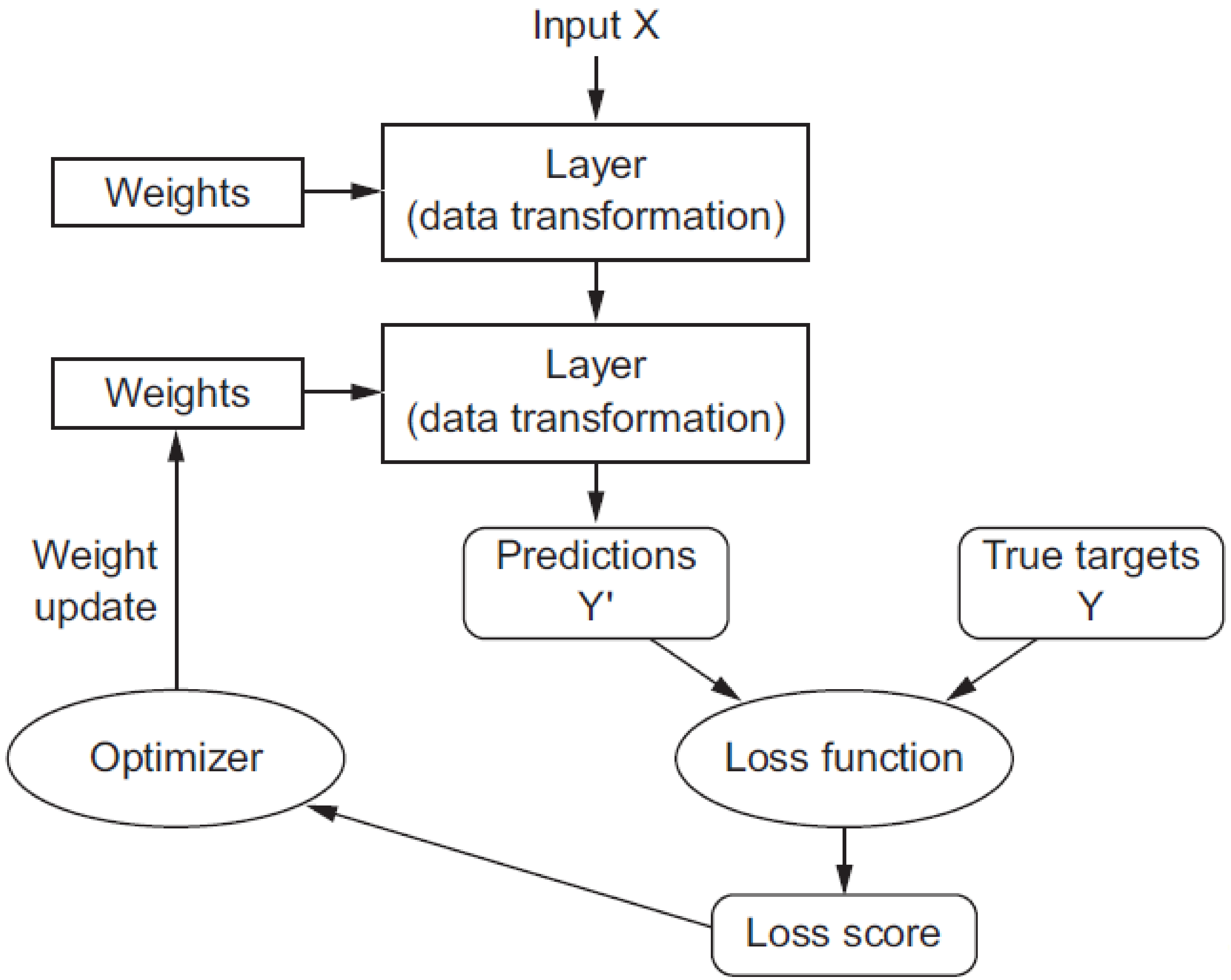
Input Features **Target**

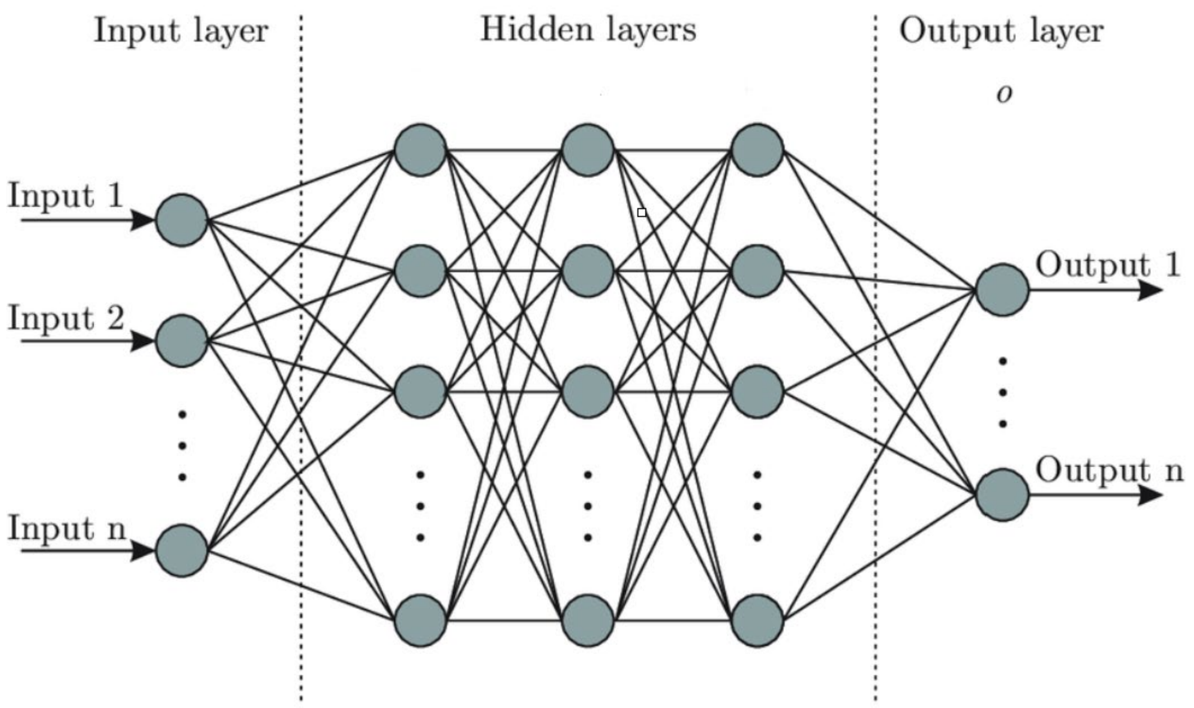
	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1
...	...	...	...	...	...	...	...	...	...
763	10	101	76	48	180	32.9	0.171	63	0
764	2	122	70	27	0	36.8	0.340	27	0
765	5	121	72	23	112	26.2	0.245	30	0
766	1	126	60	0	0	30.1	0.349	47	1
767	1	93	70	31	0	30.4	0.315	23	0

768 rows × 9 columns



How to train a Simple Neural Network

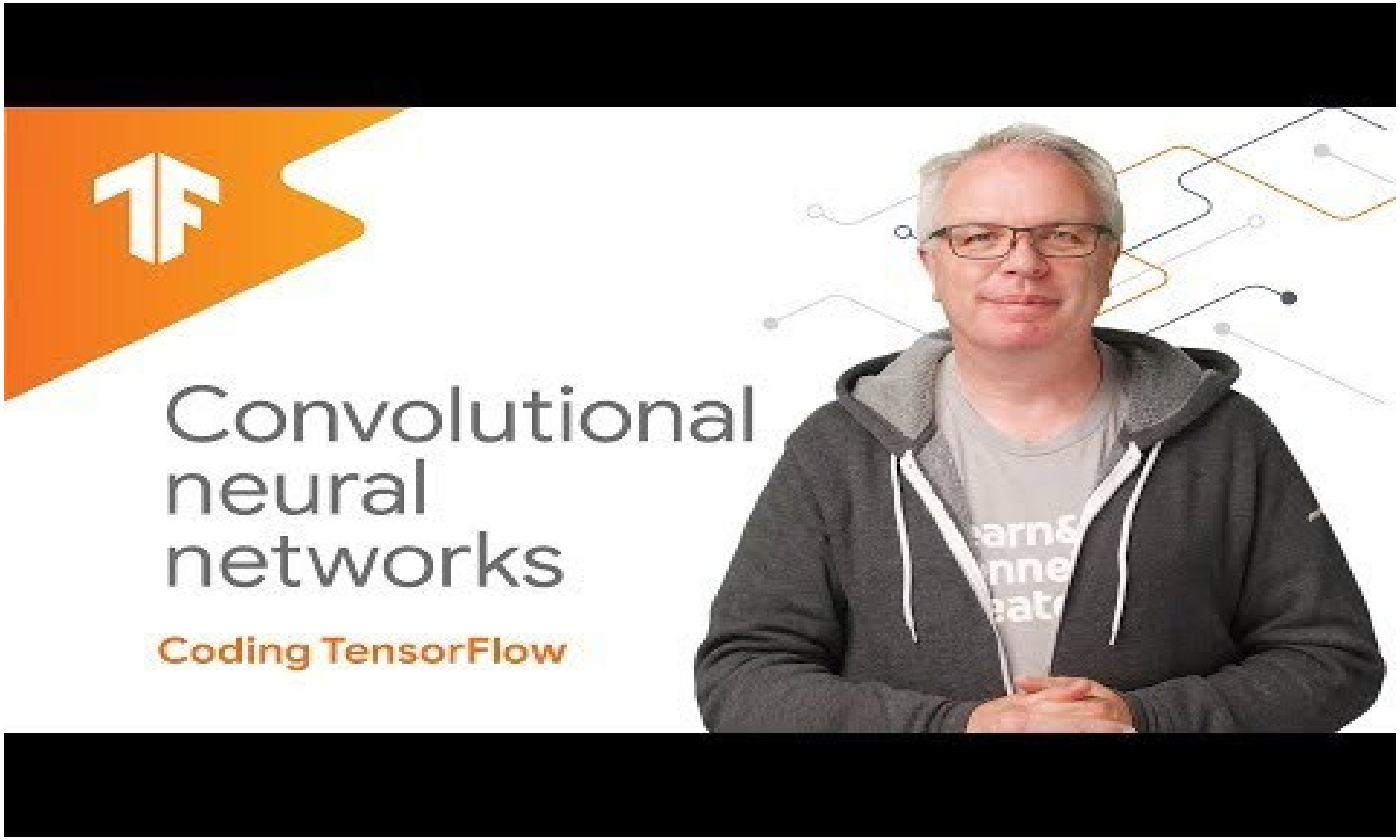




Lab 6: Neural Networks

Day 4: Deep Learning

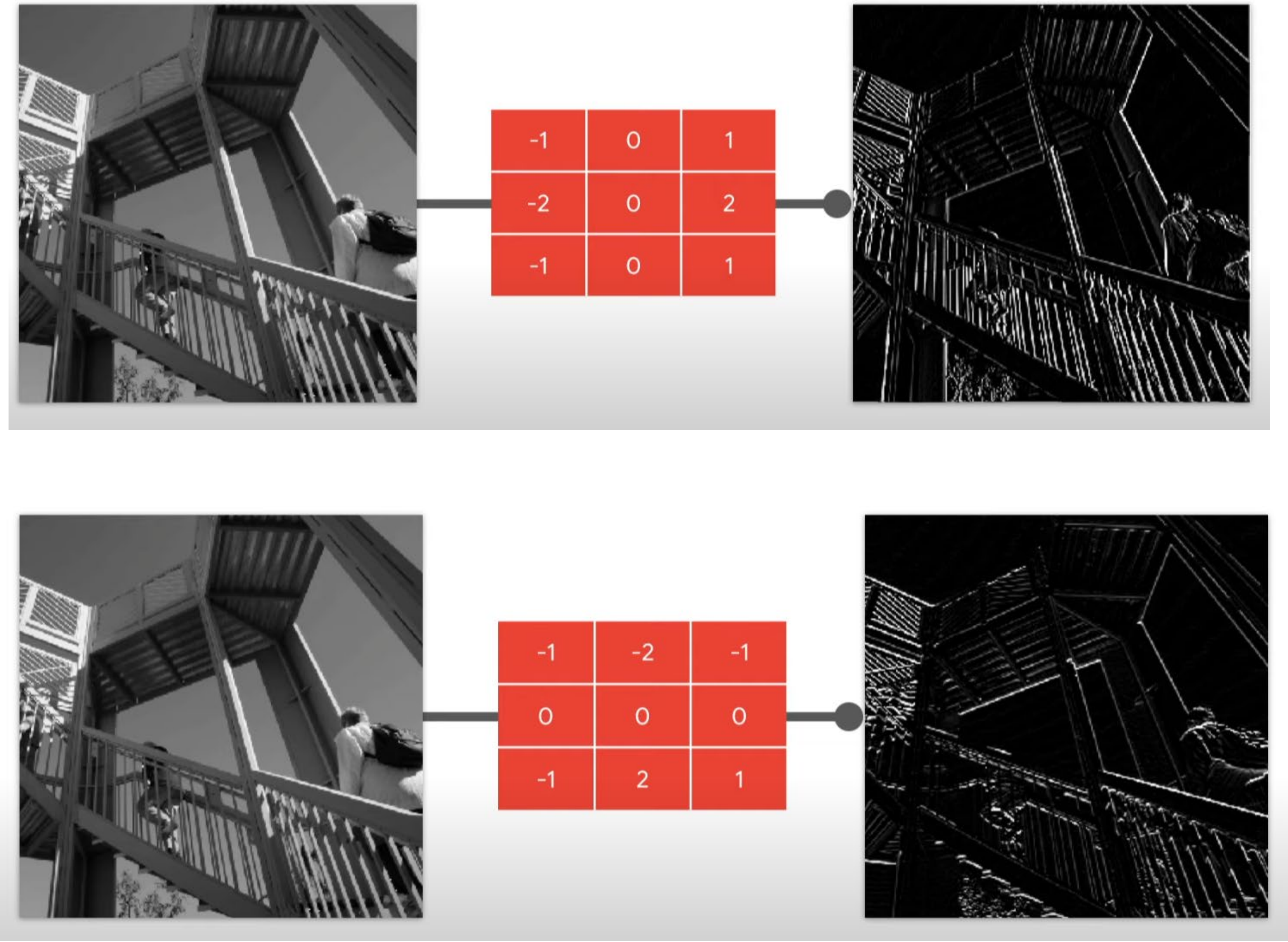
Introduction to convnets



https://youtu.be/x_VrgWTKkiM

Introduction to convnets

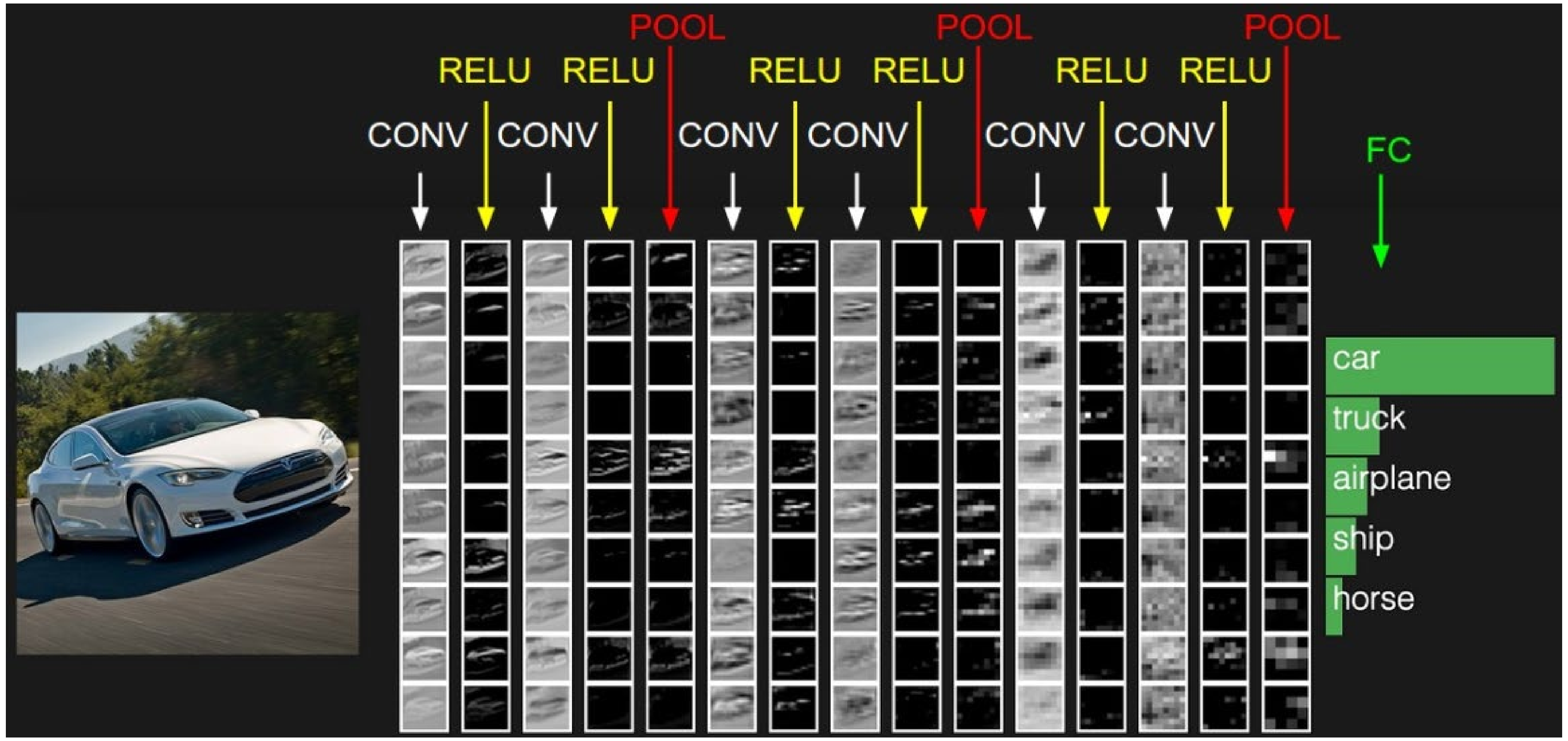
Conv2D Layer (through filters)



MaxPooling

source: https://youtu.be/x_VrgWTKkiM

An Example



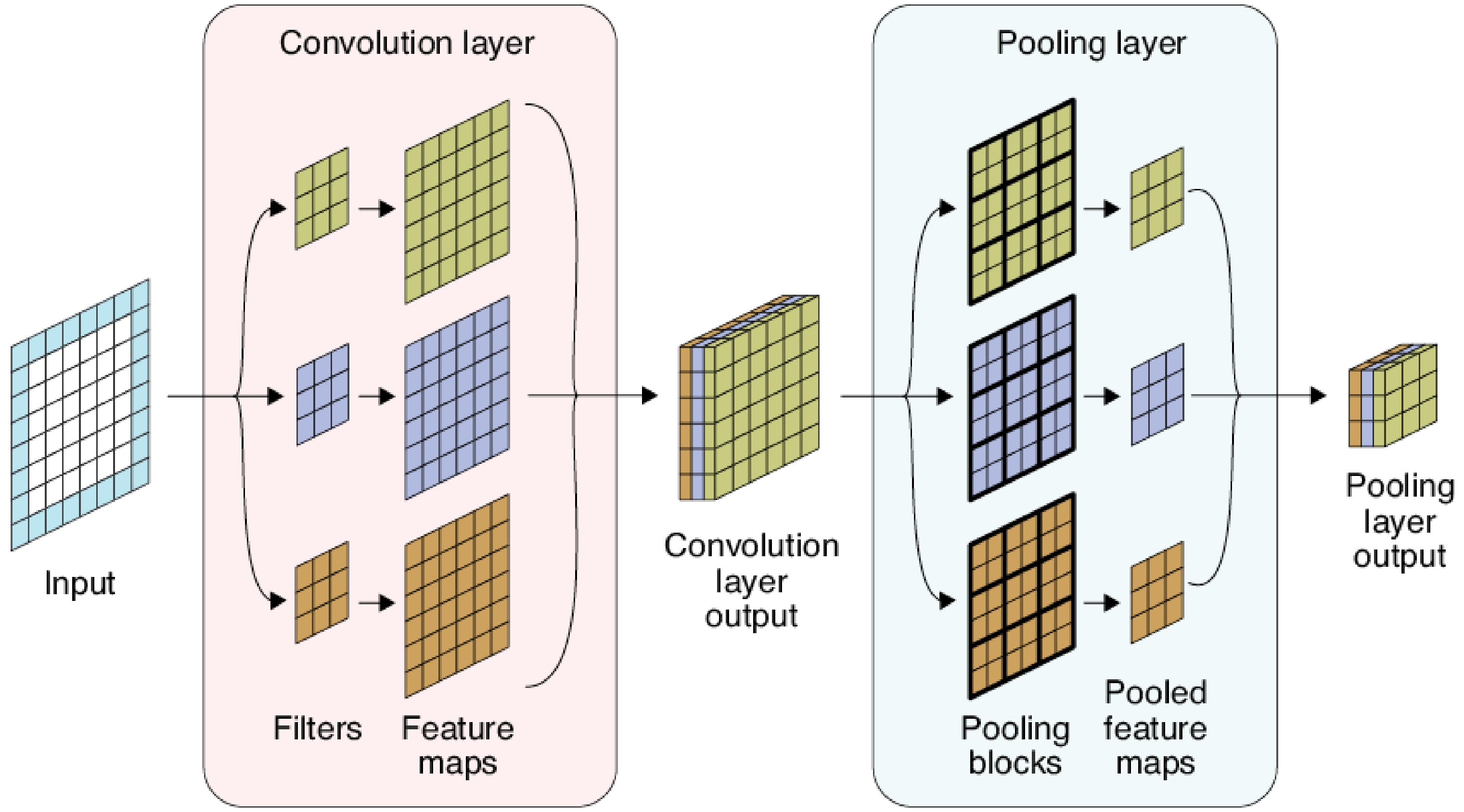
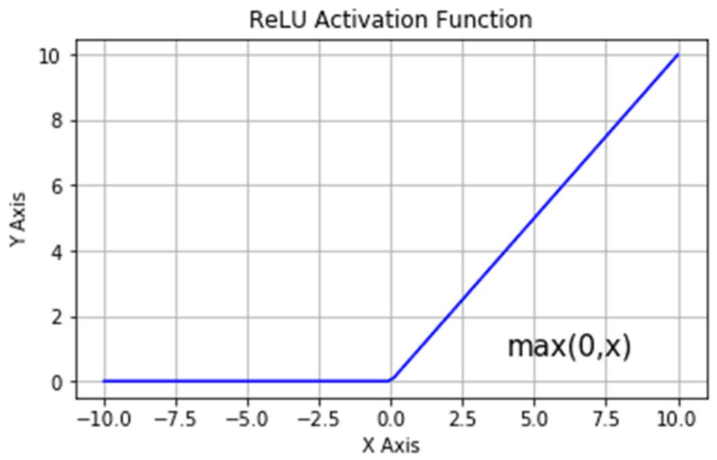


Figure 16-28: Pooling, or downsampling, with multiple filters

Deep Learning: a vision approach, Andrew Glassner, 2021

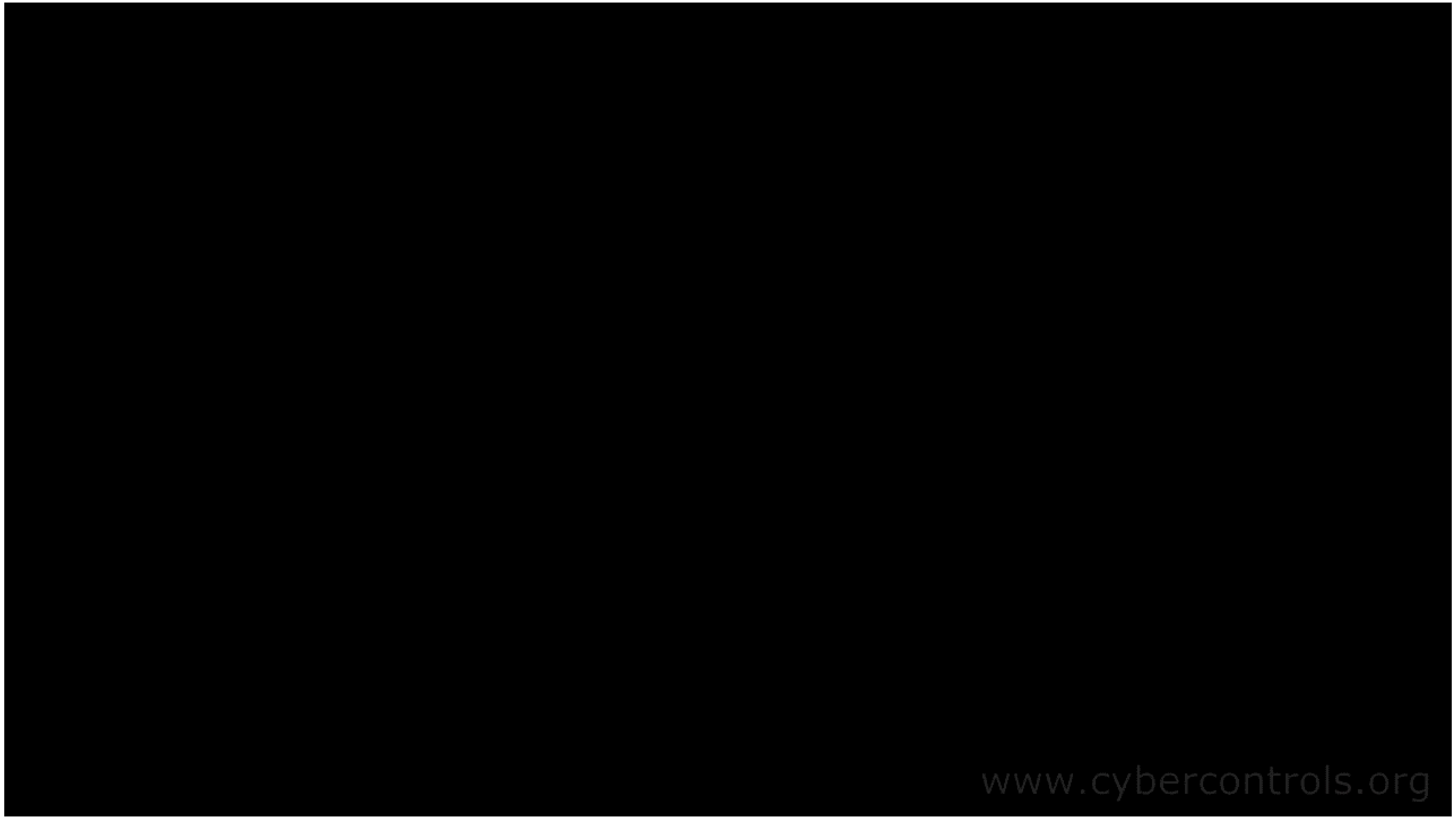
source: http://cs231n.github.io/convolutional-networks/

http://cs231n.stanford.edu/



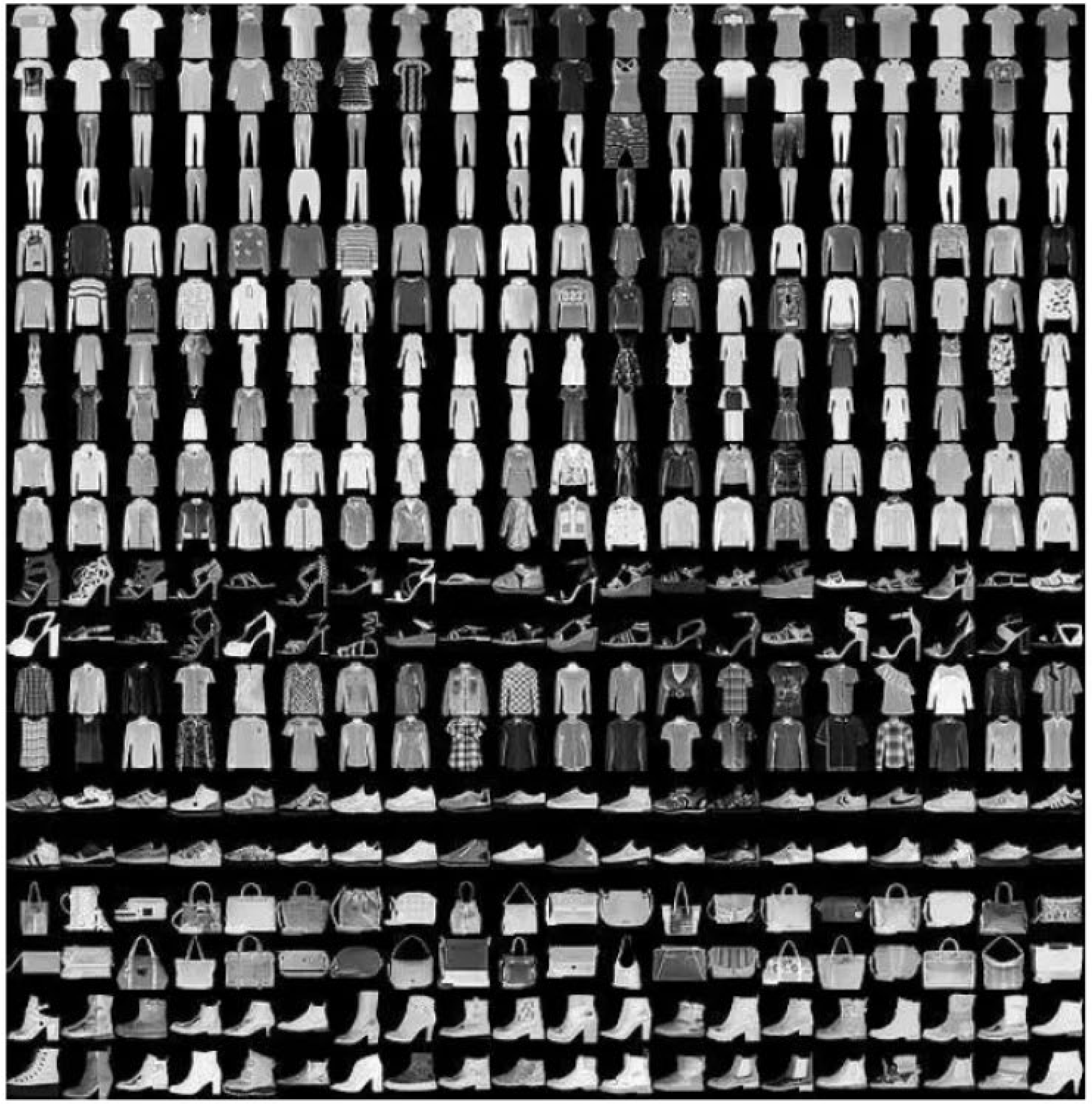
https://www.nomidl.com/deep-learning/what-is-relu-and-sigmoid-activation-function/

A VISUALIZATION …

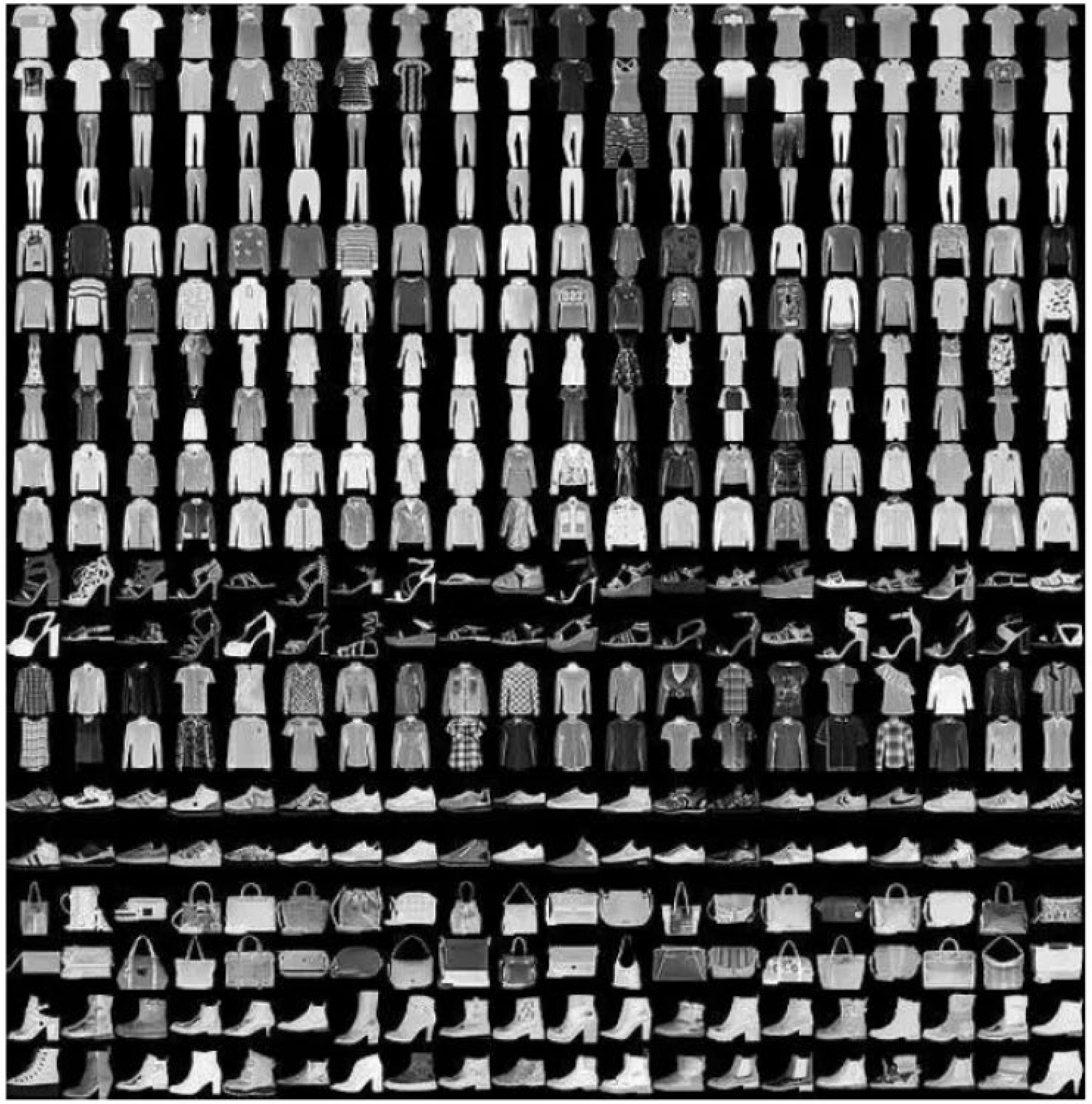


WWW.YOUTUBE.COM , S DMITRIEV

Lab 7: Fashion MNIST

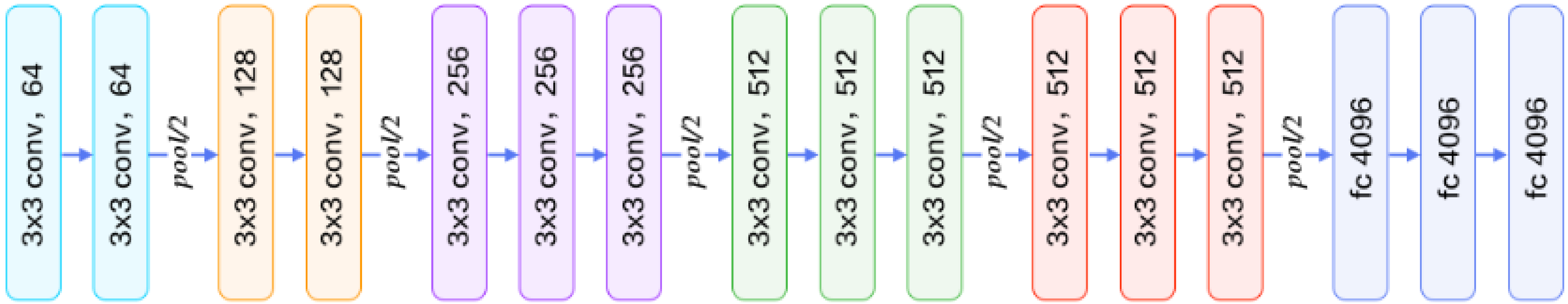
Label	Description	Examples
0	T-Shirt/Top	
1	Trouser	
2	Pullover	
3	Dress	
4	Coat	
5	Sandals	
6	Shirt	
7	Sneaker	
8	Bag	
9	Ankle boots	

Lab 8: Fashion MNIST using CNN

Label	Description	Examples
0	T-Shirt/Top	
1	Trouser	
2	Pullover	
3	Dress	
4	Coat	
5	Sandals	
6	Shirt	
7	Sneaker	
8	Bag	
9	Ankle boots	

VGG 16 and Transfer Learning

- 16 Layers: 13 conv + 3 fc
- Trained on Image-Net dataset
  - More than 14 million images and 1000 classes (e.g. animals, plants, vehicles, household items, and natural scenes)



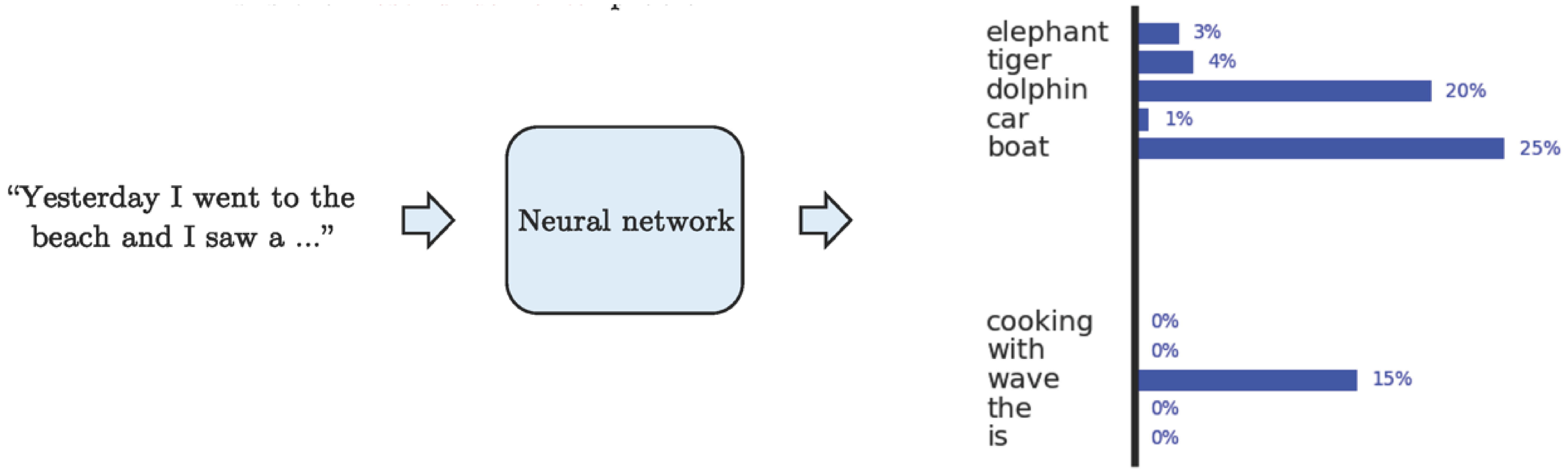
FOOD IMAGES CLASSIFIER

Language Model

Input: Sequence of Words

Output: probability distribution

Over the dictionary / vocabulary



Transformer: Attention is all you need!

The pink elephant tried to get into the car but it was too ______

big

The evolution of Transformer Model

Transformer (2017)
- Surpassed CNN & RNN on NLP task

GPT (2018)
- Pre-trained to Predict the next word
- Fine-tuning using specific task data

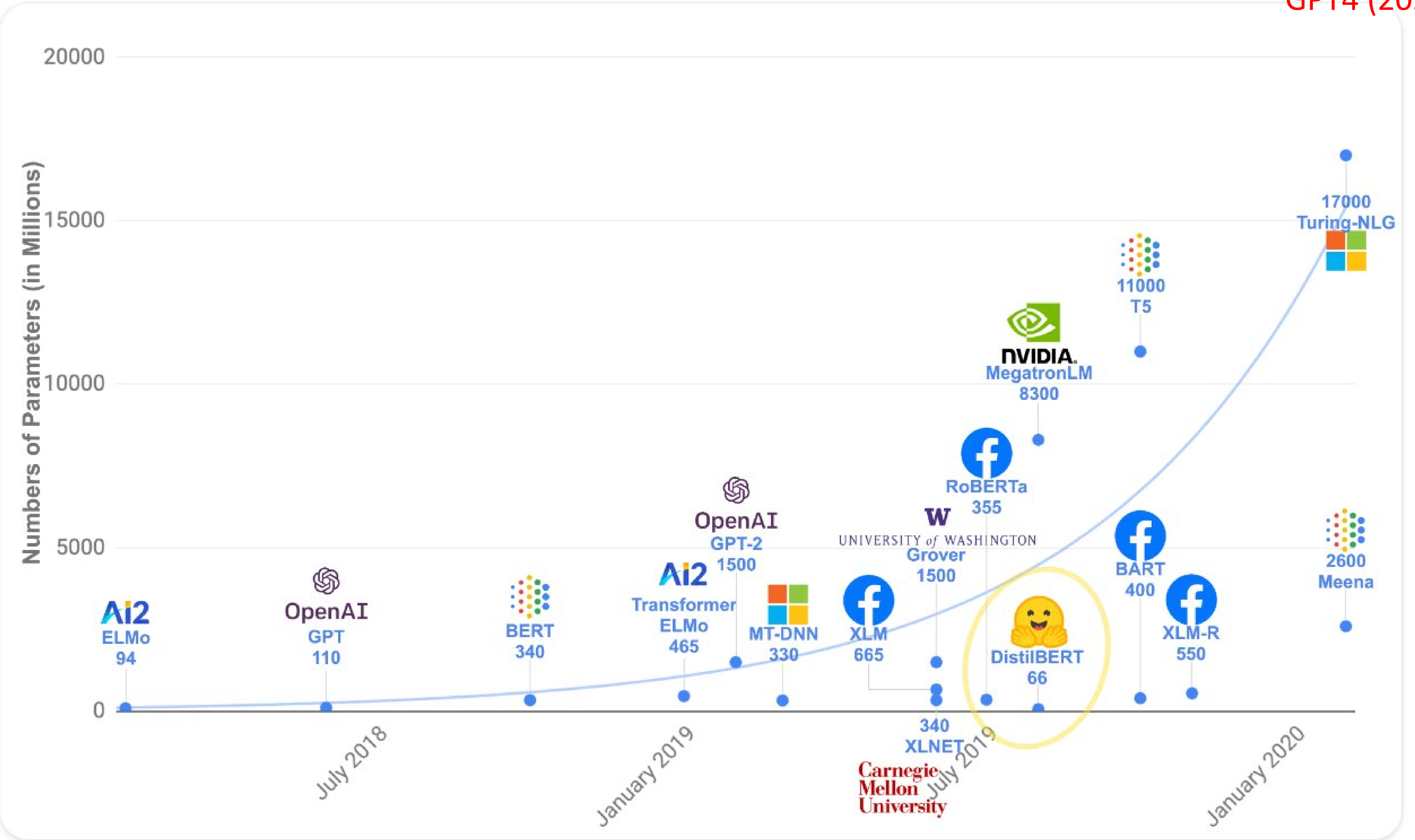
BERT (2019)
- Pre-trained to Encode the input sentence
- Fine-tuning for downstream tasks

GPT-2 (2019) & GPT-3 (2020)
- Do not require fine-tuning
- Instruction in a Prompt ➔ Detailed Response

ChatGPT (2022) & GPT-4 (2023)
- Pre-trained to predict the next word
- Fine-tuned using human feedbacks
- improvements in truthfulness and reductions in toxic outputs

Large Language Models

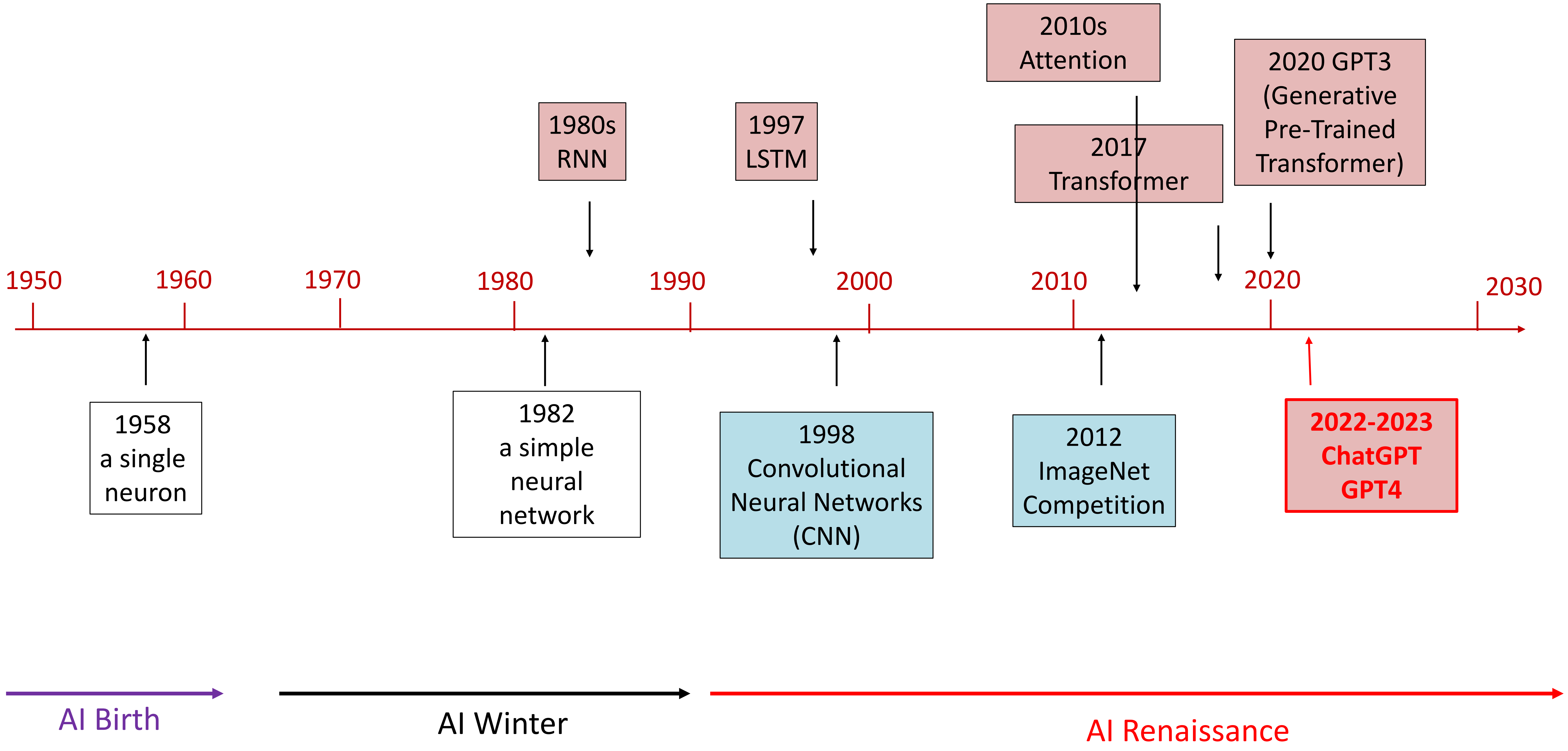
GPT3 (2020 & 175B parameters)
ChatGPT / GPT3.5 (2022)
GPT4 (2023)



https://huggingface.co/learn/nlp-course/chapter1/4?fw=pt

Lab 9: Transformer in NLP

A Brief History of Neural Networks



Project Presentation

Reference

Book

François Chollet, “Deep Learning with Python”, 2018

Bernard Marr, “Artificial Intelligence in Practice”, 2019

Stuart J. Russell and Peter Norvig, “Artificial Intelligence A Modern Approach”, 1995

Online Resource

https://www.youtube.com/watch?v=t4K6lney7Zw

https://en.wikipedia.org/wiki/Computer_chess

https://dougenterprises.com/artificial-intelligence/should-you-use-an-expert-system-instead-of-machine-learning/

https://www.saedsayad.com/decision_tree.htm

https://huggingface.co/