Large Language Model and Its Applications

LLM01: Introduction to Machine Learning

Duc-Trong Le

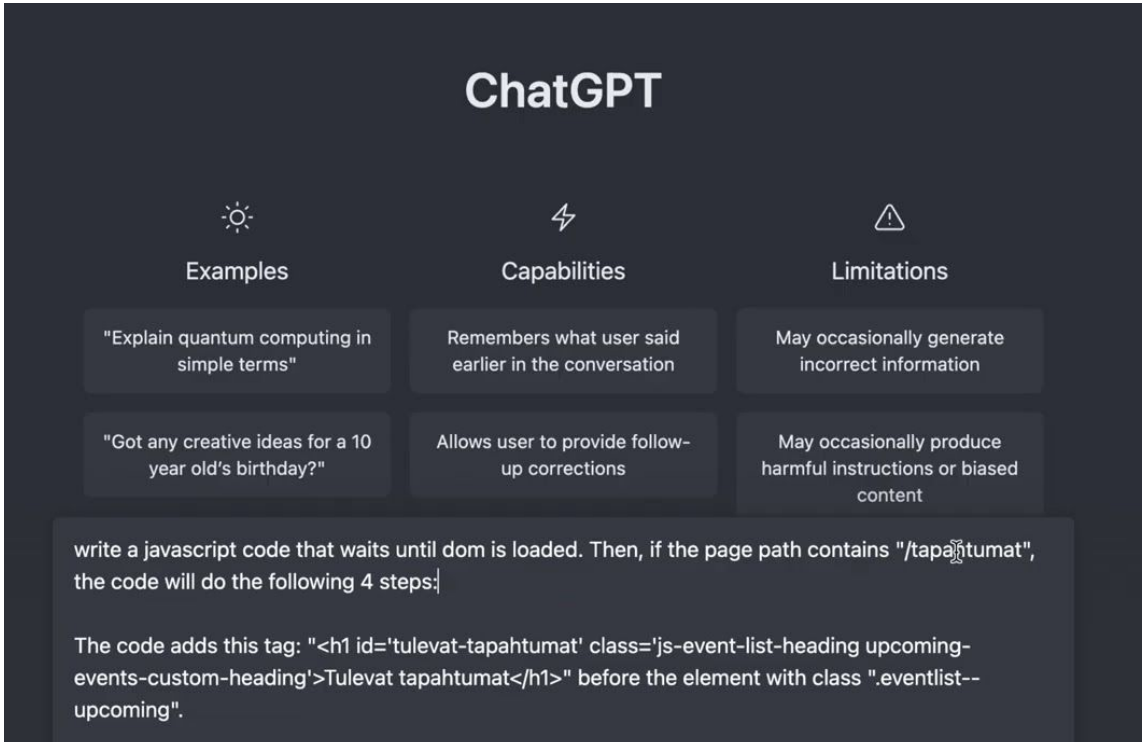
sponsored by **KEPCO KDN Co., Ltd.**
Eco-friendly & Digital Centered Energy ICT Platform Leader

Hanoi, 09/2023

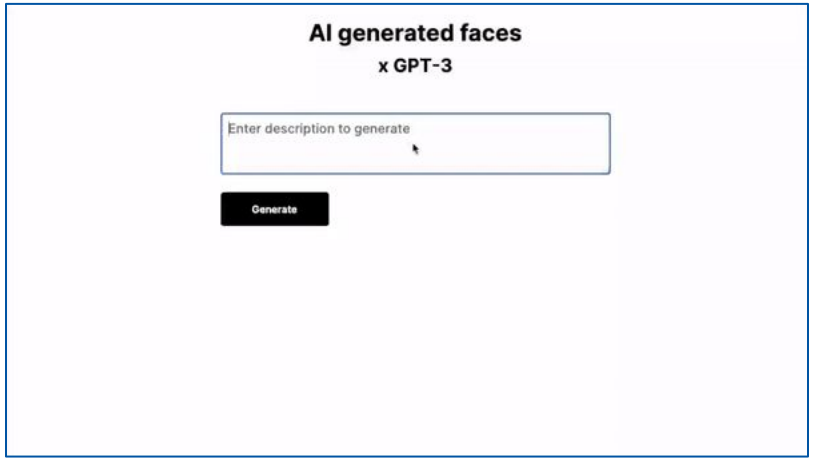
Outline

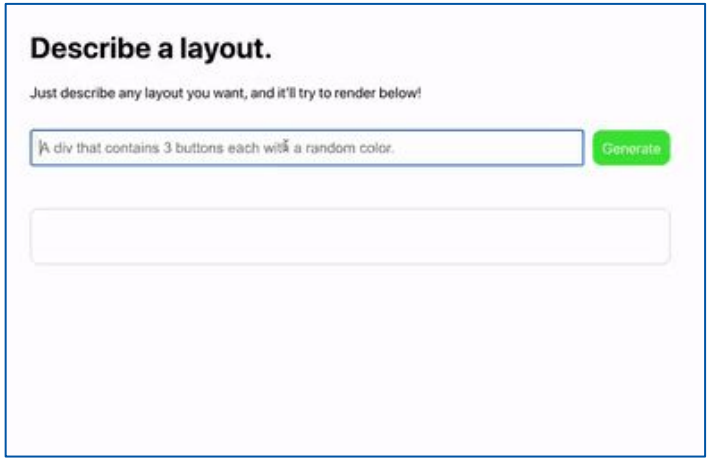
- Motivating Examples
- Introduction to Machine Learning
- Statistics vs Machine Learning
- Machine Learning vs Deep Learning
- Applications of Machine Learning
- Types of Machine Learning
  - Supervised Learning
  - Unsupervised Learning
  - Reinforcement Learning

ChatGPT


ChatGPT
Examples
Capabilities
Limitations
"Explain quantum computing in simple terms"
Remembers what user said earlier in the conversation
May occasionally generate incorrect information
"Got any creative ideas for a 10 year old's birthday?"
Allows user to provide follow-up corrections
May occasionally produce harmful instructions or biased content
write a javascript code that waits until dom is loaded. Then, if the page path contains "/tapahtumat", the code will do the following 4 steps:
The code adds this tag: "<h1 id='tulevat-tapahtumat' class='js-event-list-heading upcoming-events-custom-heading'>Tulevat tapahtumat</h1>" before the element with class ".eventlist--upcoming".

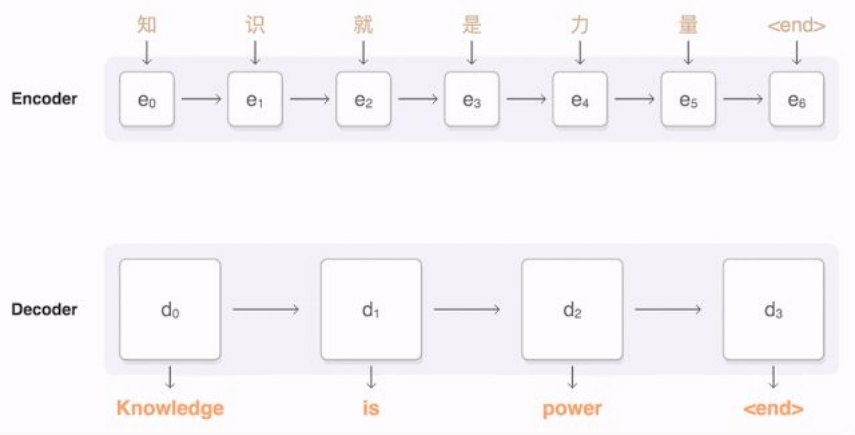
GPT-3 Applications





https://www.youtube.com/watch?v=_x9AwxfjxvE

Google Translate



https://ai.googleblog.com/2016/09/a-neural-network-for-machine.html

Voice Assistant

Hey Siri

VinFast Vivi

Automatic Driving - Tesla

See more: https://www.youtube.com/watch?v=tlThdr3O5Qo

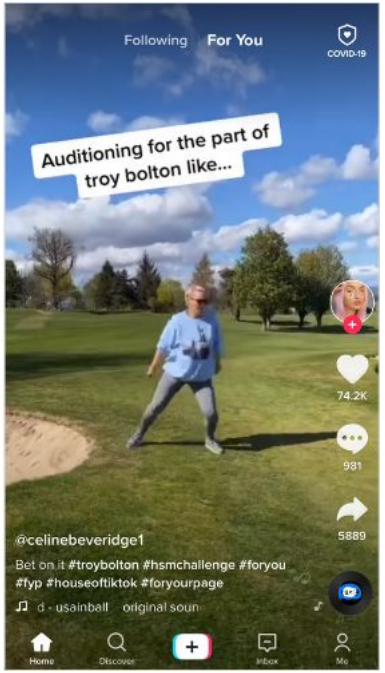
Stock Price Prediction

See more: https://patelrohan008.github.io/cx4240-stock-prediction/

Tiktok

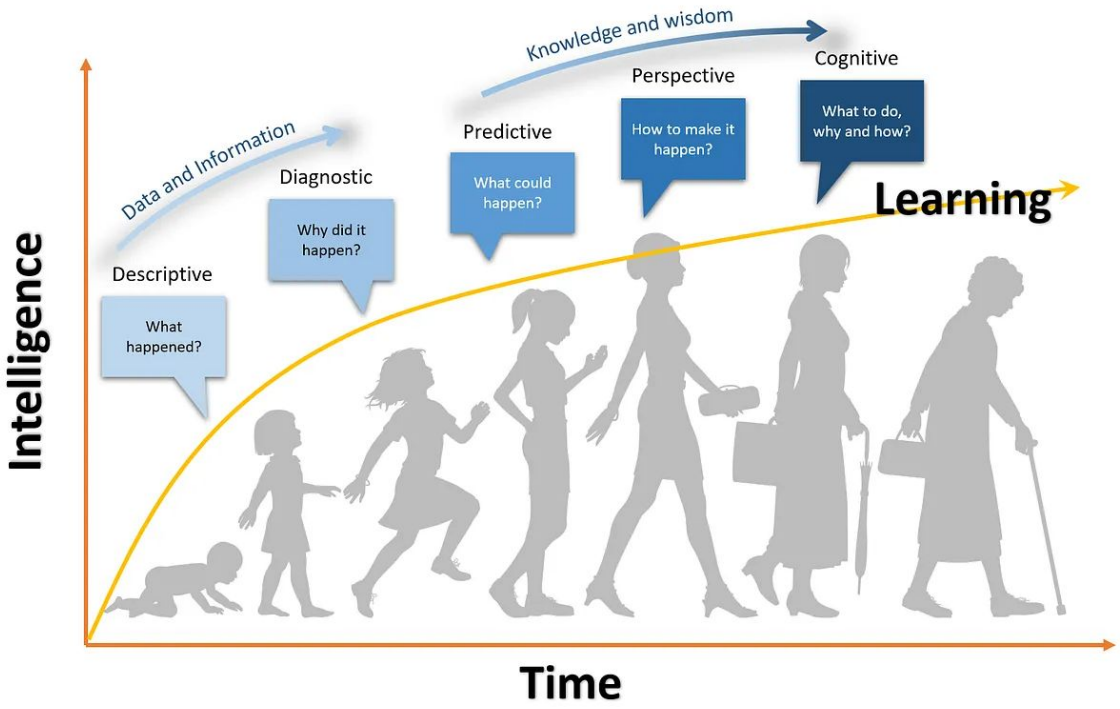

Following
For You
COVID-19
18.8K
128
284
@_sarahestherrr
So @misheyem delivered the chunkiest cookies to my doorstep & im gonna wake up with diabetes #fyp #foryou #foryoupage #cookies #nyccookie
rahestherrr original sound
Home
Discover
Inbox
Me


Following
For You
COVID-19
99.6K
756
5932
@kjp
Perfect morning for an adventure! #goldenretriever #natureathome #fyp #dog #adventure #beautiful #earth #nature #canoe #outdoors #outside #friendship
@Matt Hartke London's S
Home
Discover
Inbox
Me


Following
For You
COVID-19
Auditioning for the part of troy bolton like...
74.2K
981
5889
@celinebeveridge1
Bet on it #troybolton #hsmchallenge #foryou #fyp #houseoftiktok #foryourpage
d - usainbalt original soun
Home
Discover
Inbox
Me

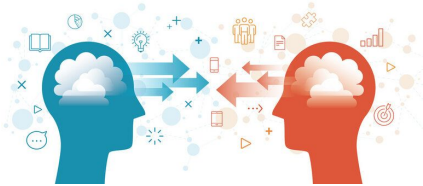
Contextual Advertisement

Human Learning



Source: https://medium.com/dartexon/ai-the-very-nature-of-human-learning-e5d93a34a54

Mean or Materials



What is Machine Learning?

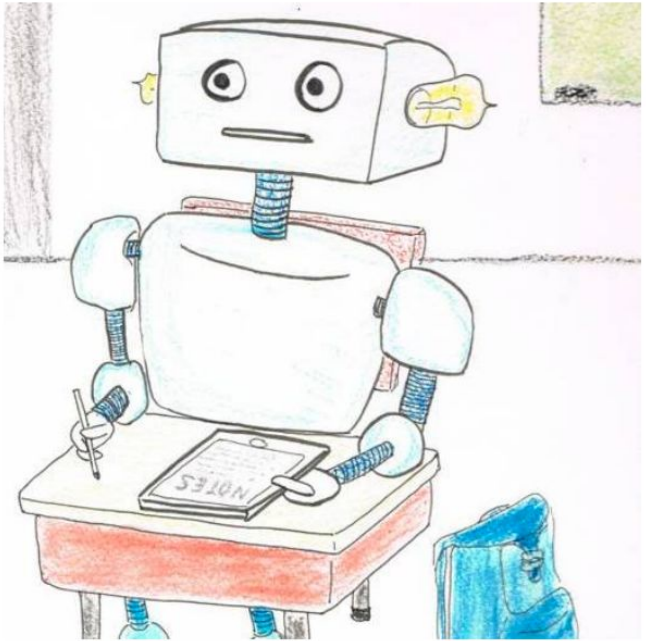


Image: www.gureckislab.org

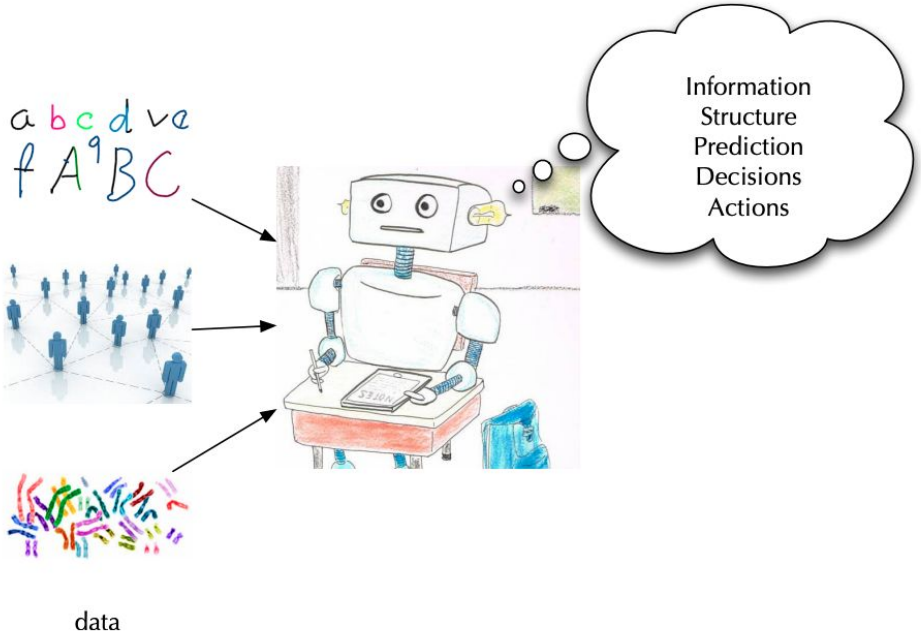
Formal Definition of Machine Learning (ML)

- **Arthur Samuel, 1959**: Field of study that gives computers the ability to **learn** without being explicitly programmed.
- Tom Mitchell, 1997: Any computer program that **improves its performance** at some task **through experience**.
- Kevin Murphy, 2012: To develop methods that can **automatically** detect **patterns in data**, and then to use the uncovered patterns to **predict** future data or other outcomes of interest.

Formal Definition of Machine Learning (ML)

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed.

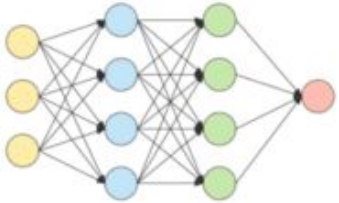
Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.



https://www.expertsystem.com/machine-learning-definition/

Human Learning vs Machine Learning

Intelligence → Models



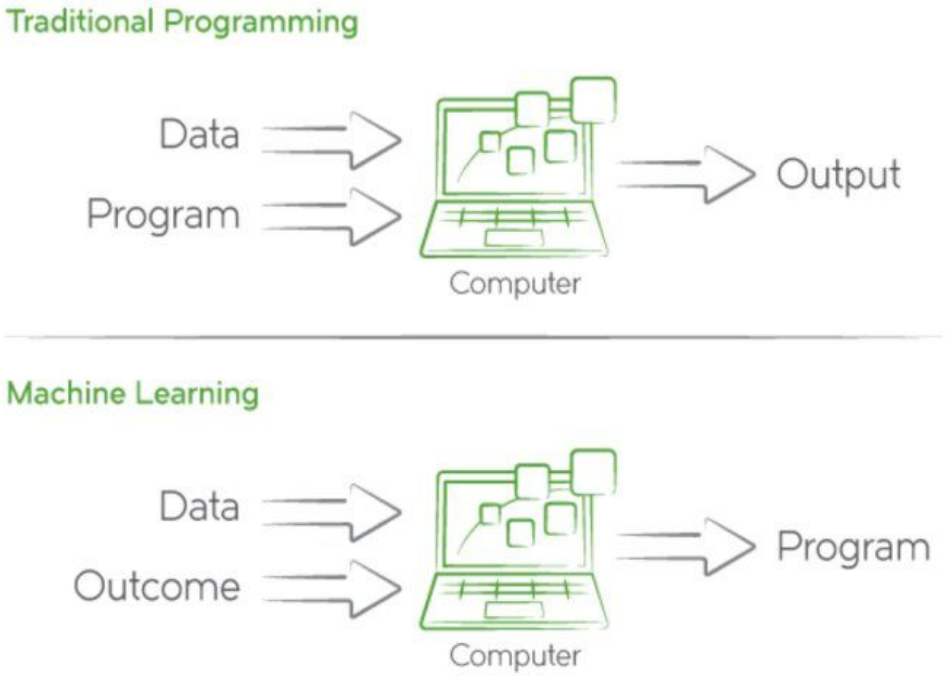
Learning materials → Data

Learning skills → Skillearn

- Learning by creating tests
- Interleaving learning
- Learning by ignoring
- ...

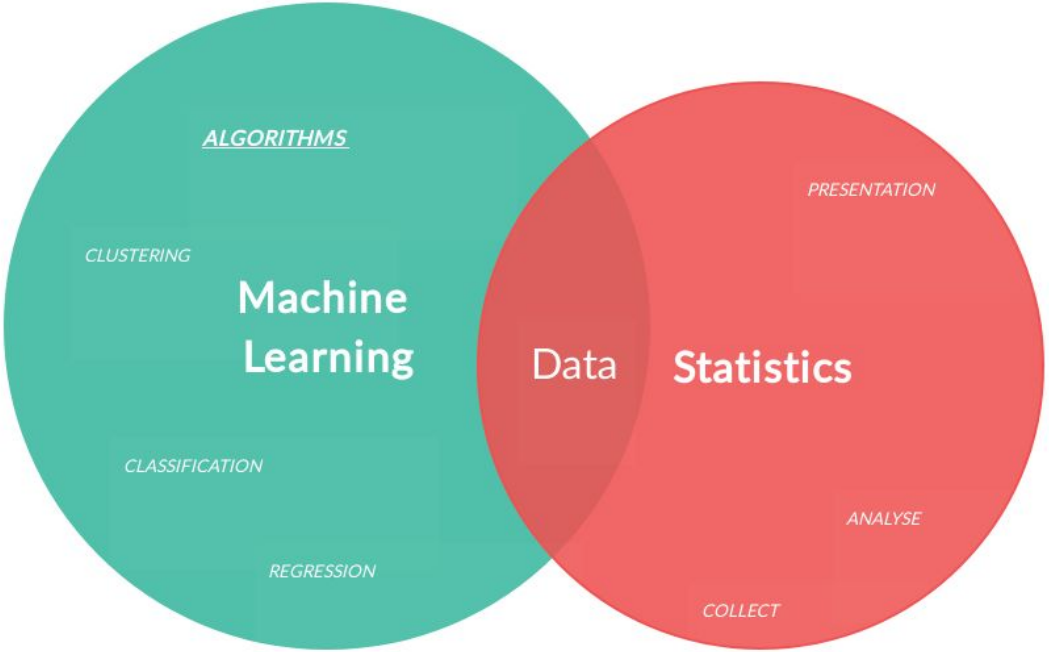
Read more: [Skillearn: Machine Learning Inspired by Humans' Learning Skills]

Traditional Programming vs Machine Learning?



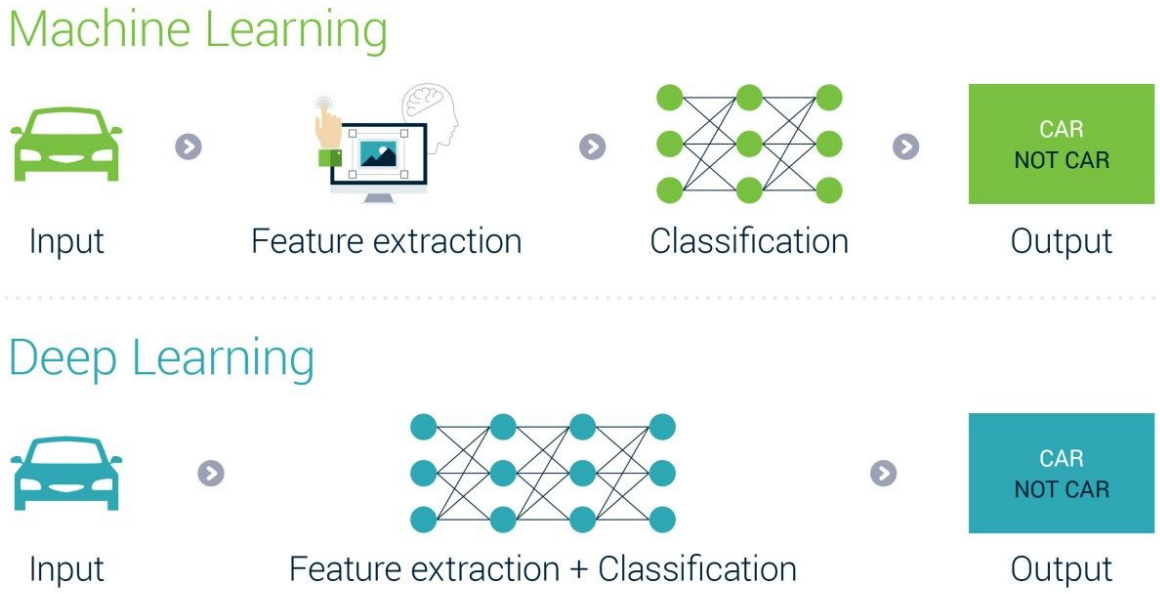
https://images.techhive.com/images/article/2017/05/traditional-programming-vs-machine-learning-100723299-large.jpg

Statistics vs Machine Learning?



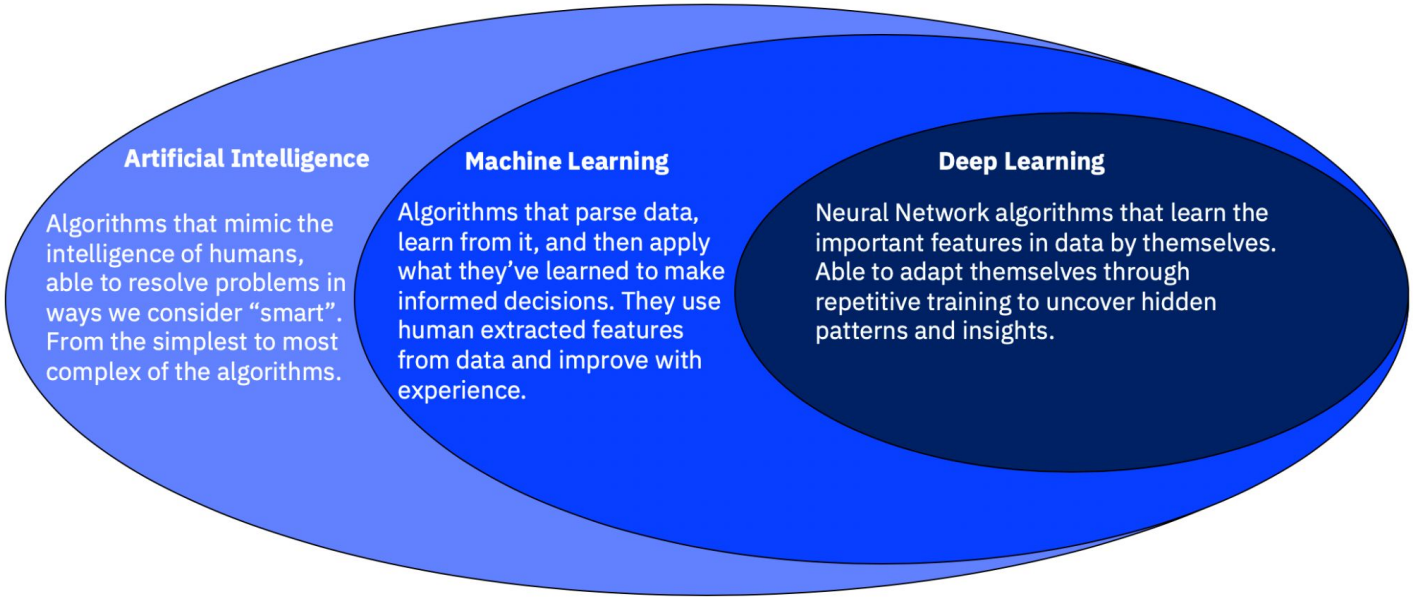
Source: https://pythonprogramminglanguage.com/what-is-the-difference-between-statistics-and-machine-learning/

Machine Learning vs. Deep Learning?



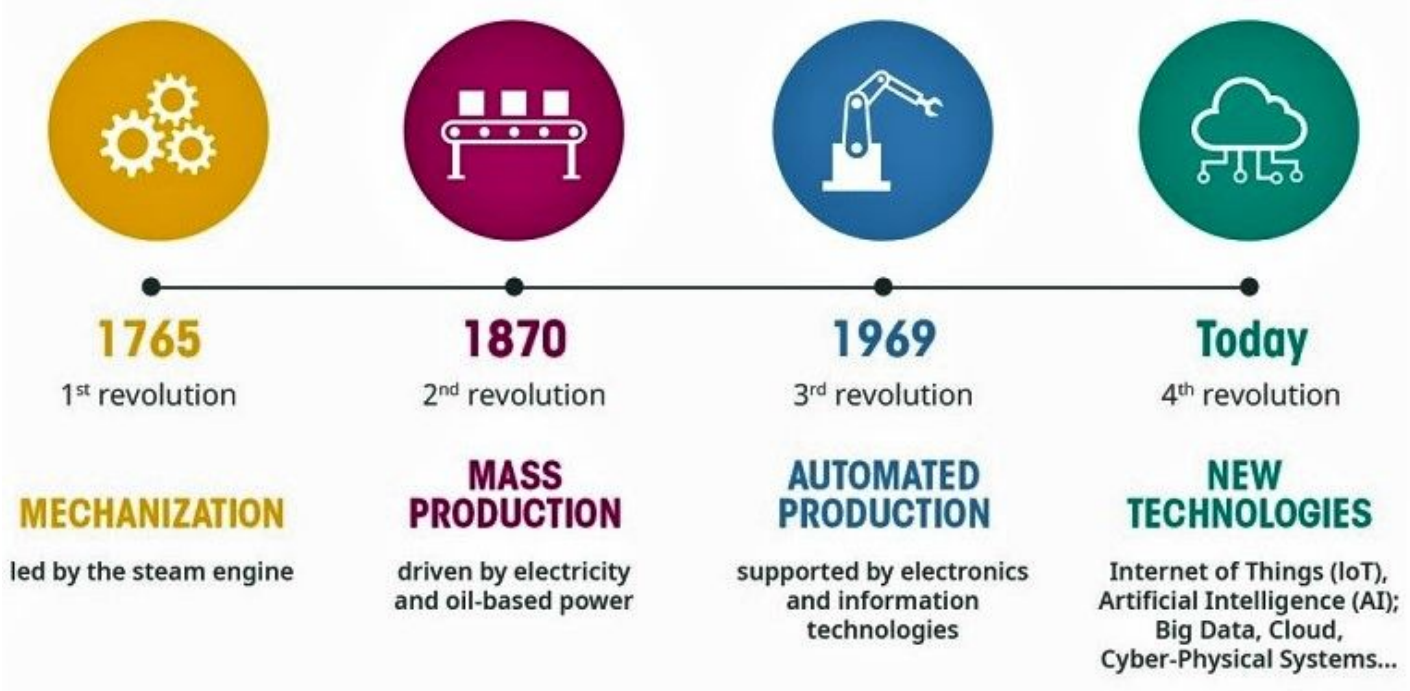
Source: https://www.linkedin.com/pulse/lets-understand-difference-between-machine-learning-vs-gauri-bapat

Machine Learning vs. Deep Learning vs AI?



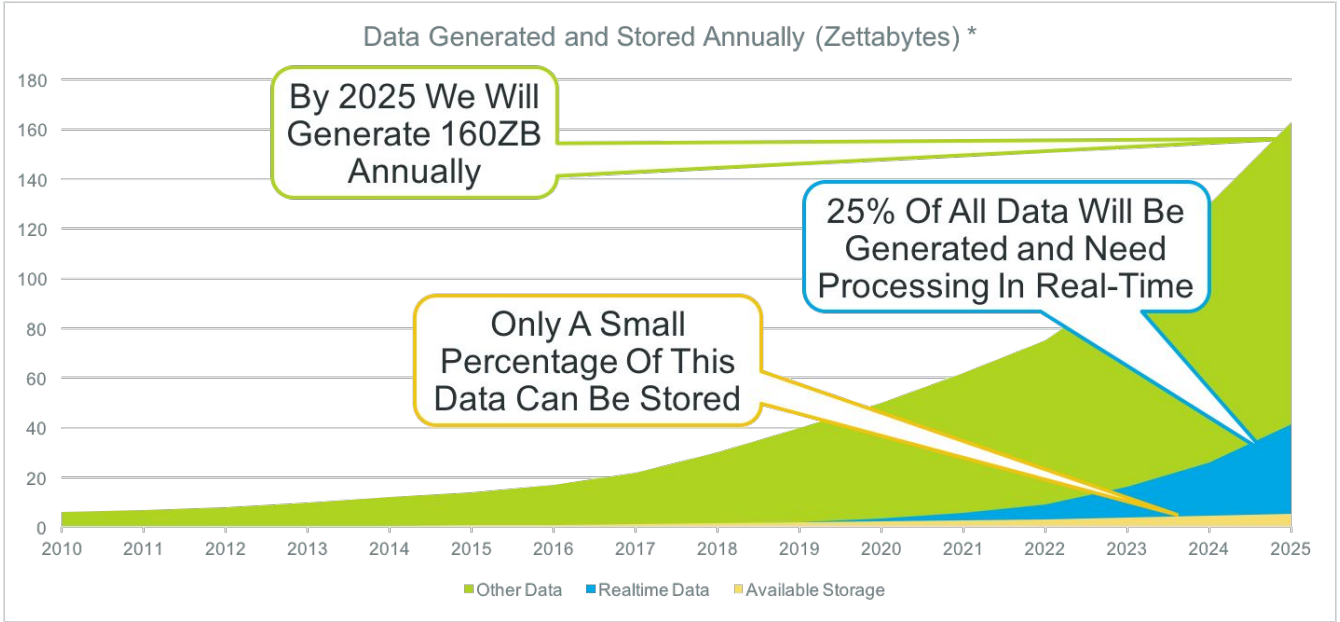
Source: https://www.ibm.com/blogs/systems/ai-machine-learning-and-deep-learning-whats-the-difference/

The Four Industrial Revolutions



Source: https://www.iasparliament.com/current-affairs/fourth-industrial-revolution-4ir

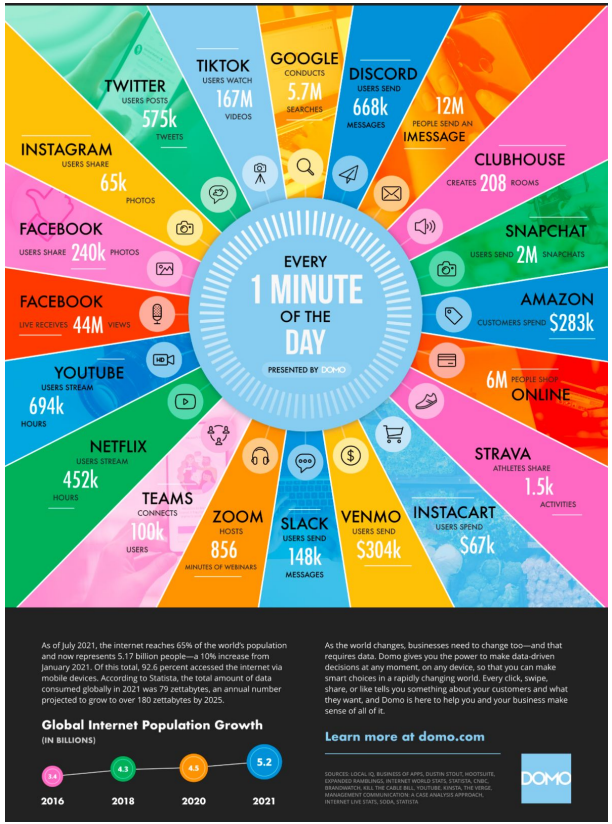
Why Machine Learning?



* Data Age 2025: The Evolution of Data to Life-Critical. An IDC White Paper, Sponsored by Seagate

https://www.striim.com/blog/real-time-data-striim-helps-prepare-for-growth/

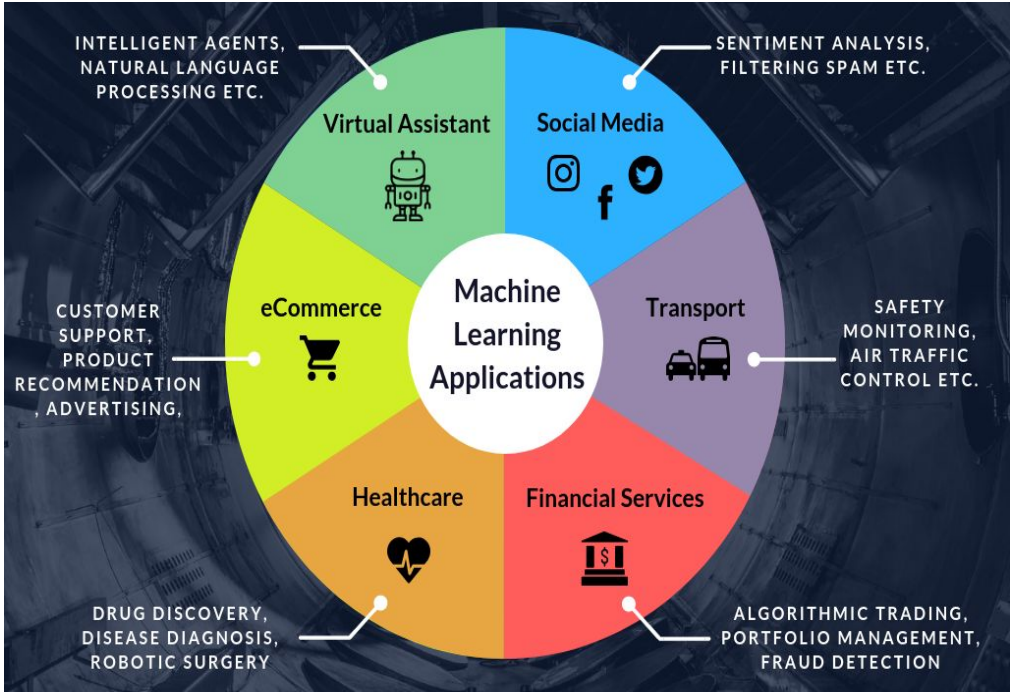
How Much Data Every Minute?



Data never sleeps

Source:
https://dailyinfographic.com/how-much-data-is-generated-every-minute

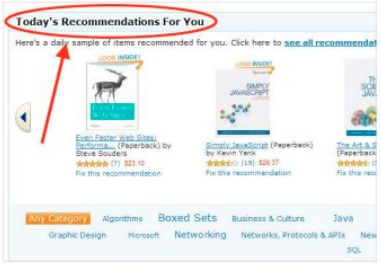
Applications of Machine Learning



Source: https://www.quantinsti.com/blog/machine-learning-basics

Applications of Machine Learning: Examples

spam filtering



recommendation systems

fraud detection

self-driving cars

image recognition

stock market analysis

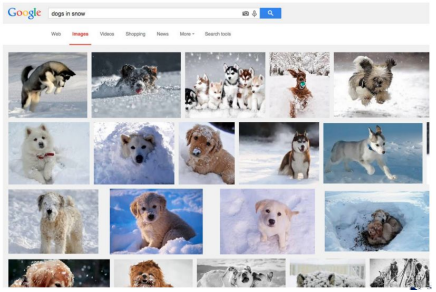
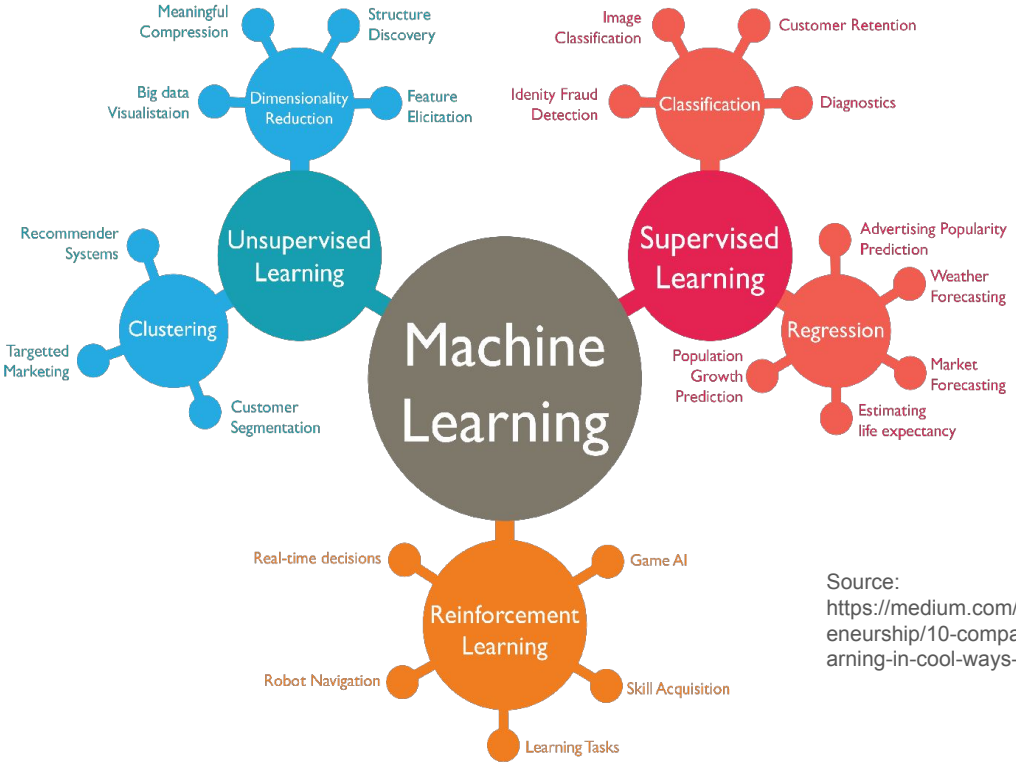


Image retrieval

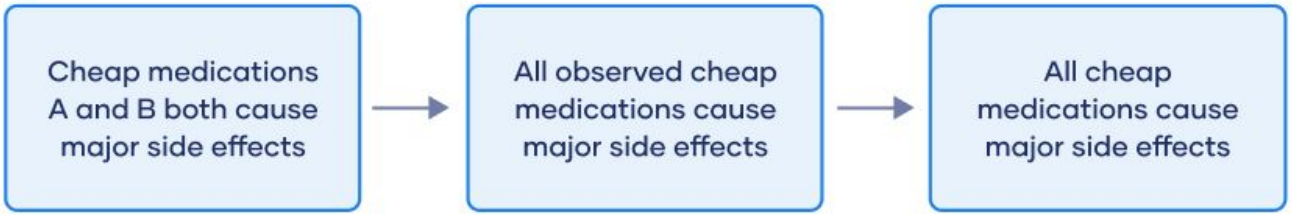
Types of Machine Learning



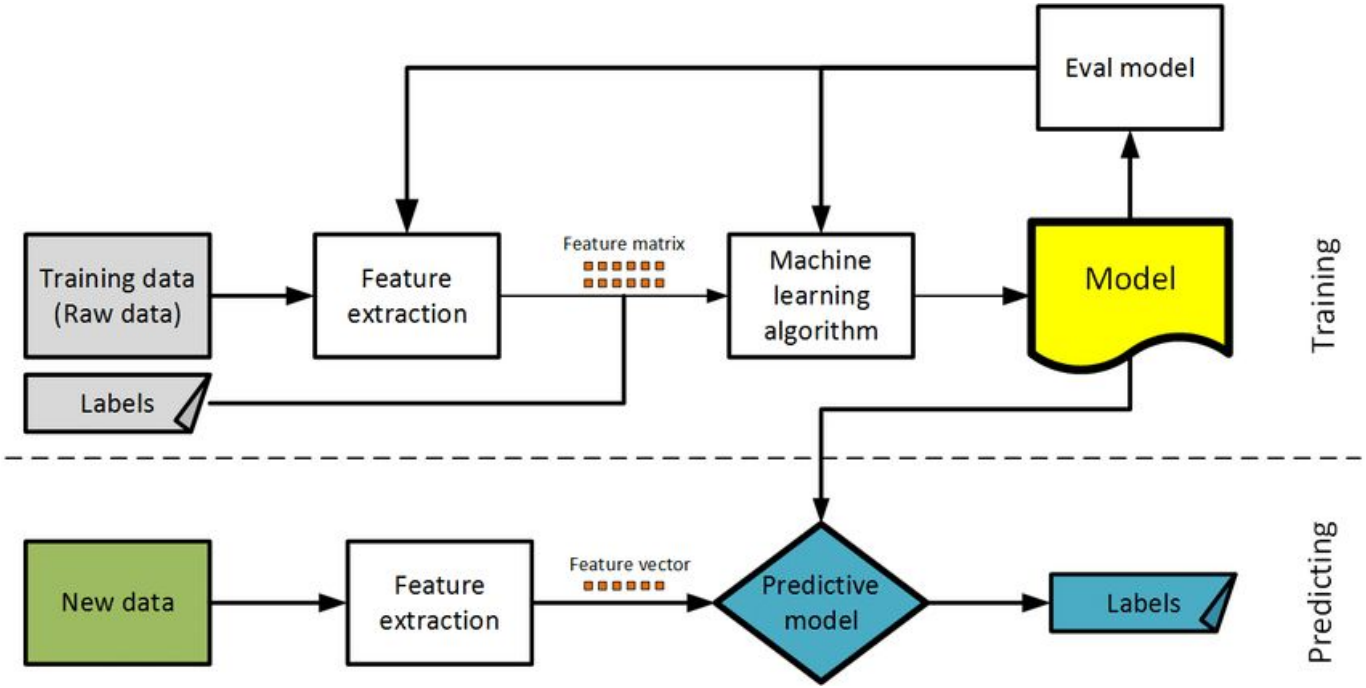
Source: https://medium.com/marketing-and-entrepreneurship/10-companies-using-machine-learning-in-cool-ways-887c25f913c3

Machine Learning ~ Inductive Reasoning

Example



Supervised Learning Workflow

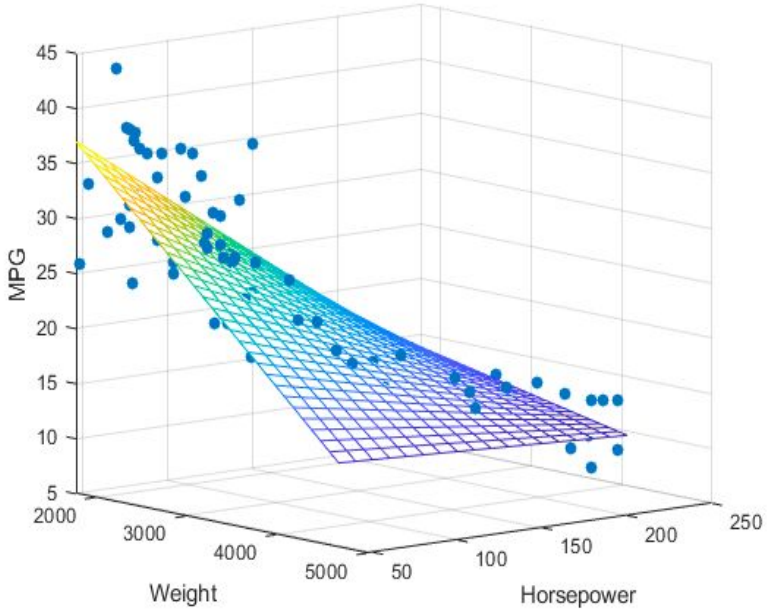


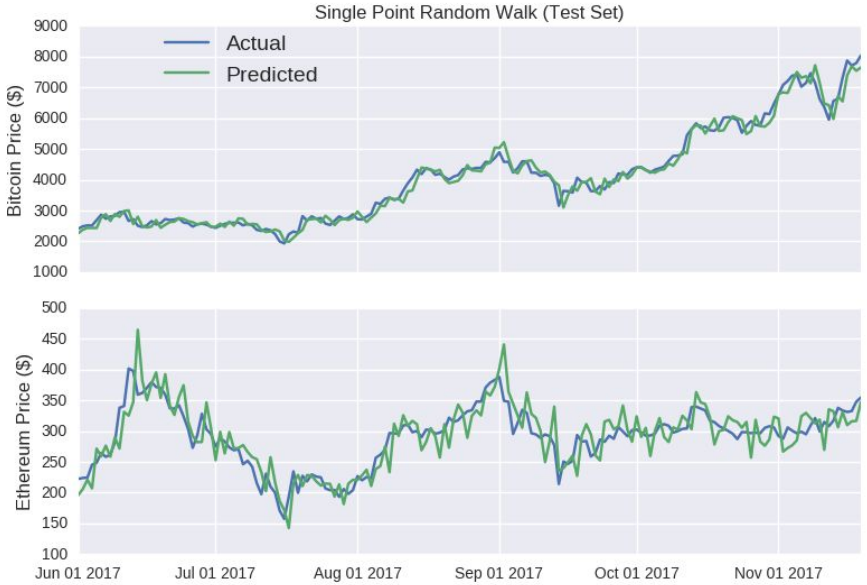
Supervised Learning

- Given a labeled set of input-output pairs $D = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{N}$
  - D is the training set
  - N is the number of labeled examples
  - $\boldsymbol{x}$ is a d-dimensional vector of features/attributes/covariates
  - y is a response variable
- Learn a function $y = f(\boldsymbol{x})$
  - Classification when y is categorical, $y \in \{1, \ldots, C\}$
  - Regression when y is real-valued, $y \in \mathbb{R}$
- Probabilistic modeling $p(y|\boldsymbol{x}, D, \theta)$
  - θ denotes the set of model parameters

Regression

- When the response variable is real-valued $y \in \mathbb{R}$





Regression

Hypothesis:

$$h(x) = w_0 + w_1 x$$

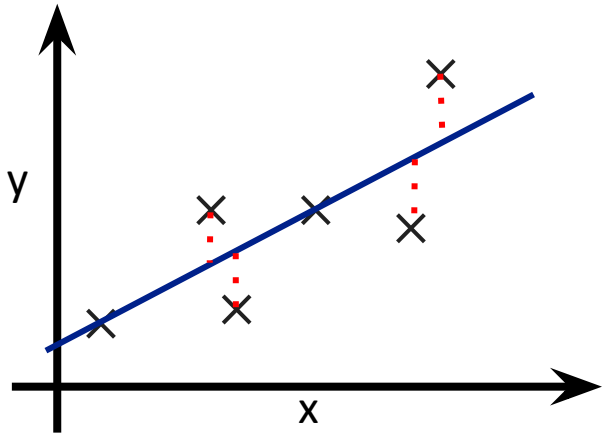
Parameters:

$$w_0, w_1$$

Cost Function: mean squared error (MSE)

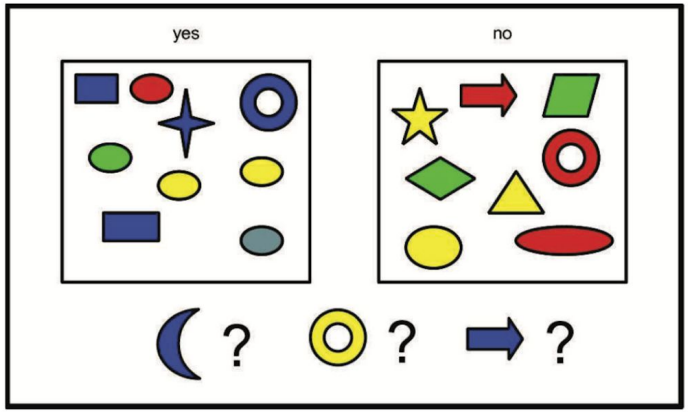
$$J(w_0, w_1) = \frac{1}{2m} \sum_{i=1}^{m} (h(x_i) - y_i)^2$$

Goal: $\min_{w_0, w_1} J(w_0, w_1)$



Classification

- When the response variable is categorical $y \in \{1, \ldots, C\}$
  - When $C = 2$, binary classification
  - When $C > 2$, multiclass classification
  - When $\boldsymbol{x}_i$ may have multiple labels, multi-label classification



(a)

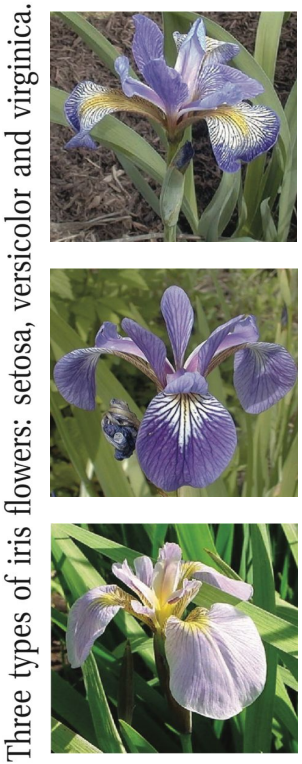
D features (attributes)

N cases

Color	Shape	Size (cm)	Label
Blue	Square	10	1
Red	Ellipse	2.4	1
Red	Ellipse	20.7	0

(b)

Classification: Classifying Flowers



Three types of iris flowers: setosa, versicolor and virginica.

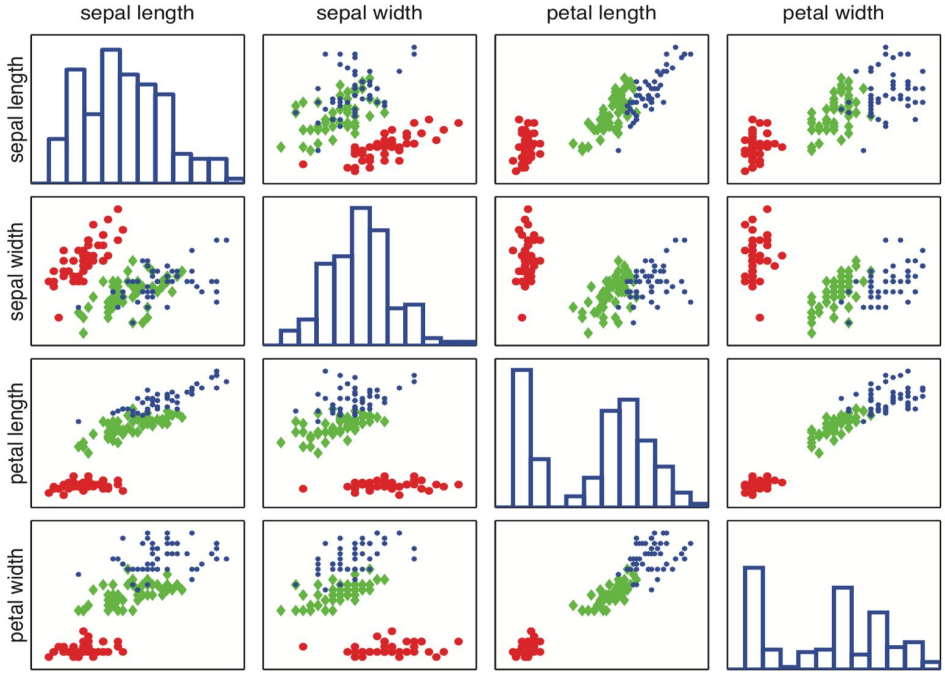
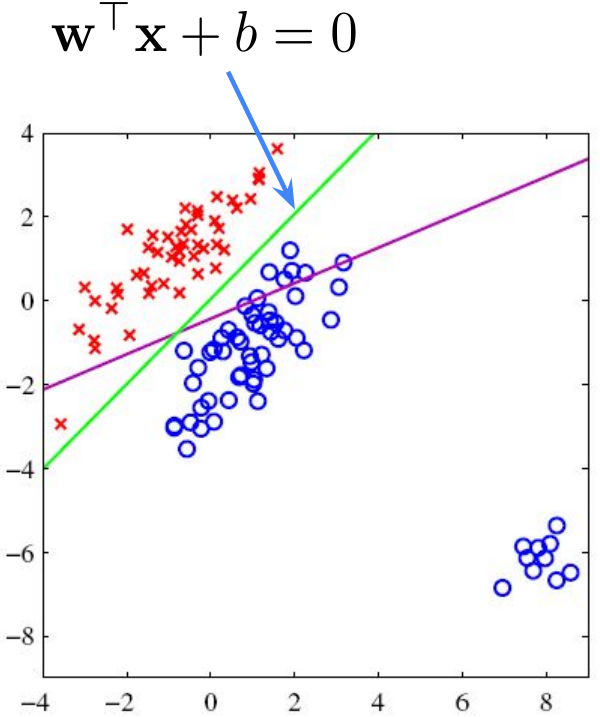


Figure 1.4 Visualization of the Iris data as a pairwise scatter plot. The diagonal plots the marginal histograms of the 4 features. The off diagonals contain scatterplots of all possible pairs of features. Red circle = setosa, green diamond = versicolor, blue star = virginica.

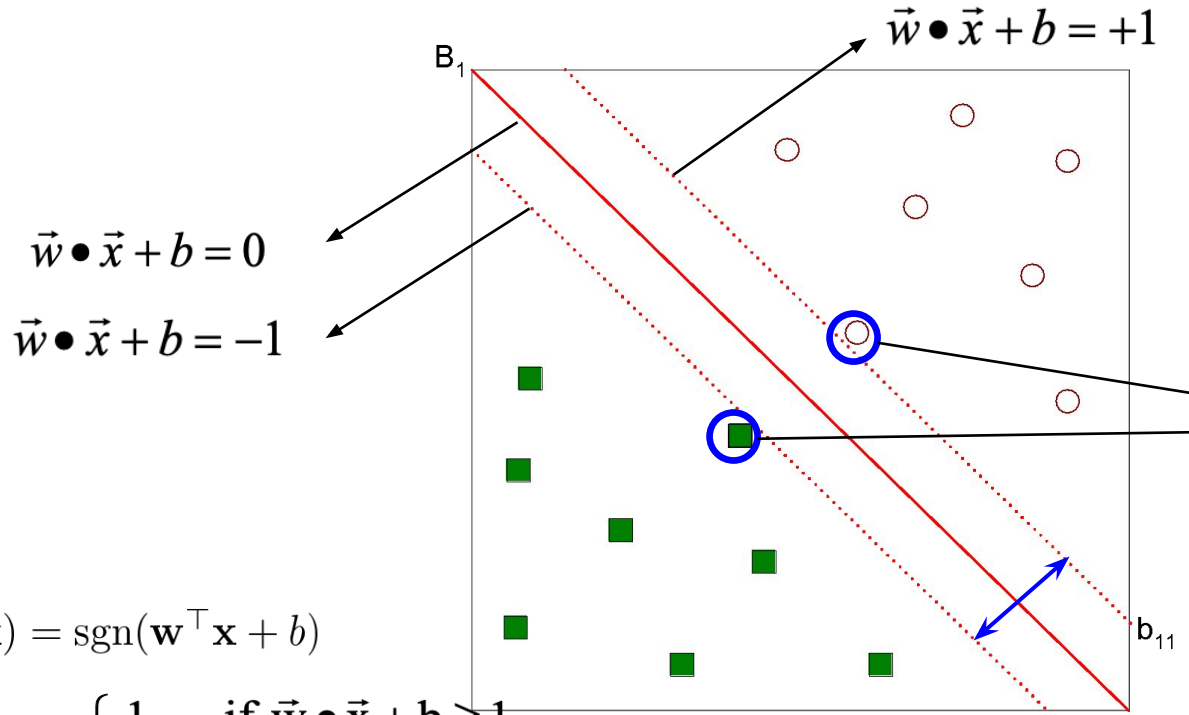
Logistic Regression

$$\ln \frac{p(y=1|\mathbf{x})}{p(y=-1|\mathbf{x})} = \mathbf{w}^\top \mathbf{x}$$

$$\begin{aligned} p(y|\mathbf{x}) &= \frac{1}{\exp(-y\mathbf{w}^\top \mathbf{x}) + 1} \\ &= \sigma(y\mathbf{w}^\top \mathbf{x}) \end{aligned}$$



Support Vector Machines



$$f(\mathbf{x}) = \text{sgn}(\mathbf{w}^\top \mathbf{x} + b)$$

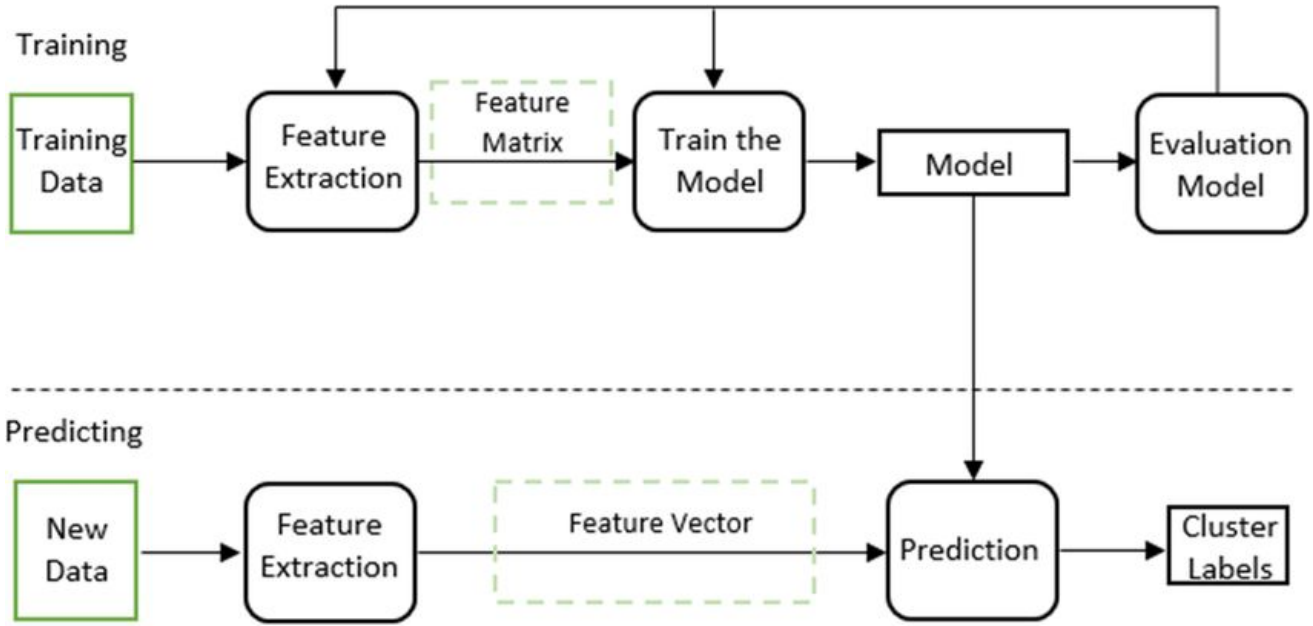
$$f(\vec{x}) = \begin{cases} 1 & \text{if } \vec{w} \bullet \vec{x} + b \geq 1 \\ -1 & \text{if } \vec{w} \bullet \vec{x} + b \leq -1 \end{cases}$$

$$\text{Margin} = \frac{2}{\| \vec{w} \|^2}$$

Unsupervised Learning

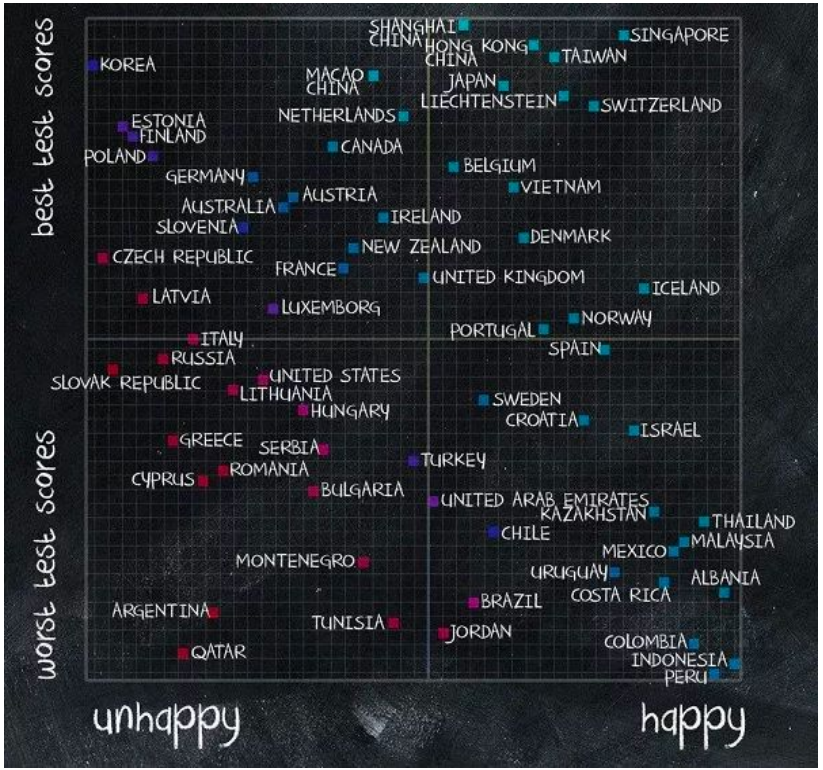
- Given a set of inputs $D = \{\boldsymbol{x}_i\}_{i=1}^{N}$
  - D is the input data
  - N is the number of examples
  - $\boldsymbol{x}$ is a d-dimensional vector of features/attributes/covariates
  - there is no response variable
- Learn a density function $p(\boldsymbol{x}_i|\theta)$
  - θ denotes the set of model parameters

Unsupervised Learning Workflow



Clustering

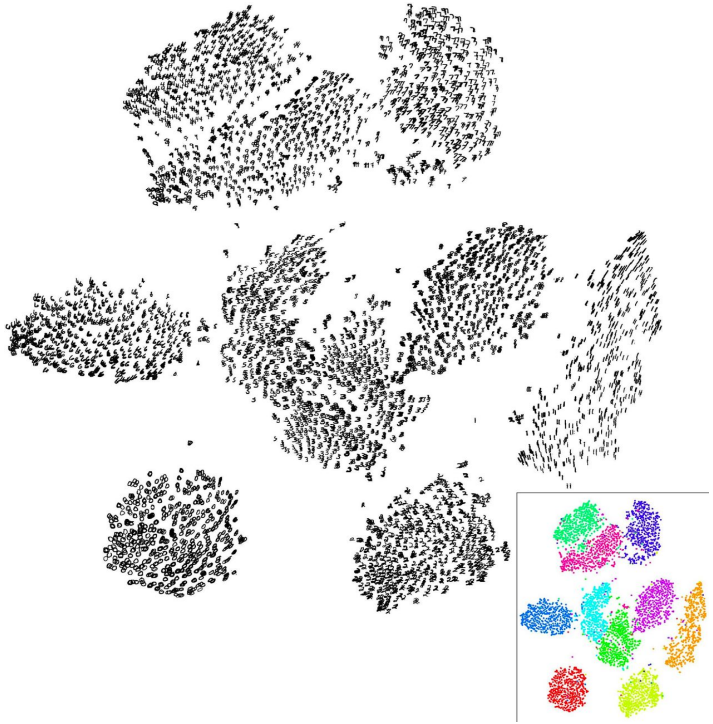
- Clusters:
  - subgroups or subpopulations in the data
- Goals:
  - Discovering the subgroups
  - Estimating which subgroup a data point belongs to



https://i2.wp.com/techielobang.com/blog/wp-content/uploads/2014/01/best-school-happiest-kids.jpg

Clustering Algorithms

Clustering: MNIST Handwritten Digits



Source: http://lvdmaaten.github.io/tsne/examples/mnist_tsne.jpg

”The original black and white (bilevel) images from NIST were size normalized to fit in a 20x20 pixel box while preserving their aspect ratio. The resulting images contain grey levels as a result of the anti-aliasing technique used by the normalization algorithm. the images were centered in a 28x28 image by computing the center of mass of the pixels, and translating the image so as to position this point at the center of the 28x28 field.”

Source: http://yann.lecun.com/exdb/mnist/

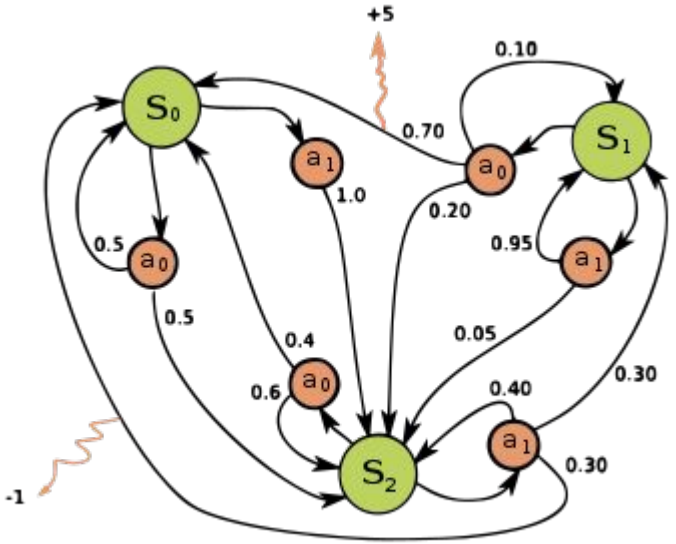
Dimensionality Reduction

Source:
https://www.flickr.com/photos/digitalanthill/4617998281/

Projecting the data to a lower-dimensional subspace that captures the “essence” of the data

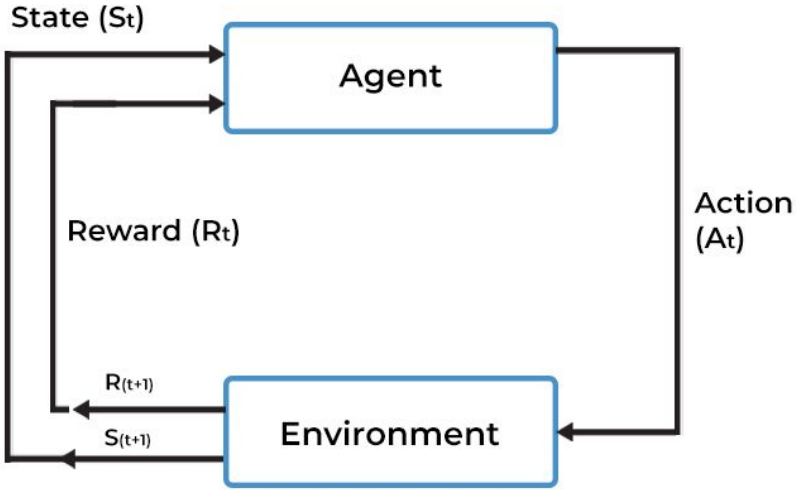
Markov Decision Process

Markov decision process (**MDP**) is a discrete-time stochastic control process. It provides a mathematical framework for modeling decision making in situations where outcomes are partly random and partly under the control of a decision maker



Source: https://en.wikipedia.org/wiki/Markov_decision_process

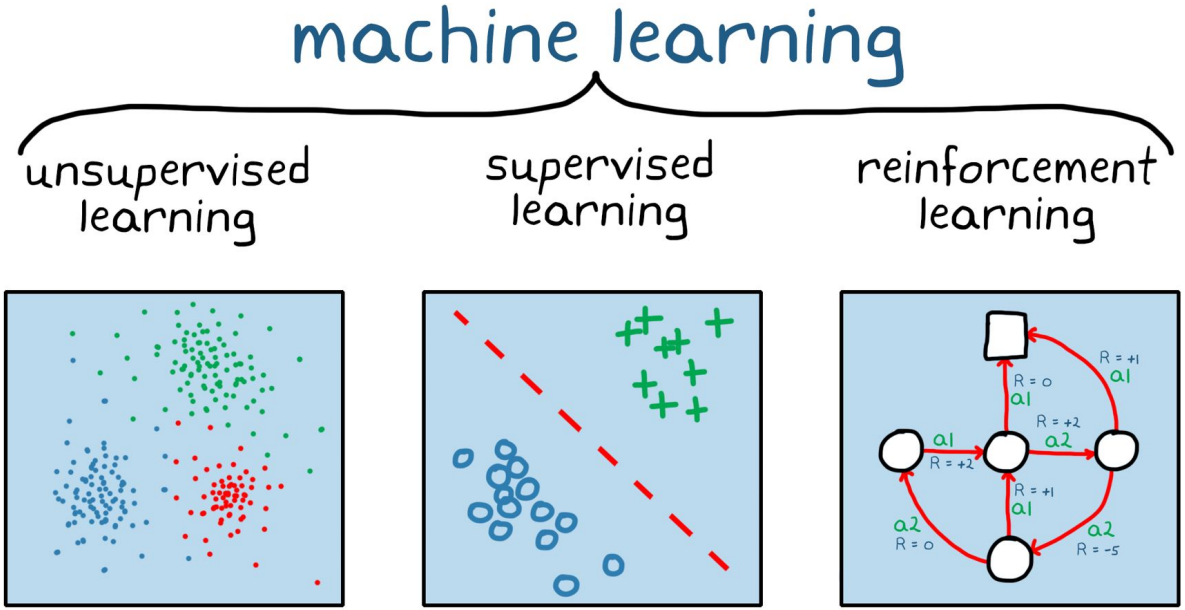
Reinforcement Learning



Reinforcement learning (**RL**) is an area of machine learning concerned with how intelligent agents ought to take actions in an environment in order to maximize the notion of cumulative reward

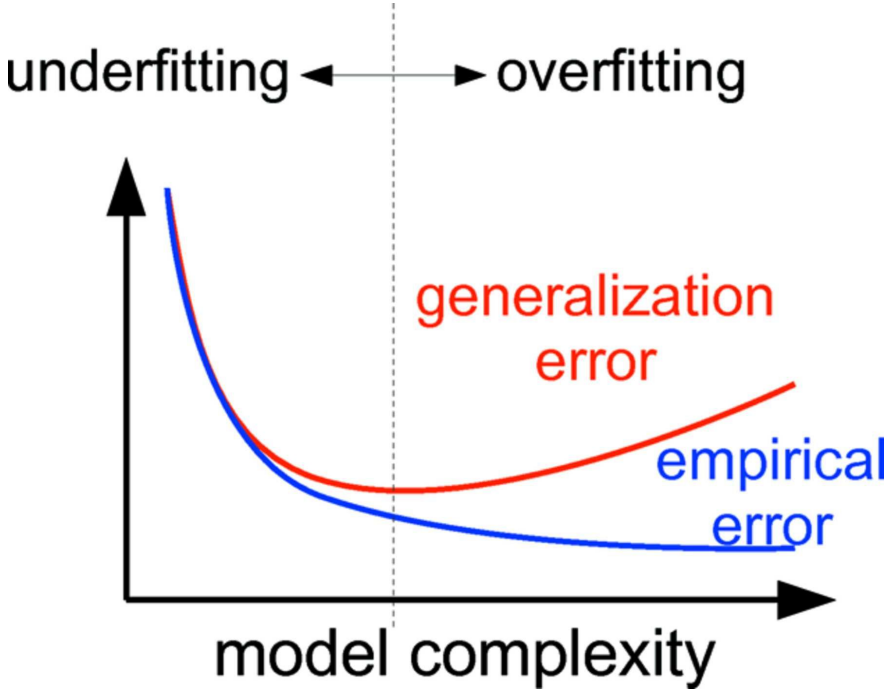
Source: https://en.wikipedia.org/wiki/Reinforcement_learning

Reinforcement Learning vs Traditional Learning

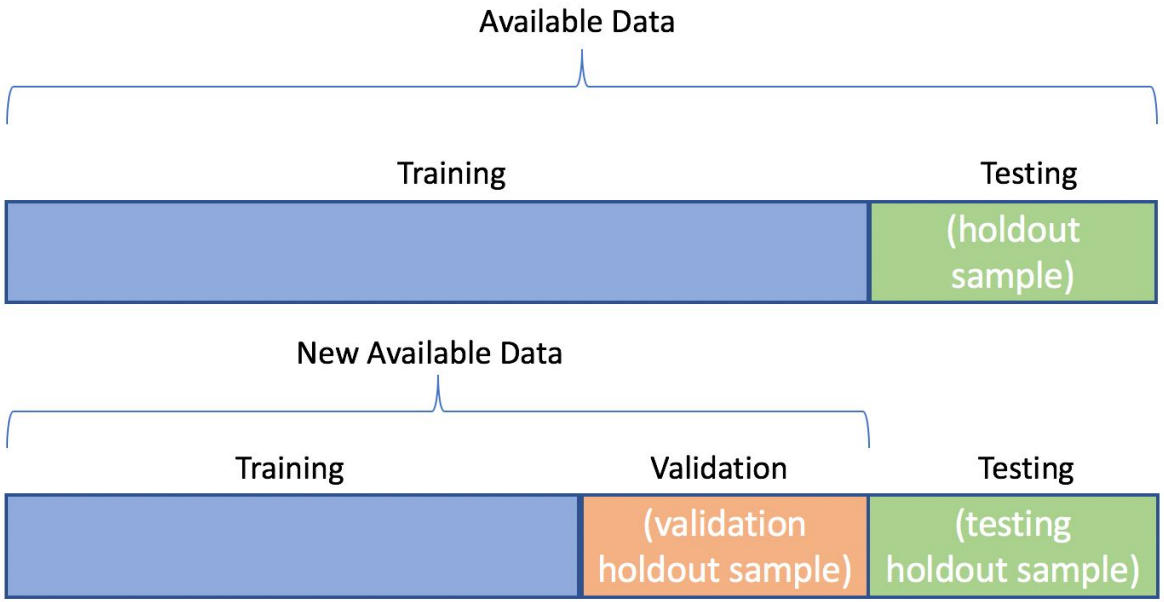


Source:
https://www.mathworks.com/discovery/reinforcement-learning.html

Generalization- and Empirical Error



Train/Validation/Test Data



Cross Validation

Summary

- Introduction to Machine Learning
- Statistics vs Machine Learning
- Machine Learning vs Deep Learning
- Applications of Machine Learning
- Types of Machine Learning
  - Supervised Learning
  - Unsupervised Learning
  - Reinforcement Learning

Thank you

Email me
trongld@vnu.edu.vn

Course: Large Language Models and Its Applications
sponsored by **KEPCO KDN Co., Ltd.**
Eco-friendly & Digital Centered Energy ICT Platform Leader