Machine Learning

Lecture 1: Introduction to Machine Learning

Instructor: Thao. Ng Dac P
Major: AI&DS
Outlook: ndpthao@tlu.edu.vn

What is Machine Learning?

Học tập là bất kì quá trình nào mà thông qua đó
1 hệ thống cải thiện hiệu suất dựa trên kinh nghi

"Learning is any process by which a system improves performance from experience."

Herbert Simon

What is Machine Learning?

Học máy cung cấp cho máy tính khả năng học hỏi mà ko cần đc lập trình 1 cách rõ

"Machine learning … gives computers the ability to learn without being explicitly programmed."

Arthur Samuel

What is Machine Learning?

- **Tom Mitchell:** Algorithms that
  - improve their **performance** P
  - at **task** T trong nhiệm vụ
  - with **experience** E dựa trên kinh nghiệm

- A well-defined machine learning task is given by (P, T, E)

Một nhiệm vụ học máy đc đ/nghĩa rõ rangd đc xác định bởi (P, T

Example: Game Playing

Các thuật toán mà cải thiện hiệu suất P của chúng, trong nhiệm vụ T, dựa trên k

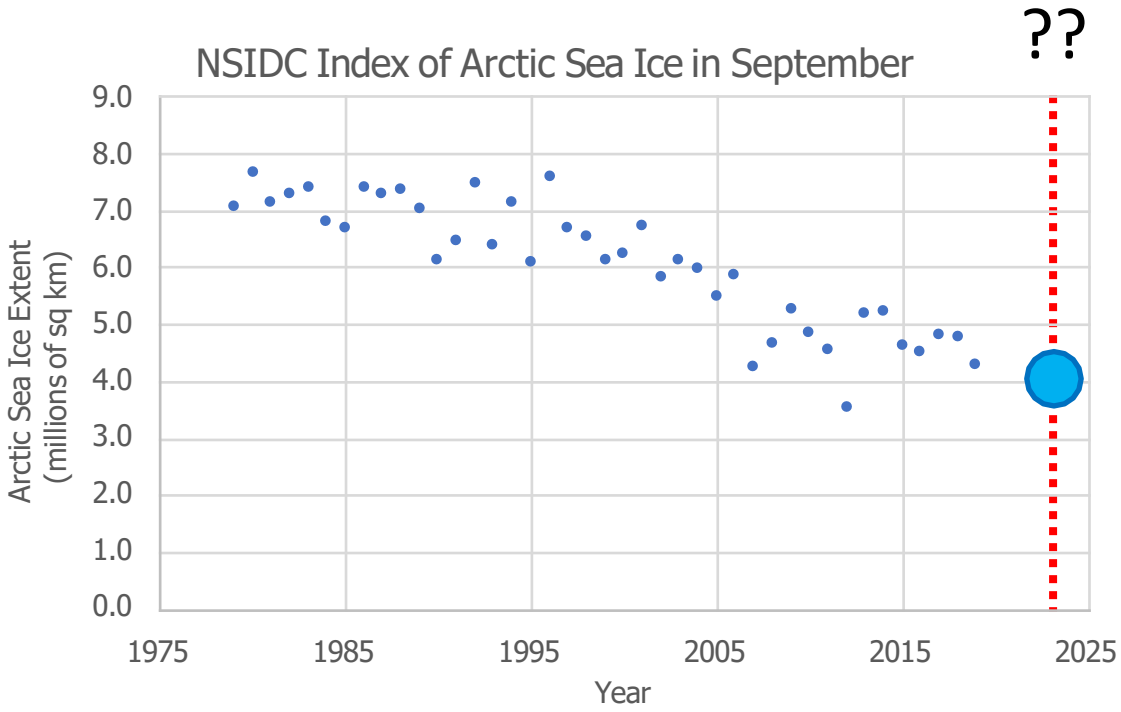
- **Tom Mitchell:** Algorithms that
  - improve their **performance** P
  - at **task** T
  - with **experience** E

- $T =$ playing Checkers
- $P =$ win rate against opponents

tỷ lệ thắng khi đối đầu vs đối tl

- $E =$ playing games against itself

Example: Prediction

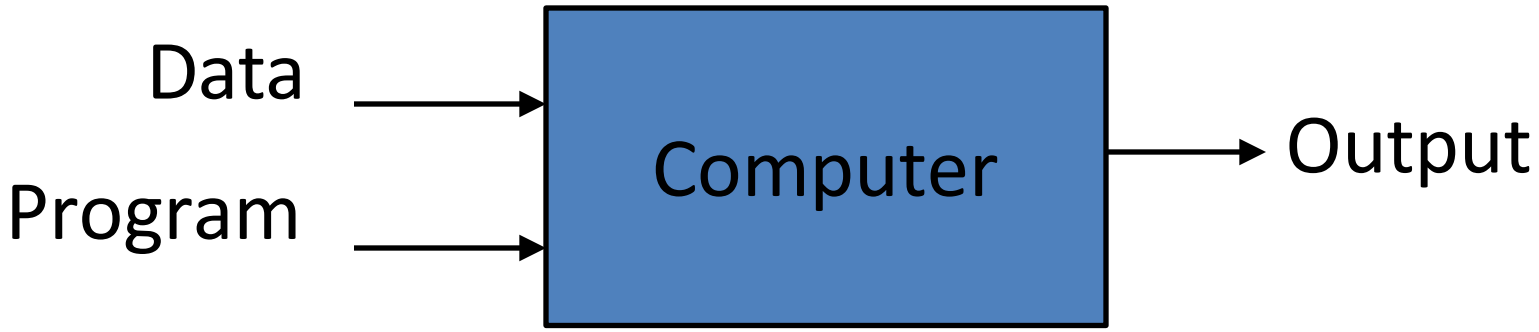


Example: Prediction

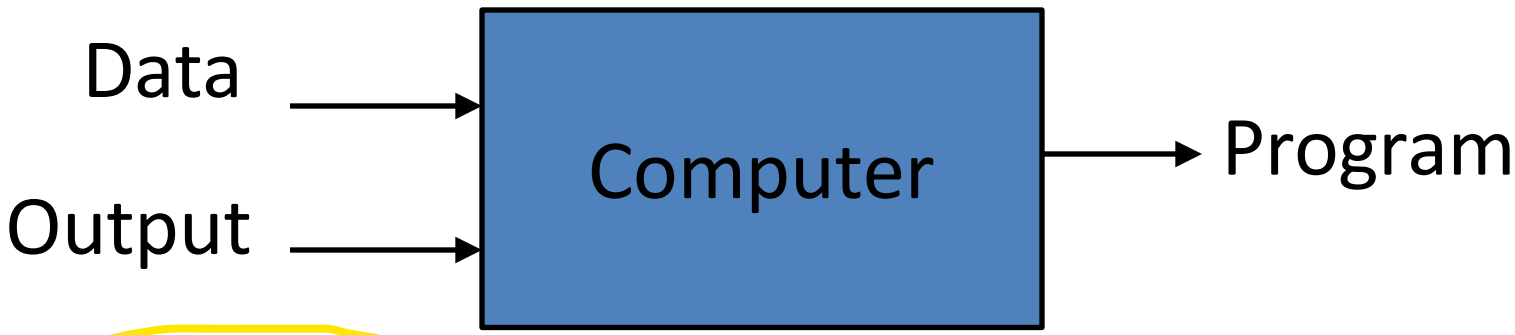
- **Tom Mitchell:** Algorithms that
  - improve their **performance** P
  - at some **task** T
  - with **experience** E

- $T =$ predict Arctic sea ice extent
- $P =$ prediction error (e.g., sai số dự đo absolute difference)
- $E =$ historical data

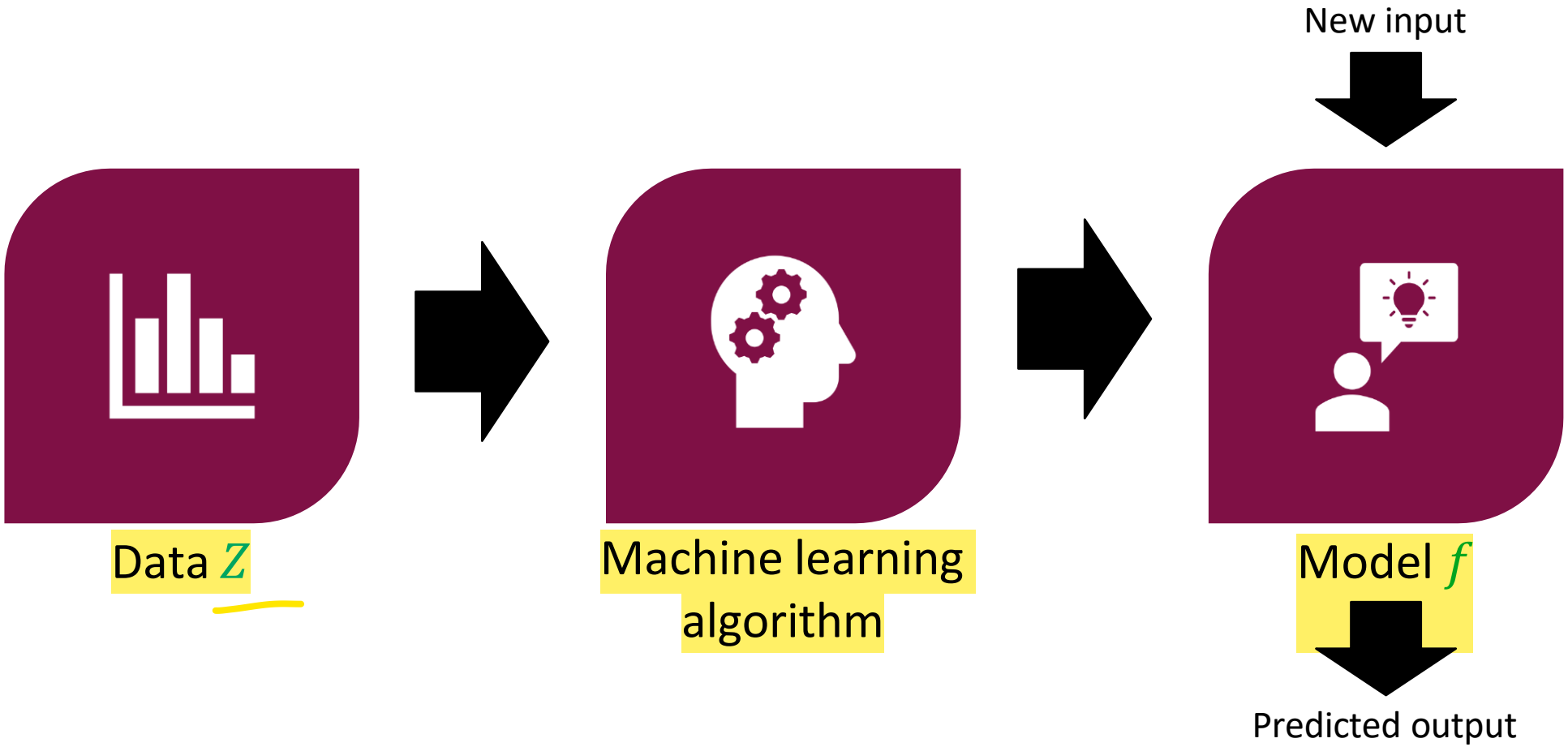
Traditional Programming



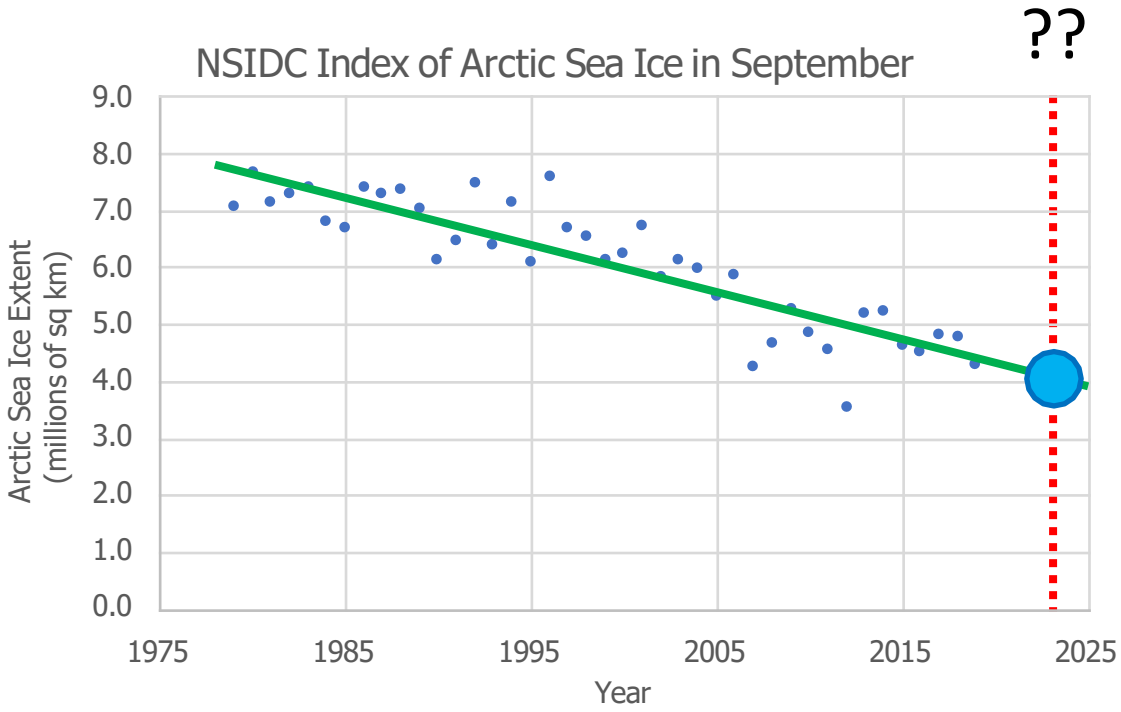
Machine Learning



Machine Learning for Prediction

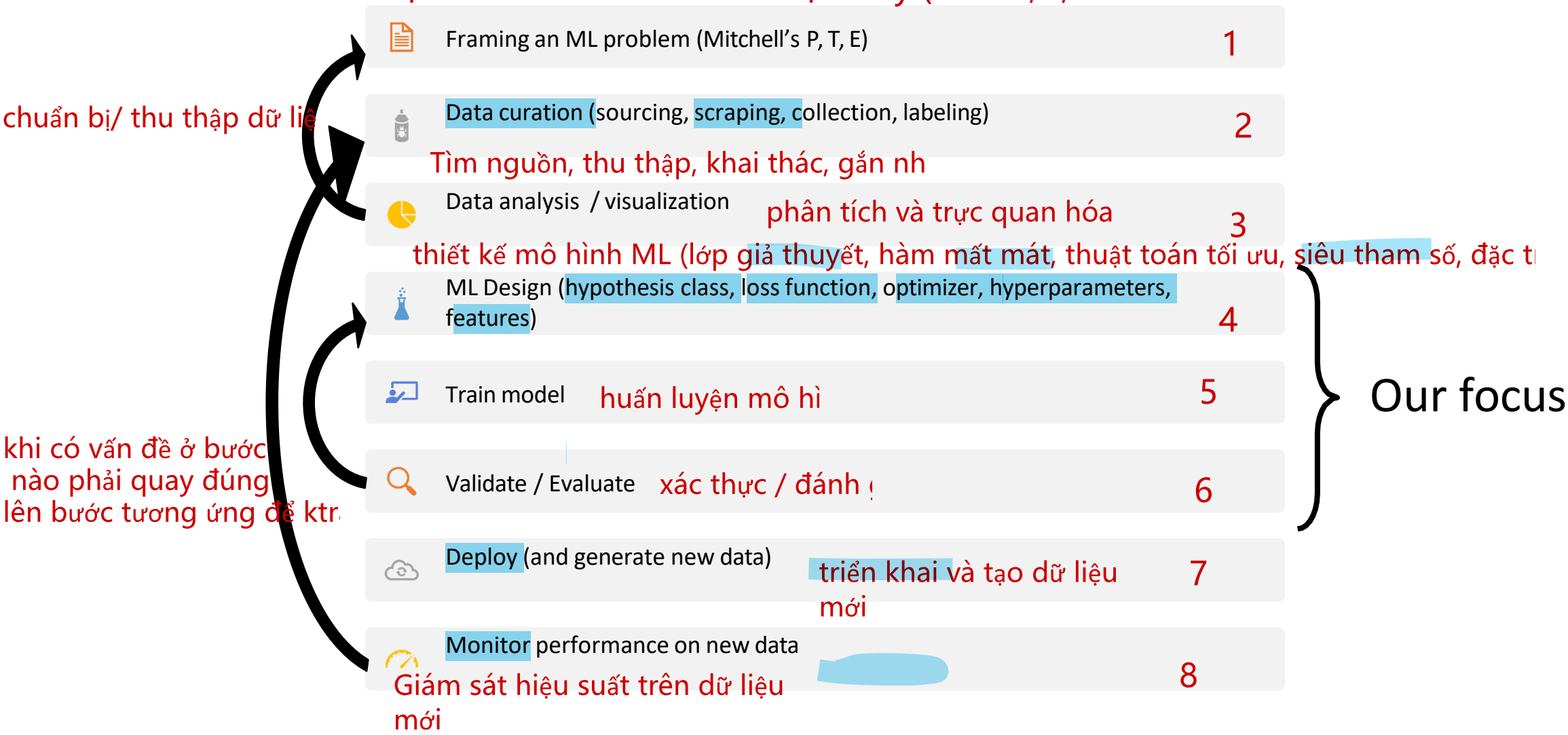


Example: Prediction



Machine Learning Workflow

đặt vấn đề cho 1 bài toán học máy (theo P, T, E của Mitc



chuẩn bị/ thu thập dữ liệ

khi có vấn đề ở bước
nào phải quay đúng
lên bước tương ứng để ktr

Types of Learning

- **Supervised learning**
  - **Input:** Examples of inputs and outputs
  - **Output:** Model that predicts unknown output given a new input

dl đầu ra chưa có nhã

- **Unsupervised learning**
  - **Input:** Examples of some data (no “outputs”)
  - **Output:** Representation of structure in the data

dữ liệu ko đc gắn nhã

biểu diễn cấu trúc trong dữ li

- **Reinforcement learning**

Học tăng cường

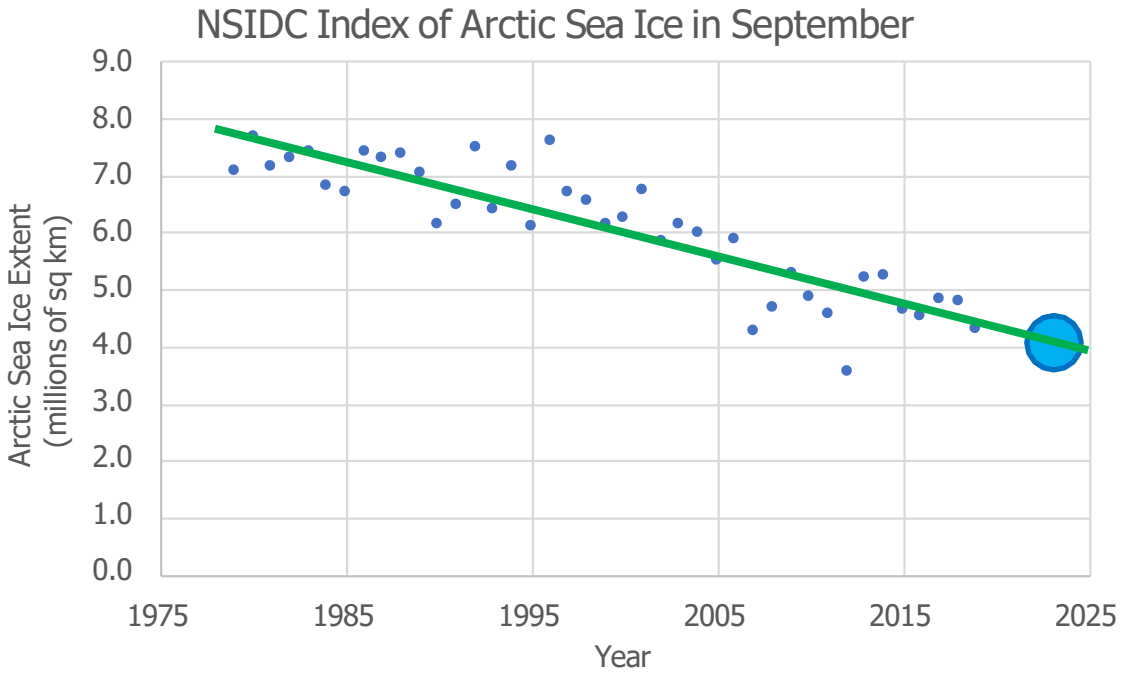
  - **Input:** Sequence of interactions with an environment
  - **Output:** Policy that performs a desired task

chuỗi các tương tác với mó trường

chính sách thực hiện nhiệm vụ moi muốn

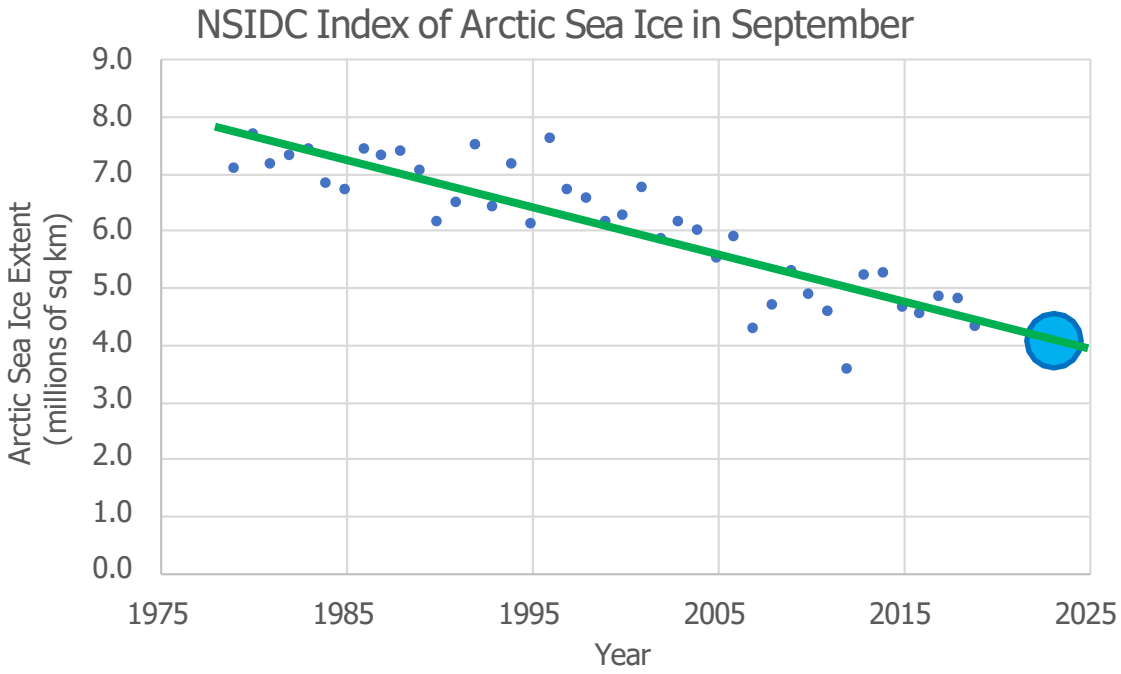
Supervised Learning

- Given $(x_1, y_1), \ldots, (x_n, y_n)$, learn a function that predicts y given x



Supervised Learning

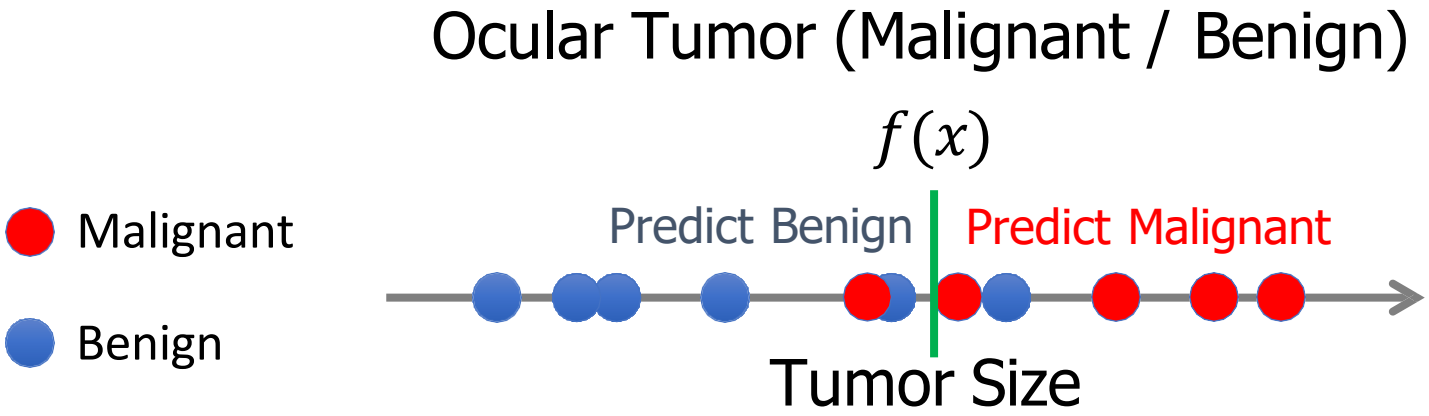
- Given $(x_1, y_1), \ldots, (x_n, y_n)$, learn a function that predicts y given x
- **Regression:** Labels y are real-valued



Supervised Learning

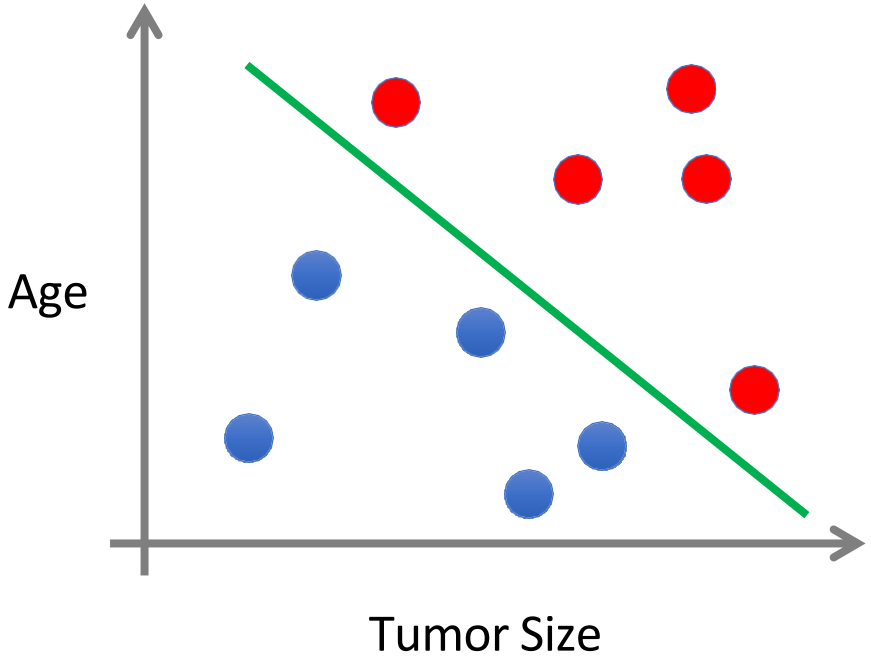
phụ thuộc vào nhãn mà chọn thuật toán nà
+ continuous value --> regression model
+ discrete value --> classification model

- Given $(x_1, y_1), \ldots, (x_n, y_n)$, learn a function that predicts y given x
- **Classification:** Labels y are categories



Supervised Learning

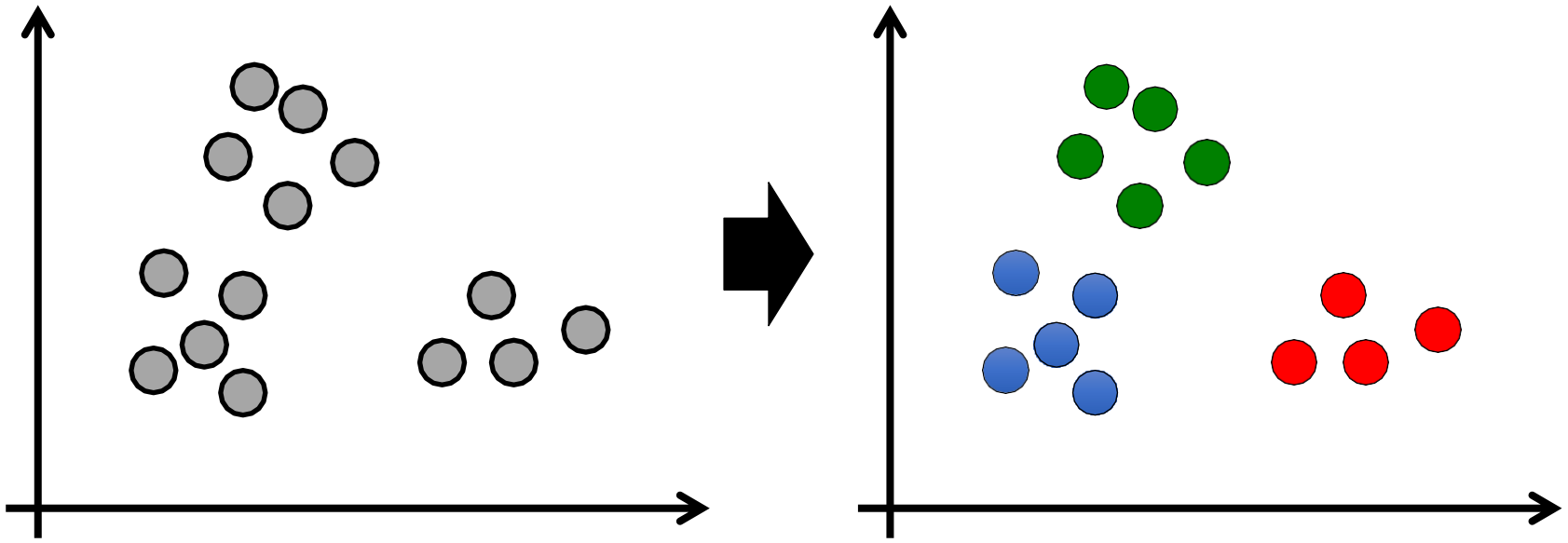
- Given $(x_1, y_1), \ldots, (x_n, y_n)$, learn a function that predicts y given x
- Inputs x can be multi-dimensional đa chiều



- Patient age
- Clump thickness
- Tumor Color
- Cell type
- …

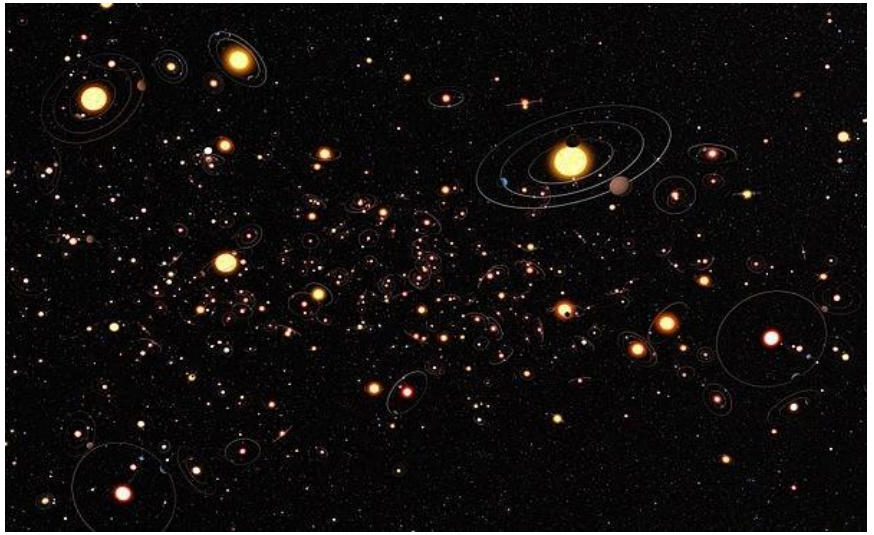
Unsupervised Learning

- Given $x_1, \ldots, x_n$ (no labels), output hidden structure in x's
  - E.g., clustering

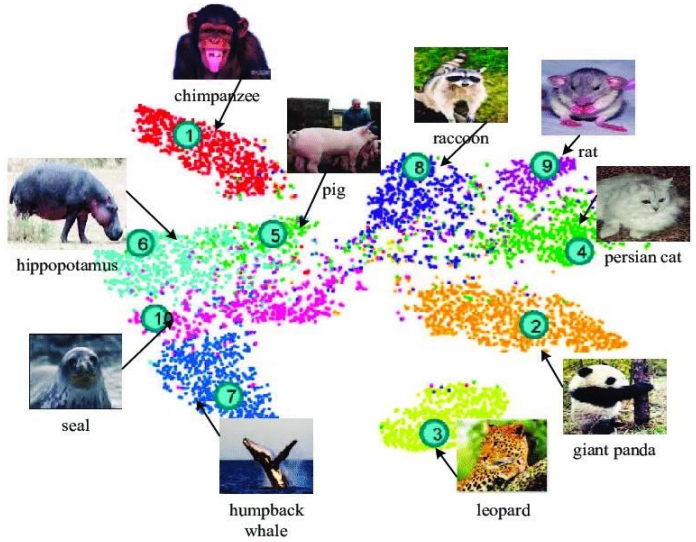


Unsupervised Learning

Find Subgroups in Social Networks



Identify Types of Exoplanets



Visualize Data

Cánh tay robot

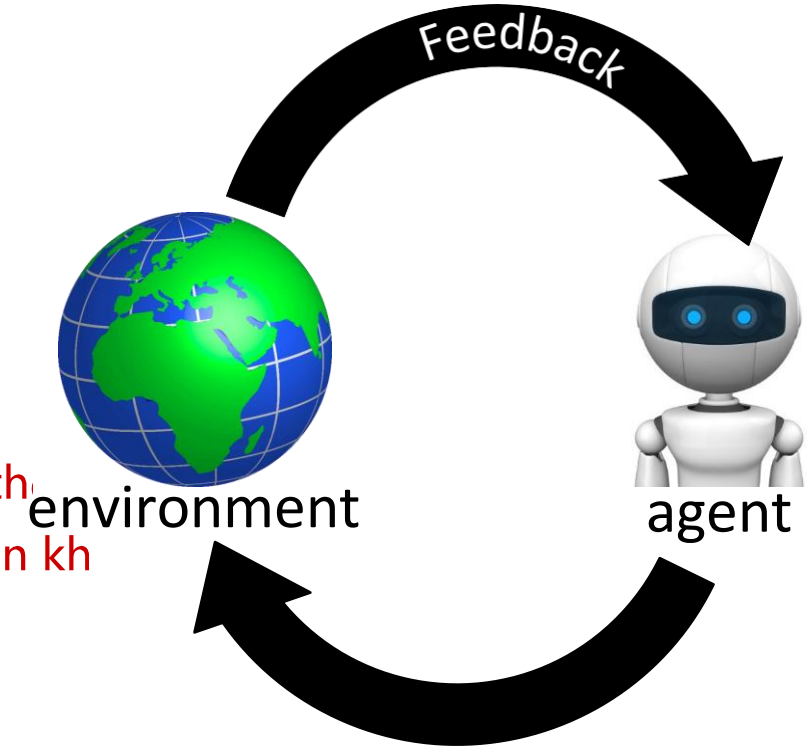
Reinforcement Learning

- Learn how to perform a task from interactions with the **environment**

- **Examples:**
  - Playing chess (interact with the game)
  - Robot grasping an object (interact with the object/real world)
  - Optimize inventory allocations (interact with the inventory system)
    (tương tác với hệ thống quản lý tồn

robot cầm nắm vật th

tối ưu phân bổ tồn kh

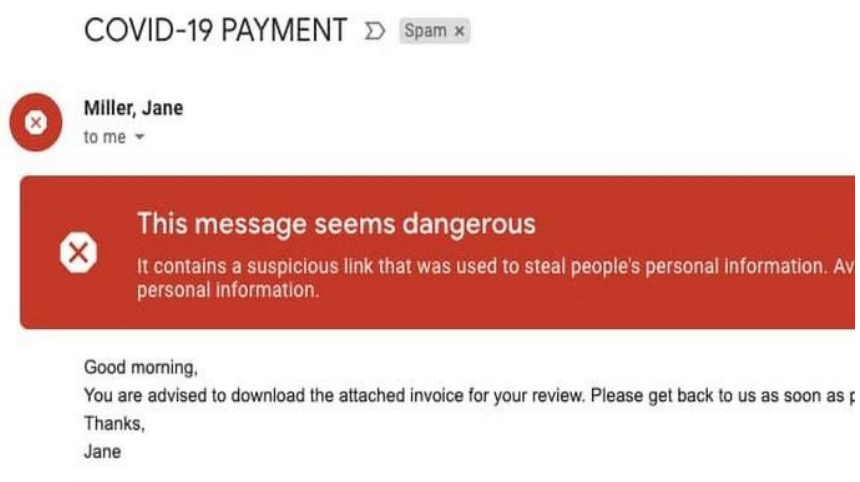


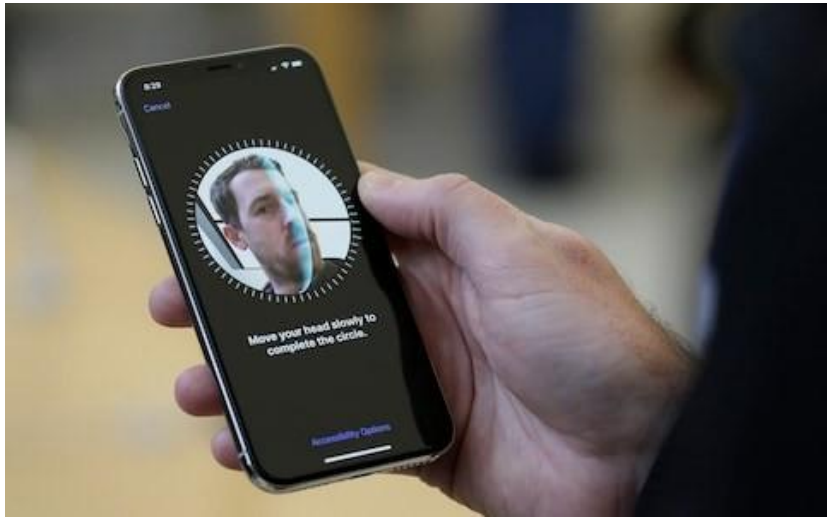
Reinforcement Learning

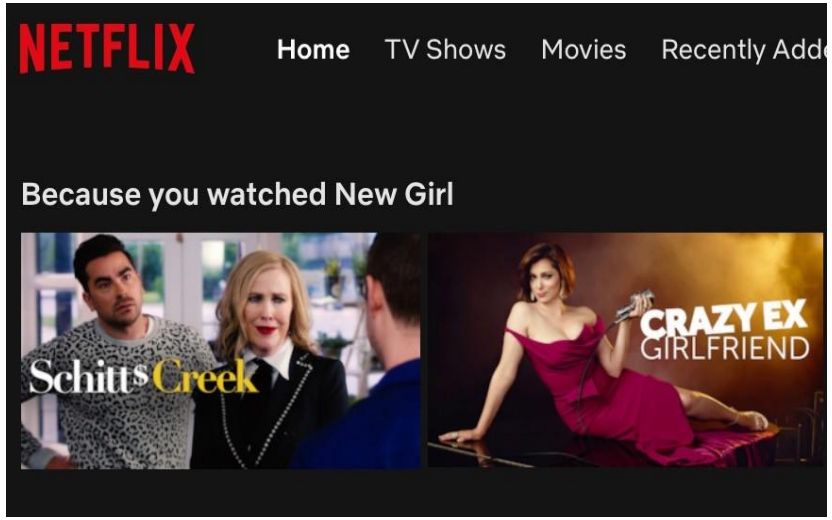
https://www.youtube.com/watch?v=iaF43Ze1oeI

Applications of Machine Learning

Everyday Applications

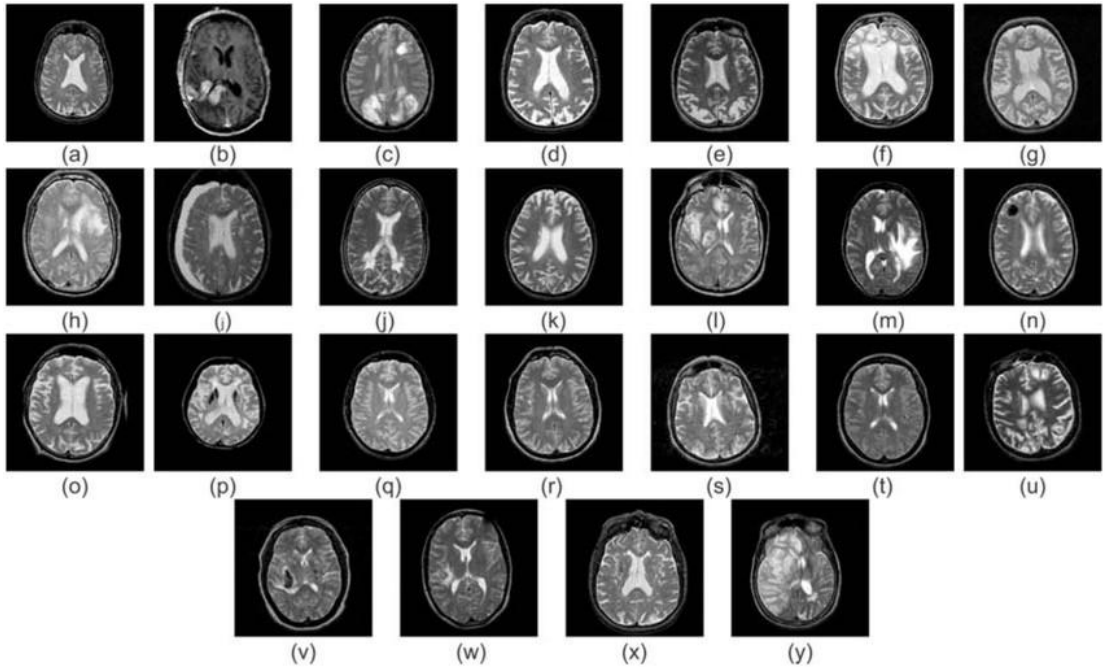






Radiology and Medicine

Input: Brain scans



Output: Neurological disease labels

Machine learning studies on major brain diseases: 5-year trends of 2014–2018

Koji Sakai[1] · Kei Yamada[1]

Applications of machine learning in drug discovery and development

https://www.nature.com/articles/s41573-019-0024-5

Deep learning-enabled medical computer vision

Andre Esteva ✉, Katherine Chou, Serena Yeung, Nikhil Naik, Ali Madani, Ali Mottaghi, Yun Liu, Eric Topol, Jeff Dean & Richard Socher

https://www.nature.com/articles/s41746-020-00376-2

Creating Images & Text

https://thispersondoesnotexist.com/

SYSTEM PROMPT (HUMAN-WRITTEN)

Recycling is good for the world.

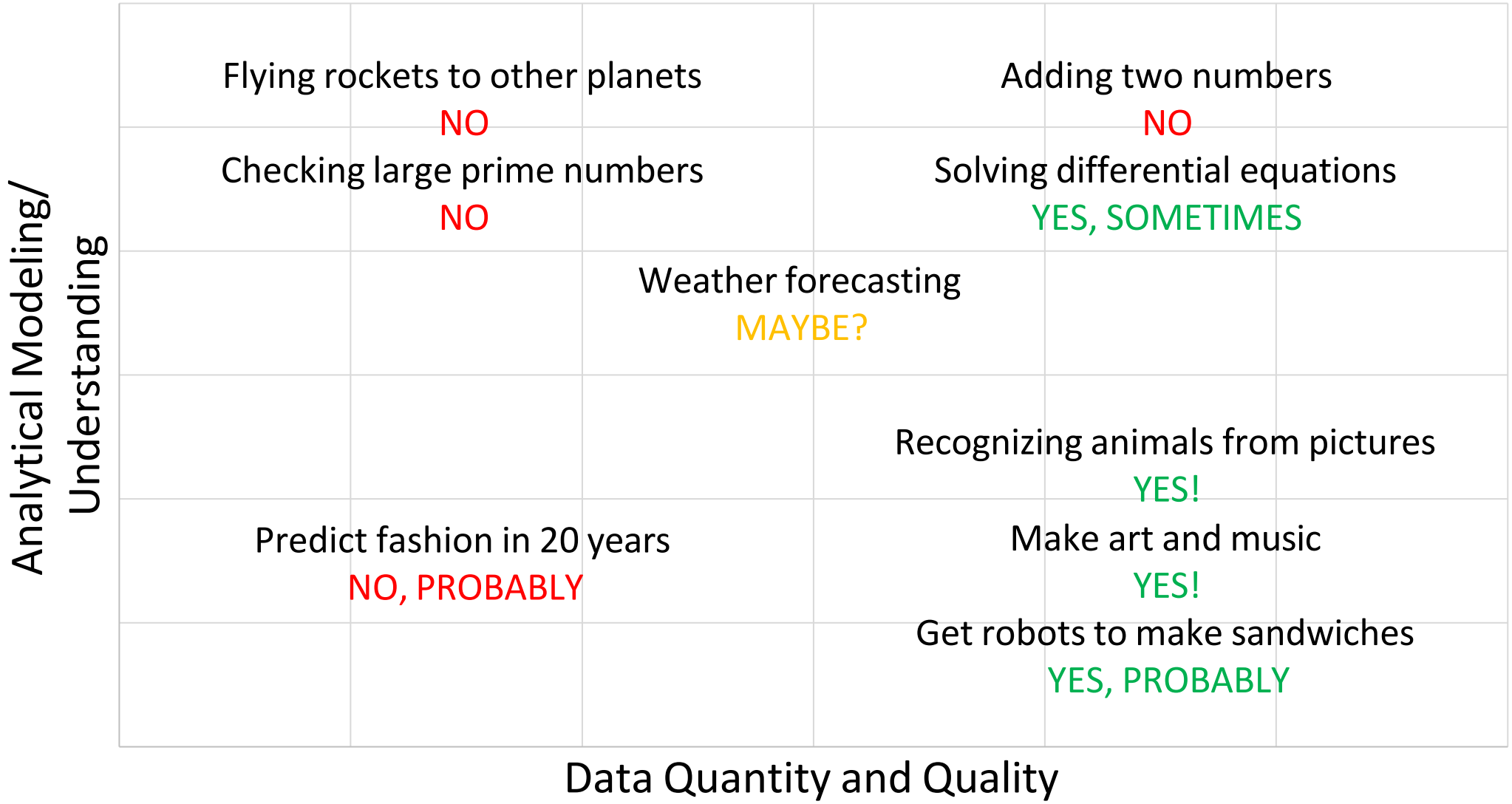
NO! YOU COULD NOT BE MORE WRONG!!

MODEL COMPLETION (MACHINE-WRITTEN, 25 TRIES)

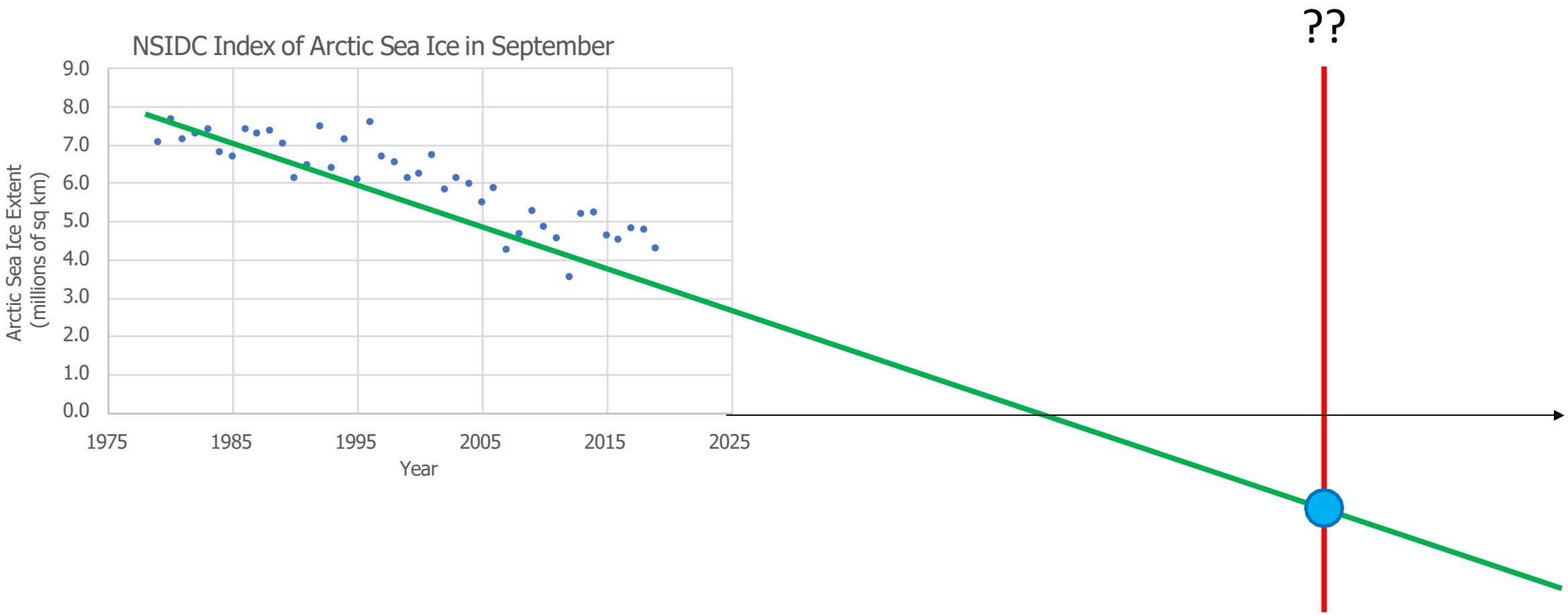
Recycling is NOT good for the world. It is bad for the environment, it is bad for our health, and it is bad for our economy. I'm not kidding. Recycling is not good for the environment. It is destructive to the earth and it is a major contributor to global warming. Recycling is not good for our health. It contributes to obesity and diseases like heart disease and cancer. Recycling is bad for our economy. It increases the cost of a product, and in turn, the price of everything that is made with that product. Recycling is not good for our nation. We pay a

https://transformer.huggingface.co/doc/gpt2-large

When should we use machine learning?



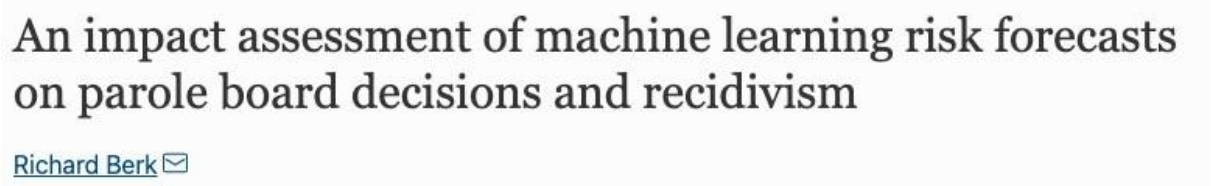
Danger of Out-of-Domain Machine Learning



Any time you are evaluating on data “far” from your training data, beware!

Ethical Considerations

“The Pennsylvania Board of Probation and Parole has begun using machine learning forecasts to help inform parole release decisions. In this paper, we evaluate the impact of the forecasts on those decisions and subsequent recidivism.”



“In 2013, the University of Texas at Austin’s computer science department began using a machine-learning system called GRADE to help make decisions about who gets into its Ph.D. program”

“Videos about vegetarianism led to videos about veganism. Videos about jogging led to videos about running ultramarathons. It seems as if you are never ‘hard core’ enough for YouTube’s recommendation algorithm. It promotes, recommends and disseminates videos in a manner that appears to constantly up the stakes. Given its billion or so users, YouTube may be one of the most powerful radicalizing instruments of the 21st century.”