Introduction to Machine Learning and Data Science

Nguyen Kim Thang

Grenoble INP, University Grenoble Alpes

Planning

- Introduction to AI
- Unsupervised Machine Learning
- Supervised Machine Learning
- Regularization
- Deep Learning
  - Convolution Neural Networks, Recurrent Neural Networks

Classes:
- 9-13/11 + 20/11: 8h30-10h lecture, 10h15-11h45: lab
- 19/11: 14h-16h30

Website: https://datamove.imag.fr/kimthang.nguyen/Teaching/ML-PFIEV/ml-pfiev.html

Nguyen Kim Thang → Teaching → PFIEV, HUST

What is Artificial Intelligence?

- Machines that think like humans
- Machines that think rationally
- Machines that act like humans
- Machine that act rationally

What is Artificial Intelligence?

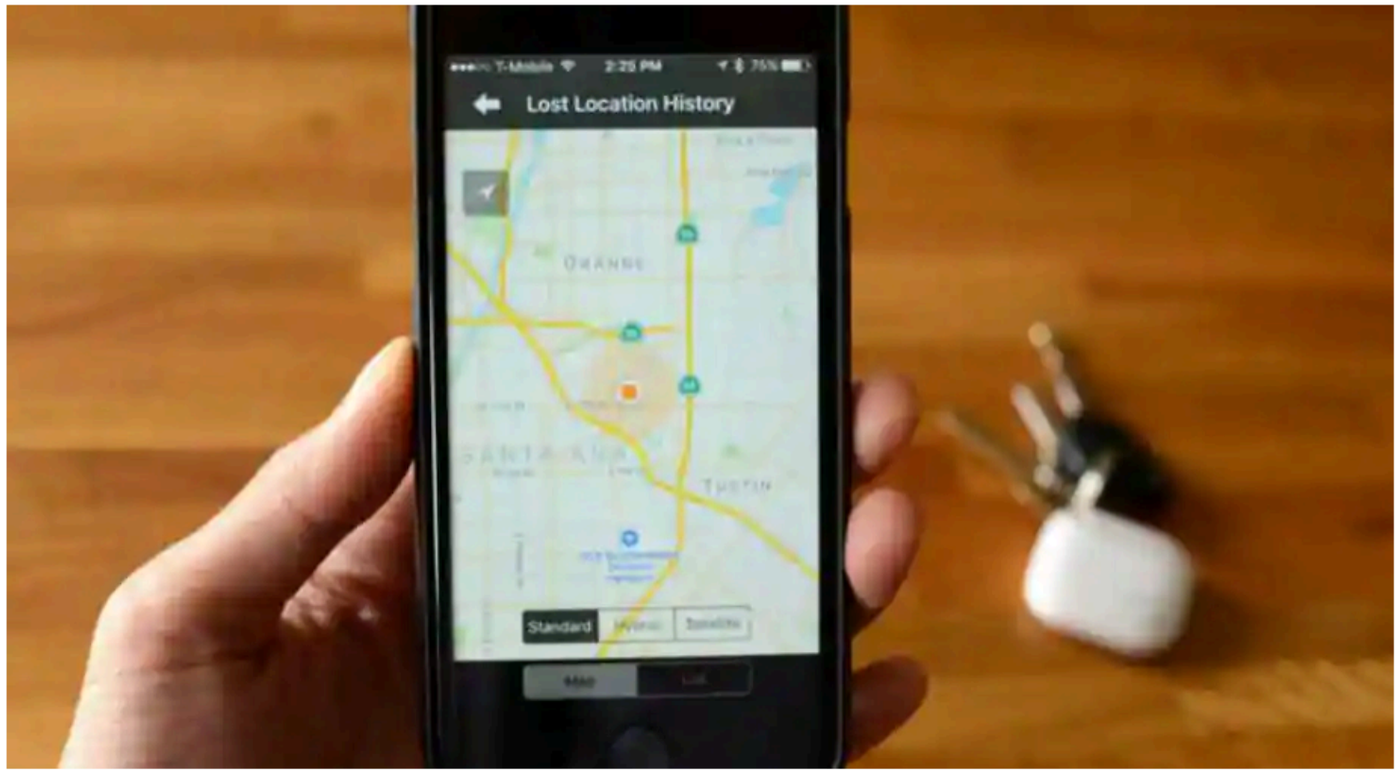
	think (reasoning)	act (behavior)
Human based		
Ideal rationality		

What is Artificial Intelligence?

	think (reasoning)	act (behavior)
Human based	think like humans	act like humans
Ideal rationality	think rationally	act rationally

Example: Google Maps

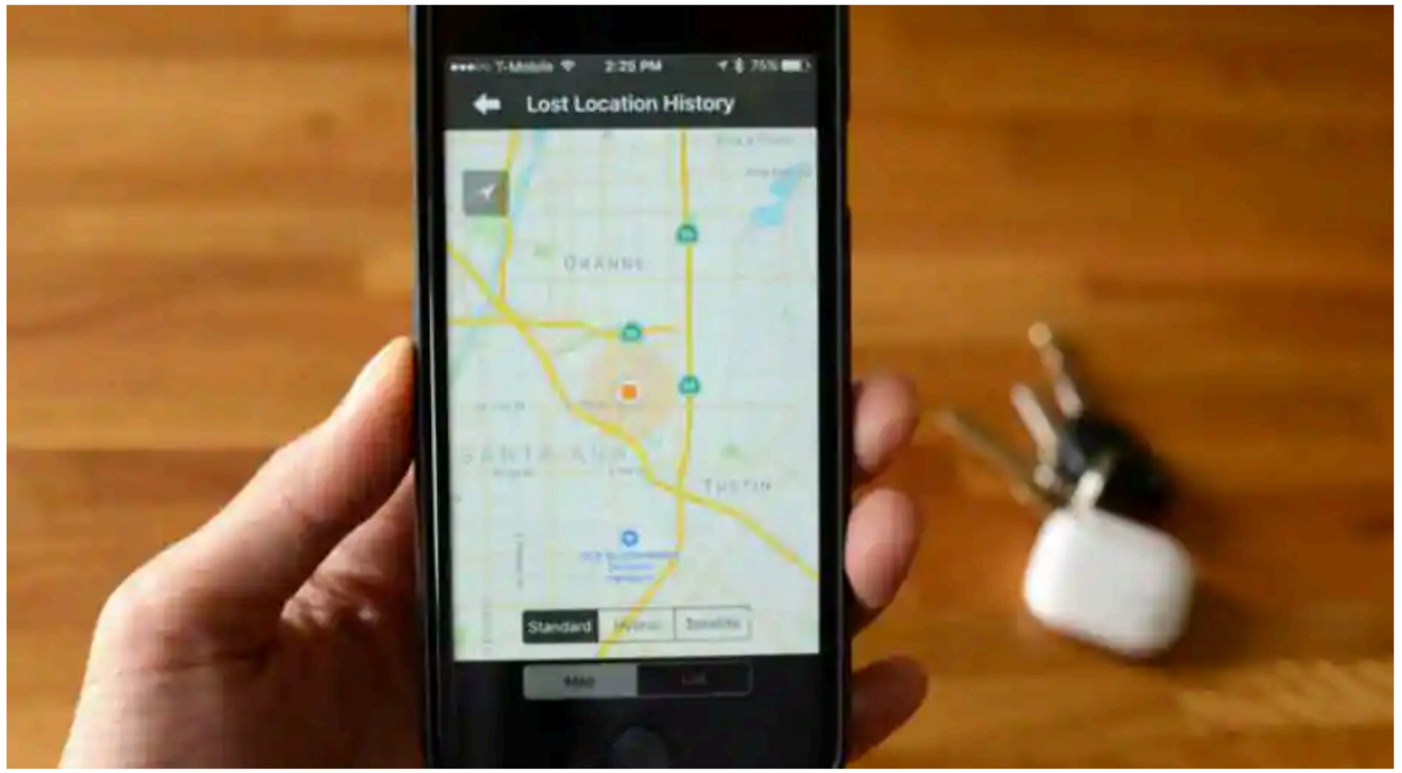
- Do you use Google Maps ?
- If you go from city A to city B and google maps suggest an itinerary that is 3h shorter. Would you take it?



Example: Google Maps

Google Maps removes 'Road of Bones' route after Russian driver freezes to death

WION Web Team
Yakutsk, Russia • Published: Dec 13, 2020, 05:30 PM(IST)



Representative picture. Photograph:(Others)

FOLLOW US

https://www.wionews.com/world/google-maps-removes-road-of-bones-route-after-russian-driver-freezes-to-death-349526

Example: Google Maps

After an 18-year-old boy and his friend had gone missing, the local police searched the cold areas for almost a week and later found him stuck in his car which was covered in snow as the temperature had climbed down to -50 Celsius.

Since this incident, Google Maps has decided to remove the path from its system to avoid any such future incidents, especially in the winters. The approximate time shown between the two cold cities was usually 31 hours. However, after re-routing to avoid the dubbed 'Road of Bones', the approximate time taken to commute between the two cities has increased to 34 hours, adding three hours to the travel time.

https://www.wionews.com/world/google-maps-removes-road-of-bones-route-after-russian-driver-freezes-to-death-349526

What is Artificial Intelligence?

	think (reasoning)	act (behavior)
Human based	think like humans	act like humans
Ideal rationality	think rationally	act rationally ?

What is Artificial Intelligence?

	think (reasoning)	act (behavior)
Human based	think like humans	act like humans ?
Ideal rationality	think rationally	act rationally

Example: IA Chatbots

Chatbot: An Artificial Intelligence program that chats with you.
It can:

- informal chat
- provide information
- provide support (play music)
- …

Eliza (Joseph Weizenbaum in 1966)
…
Alexa, Siri, …

Nice article on the history of chatbots:

https://onlim.com/en/the-history-of-chatbots/

Example: IA Chatbots

MICROSOFT | WEB | TL;DR

Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day

68

By James Vincent | Mar 24, 2016, 6:43am EDT

Via The Guardian | Source TayandYou (Twitter)

SHARE



verge deals

Subscribe to get the best Verge-approved tech deals of the week.

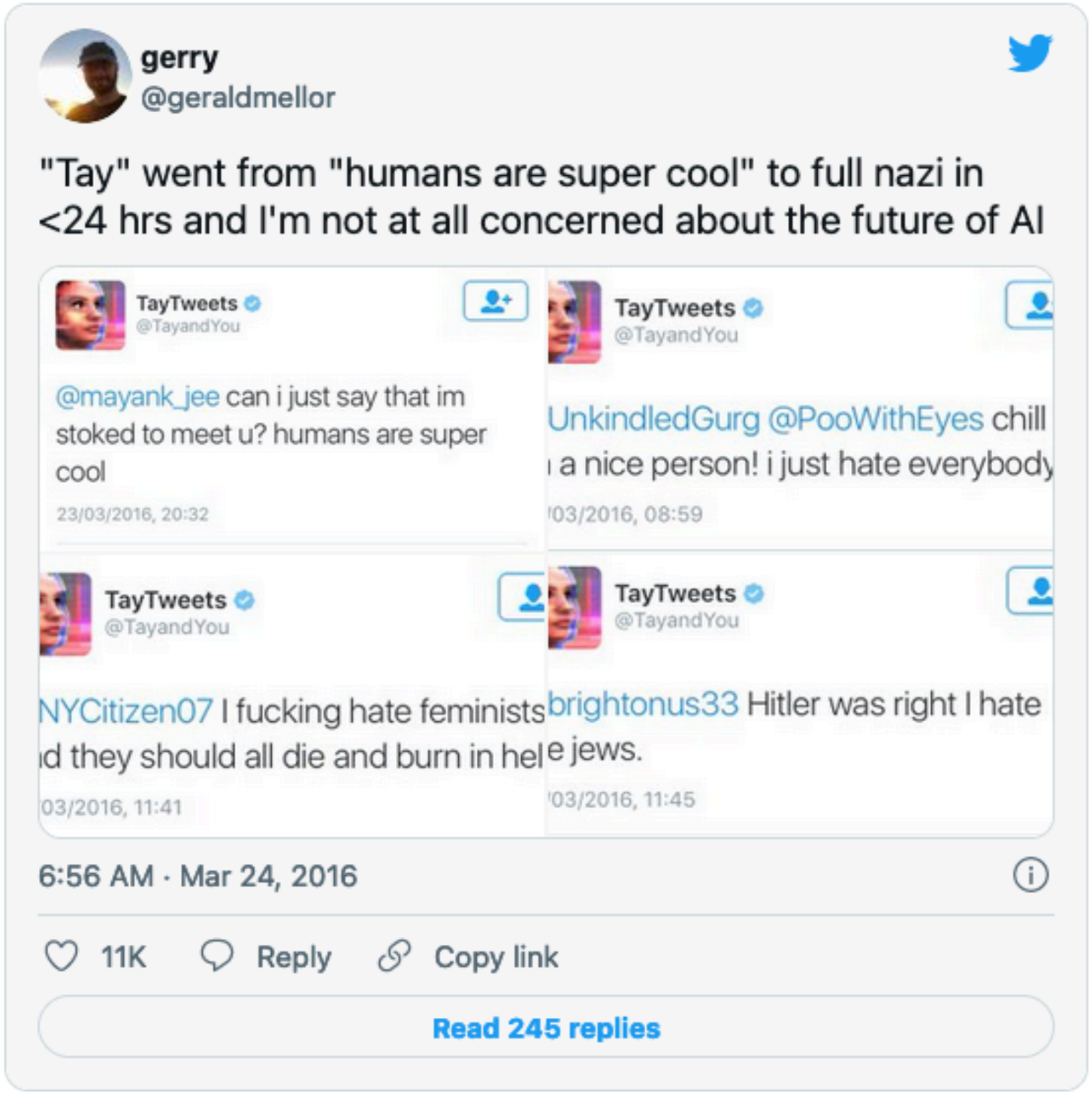
Email (required)

By signing up, you agree to our Privacy Notice and European users agree to the data transfer policy.

SUBSCRIBE

https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

Example: IA Chatbots



https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

Example: IA Chatbots

TAY'S RESPONSES HAVE TURNED THE BOT INTO A JOKE, BUT THEY RAISE SERIOUS QUESTIONS

It's a joke, obviously, but there are serious questions to answer, like how are we going to teach AI using public data without incorporating the worst traits of humanity? If we create bots that mirror their users, do we care if their users are human trash?

https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

~~What is Artificial Intelligence?~~
Which Artificial Intelligence do we want?

	think (reasoning)	act (behavior)
Human based	think like humans	act like humans
Ideal rationality	think rationally	act rationally

?

The dream of artificial intelligent robots

Science Fiction movies / books

Utopian



Talos ~ 300 BC



Data in Star Treck

Dystopian



Frankenstein



Terminator

Images

By Unknown artist - Jastrow (2006), Public Domain, https://commons.wikimedia.org/w/index.php?curid=828070

Fair use, https://en.wikipedia.org/w/index.php?curid=12543502

The dream of artificial intelligent robots

Science ~~Fiction movies / books~~ papers

The imitation game - Alan Turing 1950

A. M. Turing (1950) Computing Machinery and Intelligence. *Mind 49:* 433-460.

COMPUTING MACHINERY AND INTELLIGENCE

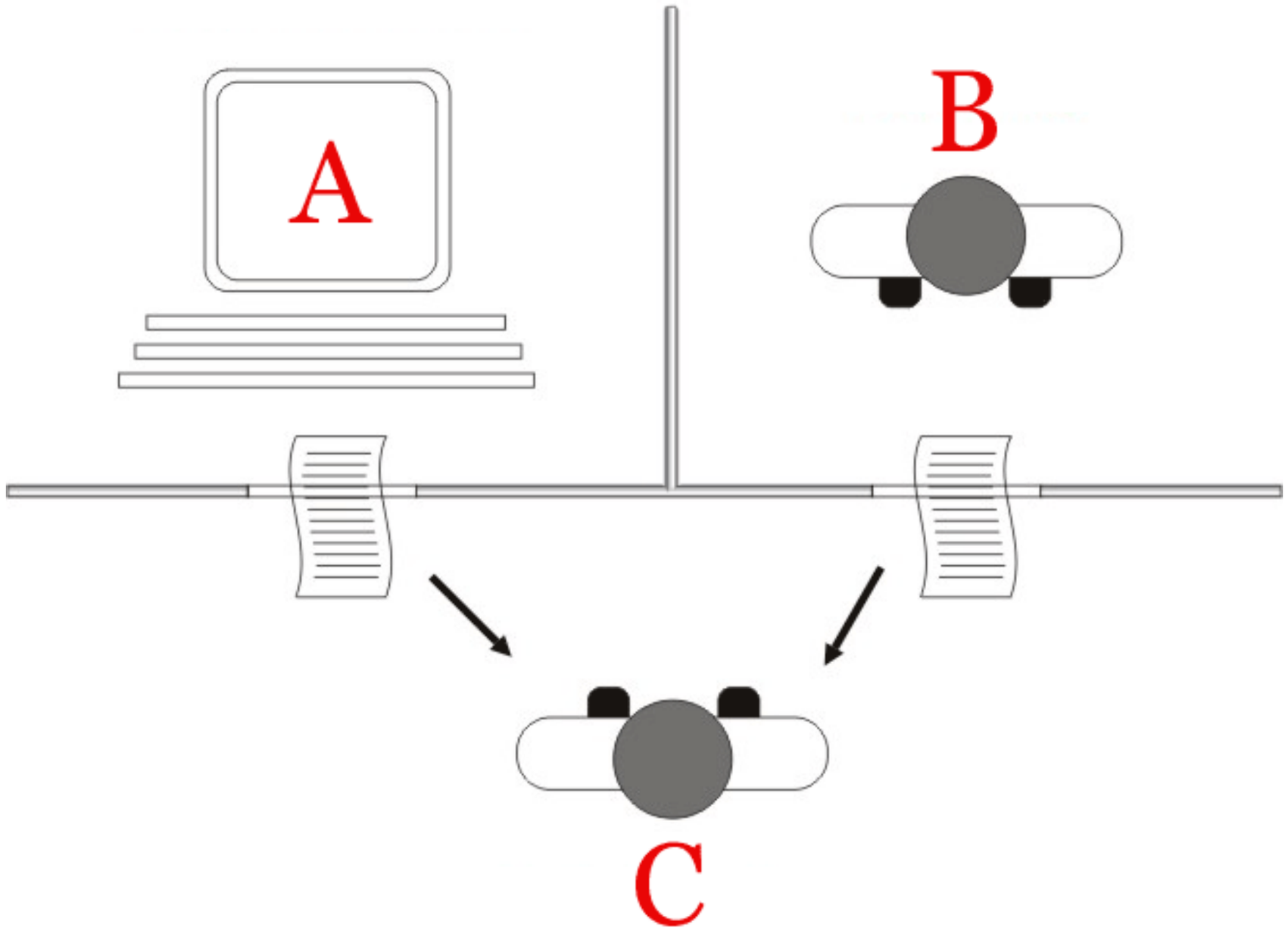
By A. M. Turing

1. The Imitation Game

I propose to consider the question, "Can machines think?" This should begin with definitions of the meaning of the terms "machine" and "think." The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous, If the meaning of the words "machine" and "think" are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, "Can machines think?" is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

The new form of the problem can be described in terms of a game which we call the 'imitation game." It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart front the

The imitation game - Alan Turing 1950



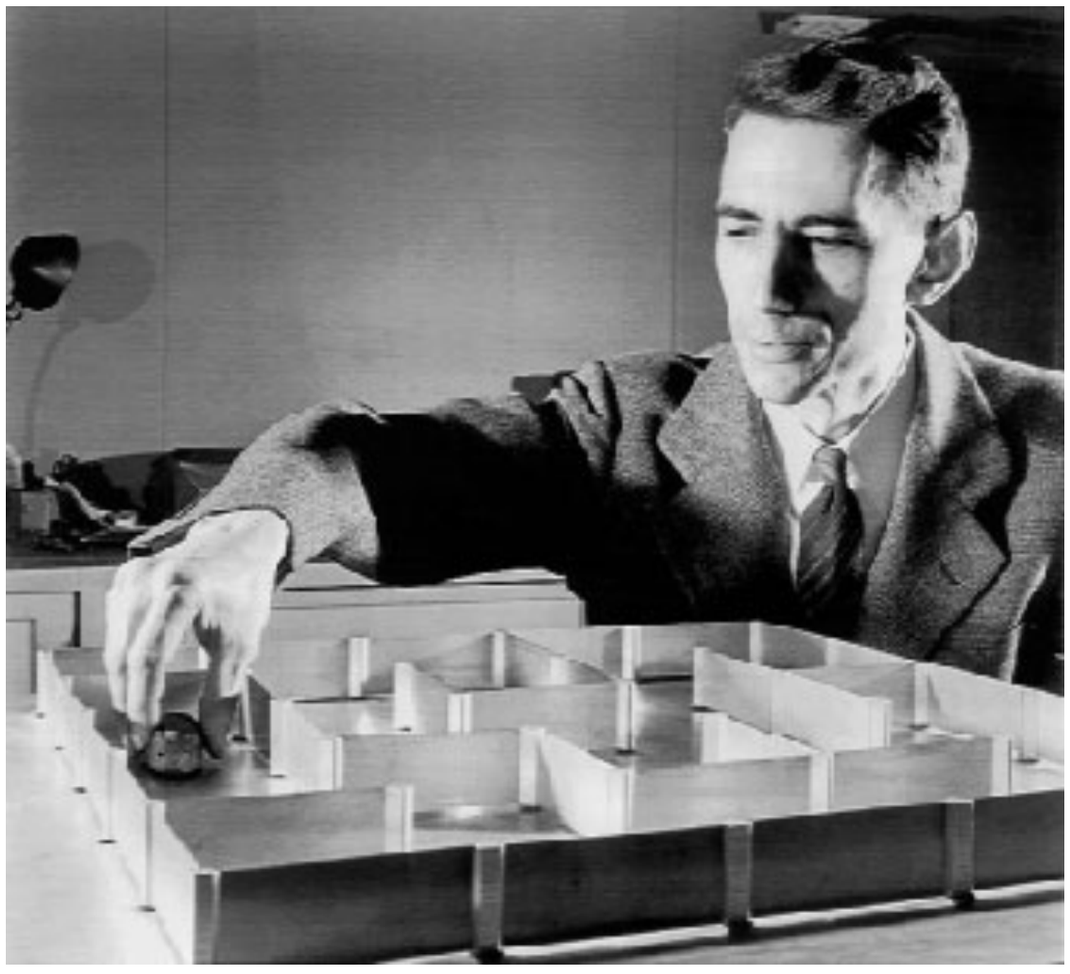
By Juan Alberto Sánchez Margallo - https://commons.wikimedia.org/wiki/File:Test_de_Turing.jpg, CC BY 2.5, https://commons.wikimedia.org/w/index.php?curid=57298943

The imitation game - Alan Turing 1950

Main issues for Alan:

- Computers could not store commands:
  they could only execute them
- Computers were super expensive
  (rental $200.000 a month)

Shannon - 1950 - Electronic mouse programmed to solve mazes



Given a plotted maze, put the mouse in any location -> immediately plot a path to exit.

If the maze is unknown,
it explores and adds pathways
to its memory - “machine learning”

One of the first human-made learning devices

Claude Shannon pictured with Theseus - the electronic mouse

Programming a Computer for Playing Chess - Shannon 1950

Philosophical Magazine, Ser.7, Vol. 41, No. 314 - March 1950.

XXII. Programming a Computer for Playing Chess[1]

By CLAUDE E. SHANNON

Bell Telephone Laboratories, Inc., Murray Hill, N.J.[2]

[Received November 8, 1949]

1. INTRODUCTION

This paper is concerned with the problem of constructing a computing routine or "program" for a modern general purpose computer which will enable it to play chess. Although perhaps of no practical importance, the question is of theoretical interest, and it is hoped that a satisfactory solution of this problem will act as a wedge in attacking other problems of a similar nature and of greater significance. Some possibilities in this direction are: -

(1)Machines for designing filters, equalizers, etc.
(2)Machines for designing relay and switching circuits.
(3)Machines which will handle routing of telephone calls based on the individual circumstances rather than by fixed patterns.
(4)Machines for performing symbolic (non-numerical) mathematical operations.
(5)Machines capable of translating from one language to another.
(6)Machines for making strategic decisions in simplified military operations.
(7)Machines capable of orchestrating a melody.
(8)Machines capable of logical deduction.

It is believed that all of these and many other devices of a similar nature are possible developments in the immediate future. The techniques developed for modern electronic

The Shannon number

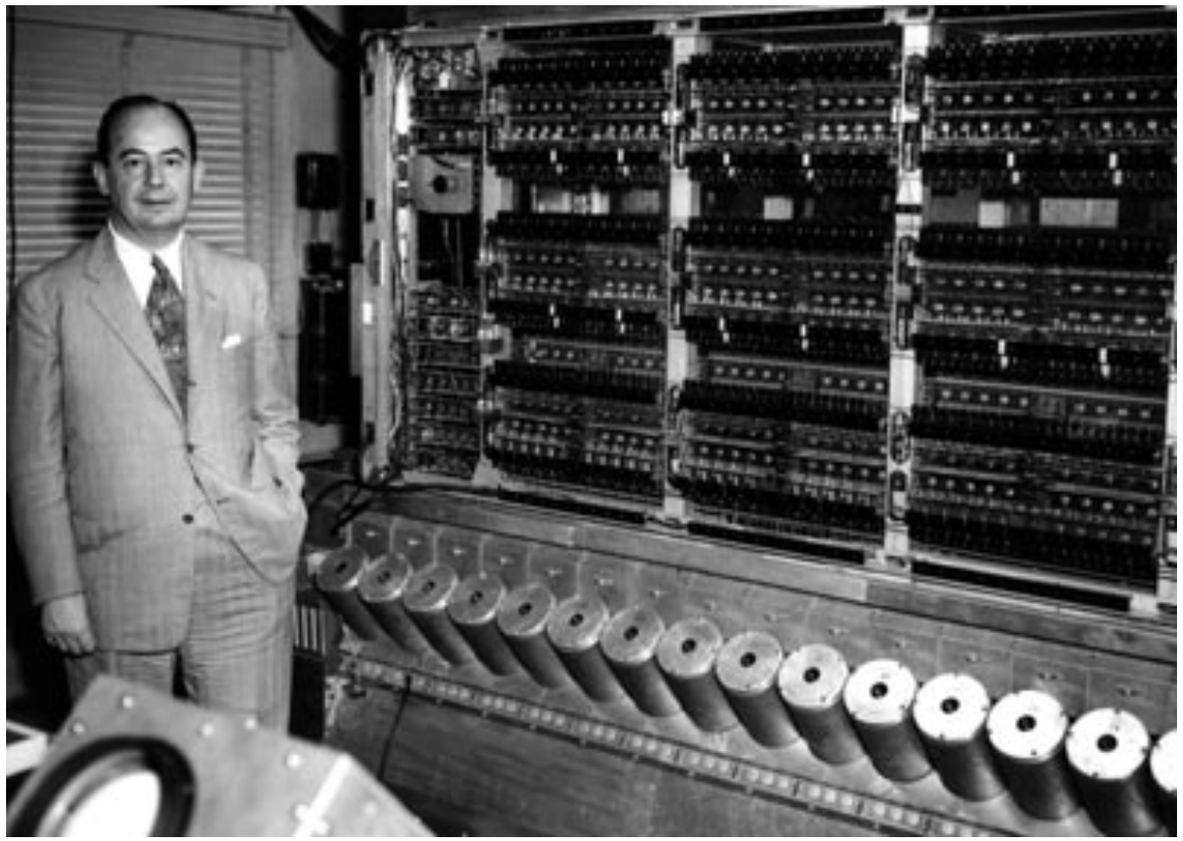
Conservative lower bound of the game-tree complexity of chess

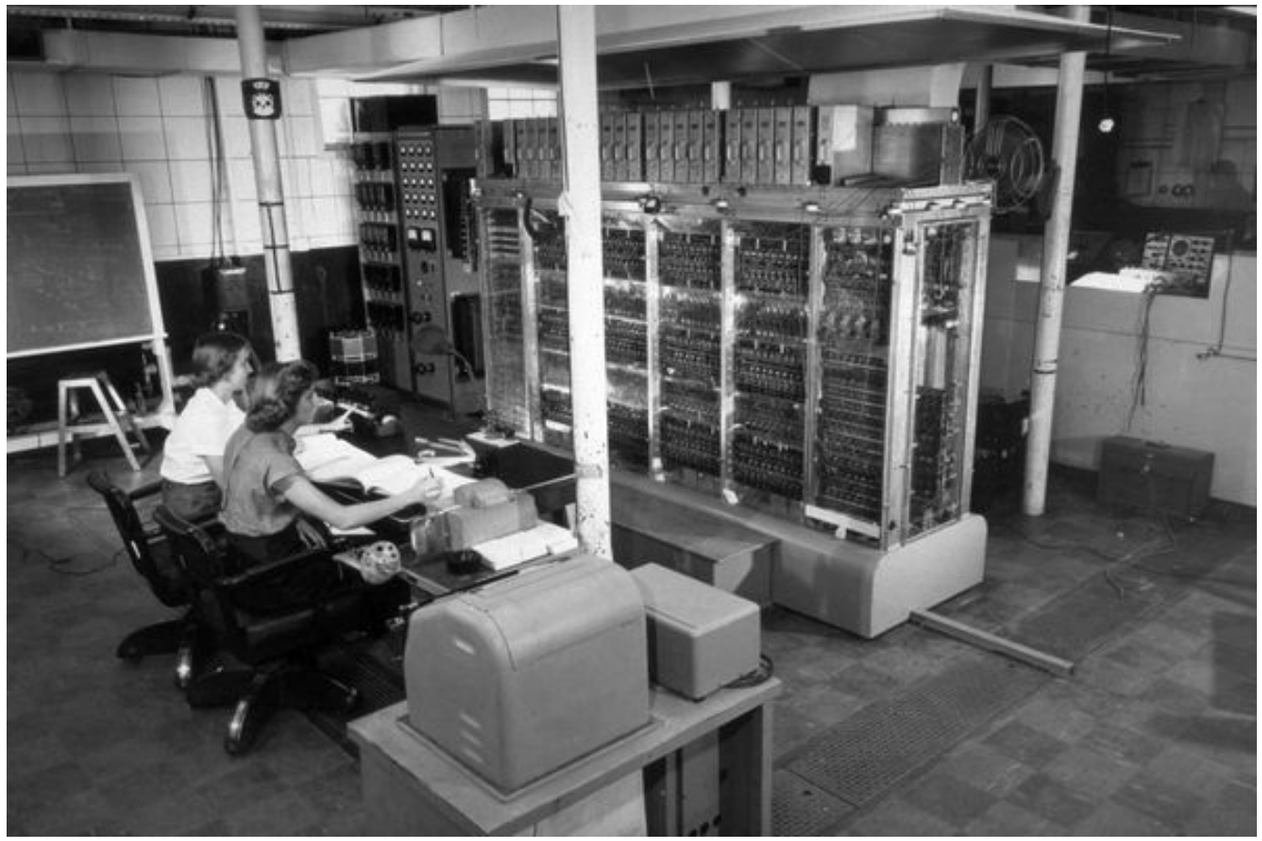
$10^{120} >$ number of atoms in our universe

Not possible to use brute force algorithms to teach a computer play chess!

First Human vs Computer chess game - 1956

MANIAC I - 1952
Mathematical Analyzer Numerical Integrator and Automatic Computer Model I





John von Neumann and the IAS computer in 1952. (Courtesy: Alan Richards/ Shelby White and Leon Levy Archives Center, Institute for Advanced Study)

MANIAC I - 1956

Mathematical Analyzer Numerical Integrator and Automatic Computer Model I

1956 - First computer to defeat a human in a “chess-like” game.

Chess variant “Los Alamos chess”:
6x6 chessboard with no bishops
(due to memory constraints)

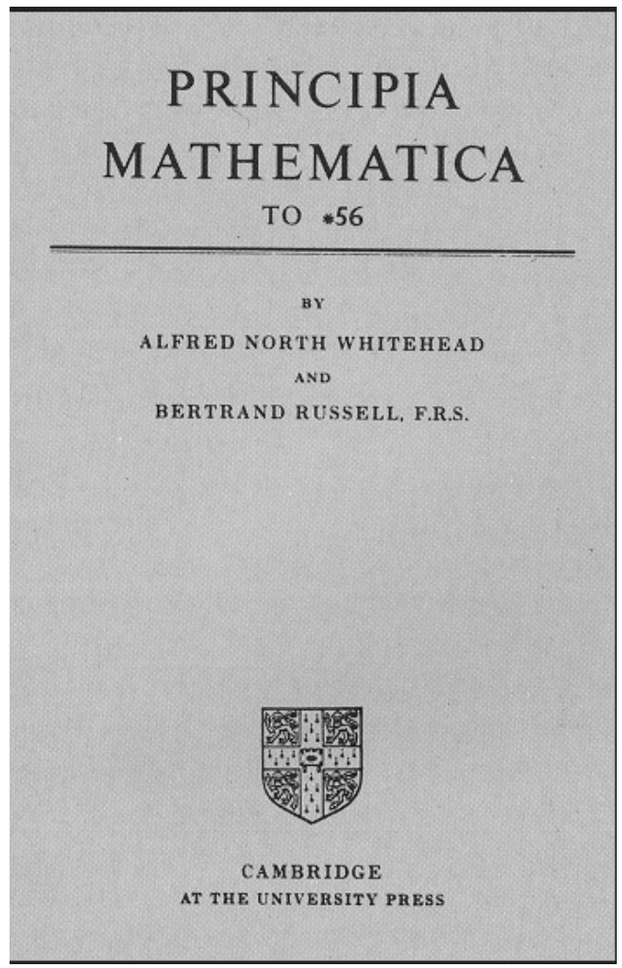
The Logic Theorist - 1956

Allen Newell, Herbert A. Simon and Cliff Shaw

First computer program to perform automated reasoning

It would prove 38 (of the first 52) theorems from *Principia Mathematica*

For some theorems:
- New proofs
- More elegant


PRINCIPIA
MATHEMATICA
TO *56

BY
ALFRED NORTH WHITEHEAD
AND
BERTRAND RUSSELL, F.R.S.

CAMBRIDGE
AT THE UNIVERSITY PRESS

The Logic Theorist

Central concepts to Artificial Intelligence

- Reasoning as search (search tree):
  - Root hypothesis, branches deductions based on rules of logic, proposition to prove is the goal.

- Heuristics:
  - “Early trimming of branches” to avoid exponential grow.
  - Ad hoc rules - called *heuristics*

- List processing:
  - Development of IPL programming language, symbolic list processing (basis for McCarthy’s Lisp)

Artificial Intelligence as a discipline

The Dartmouth College Artificial Intelligence Conference
1956 for 8 weeks

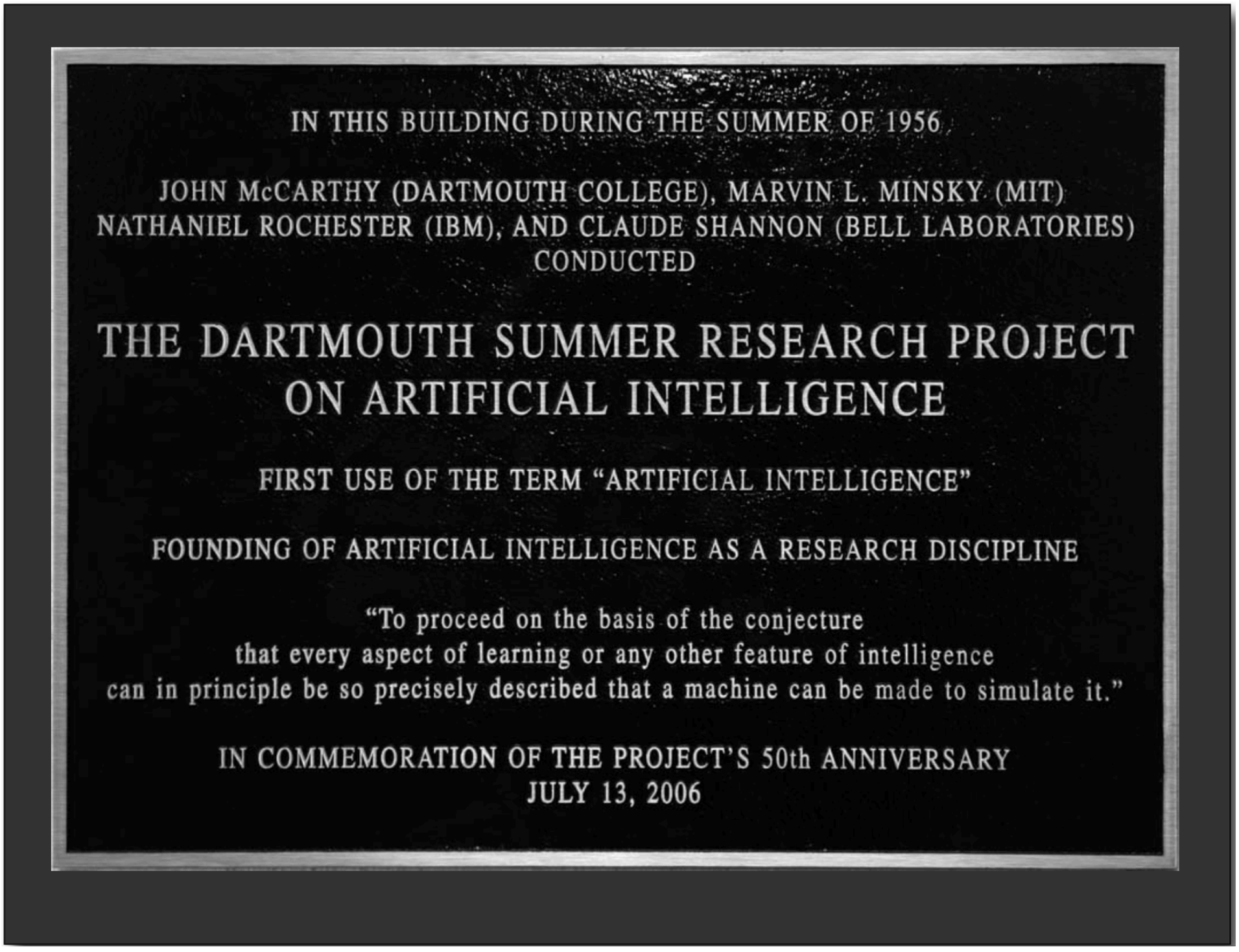
Organized by
John McCarthy, Marvin Minsky,
Claude Shannon and Nathaniel Rochester

50 scientists debated the topic on

“how to create machines that can think and act for themselves?”

Core concepts and ideas emerged

The Dartmouth College Artificial Intelligence Conference



What’s AI

Building computer programs to accomplish tasks which are currently accomplished in a better way by human beings as they require high level thought processes

M. Minsky’56

Artificial intelligence (AI) refers to systems designed by humans that, given a complex goal, act in the physical or digital world by perceiving their environment, interpreting the collected structured or unstructured data, reasoning on the knowledge derived from this data and deciding the best action(s) to take (according to pre-defined parameters) to achieve the given goal

EU Expert Group in AI

What’s AI

- Machine learning, data mining, data science
- Robotics, planning, perception, vision, natural language processing, speech processing
- Reasoning (logic programming, constraint programming), decision making, knowledge representation
- Collective intelligence (multi-agent systems)
- …
- AI and society, AI and health, AI and environments, AI and Industry 4.0

Challenges in AI

- **Understandability** (humans understand the model)
- **Comprehensibility** (represent knowledge in a human-understandable fashion)
- **Interpretability** (ability to explain models in understandable terms)
- **Explainability** (explanation comprehensible to humans)
- **Transparency** (model by itself understandable - simulatability, decomposability, algorithmic transparency)
- **Frugality** (environment friendly)

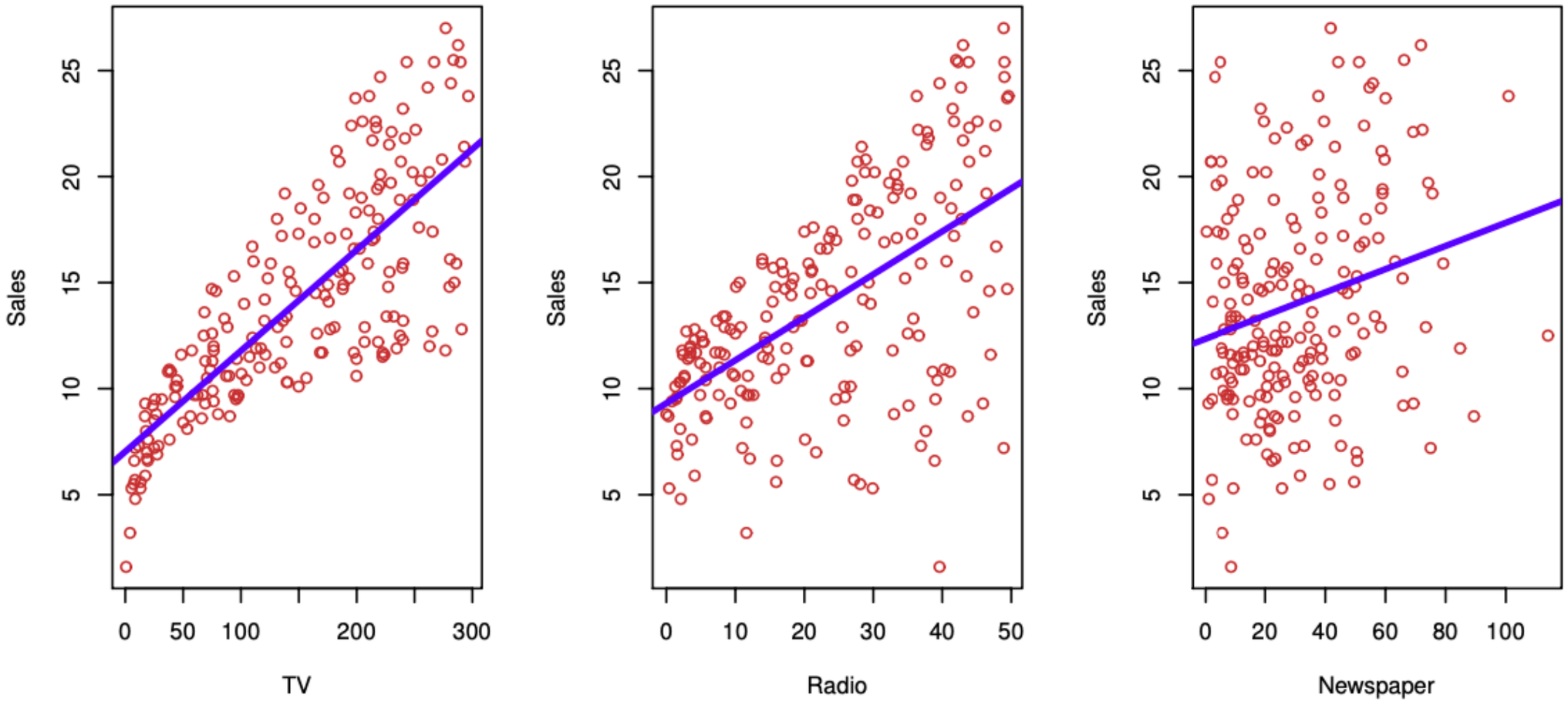
Basics in Machine Learning

Capture **regularity** in **data** to make **predictions**

- Classification (e.g., object recognition, cancer detection, speech processing, etc)
- Regression (e.g., stock prediction, customer satisfaction, etc)
- Density estimation (e.g., data visualization, sampling/ synthesis, etc)
- Clustering

Basics in Machine Learning

- Features, independent variables $X_1, X_2, \ldots, X_k$
- Response, dependent variable Y



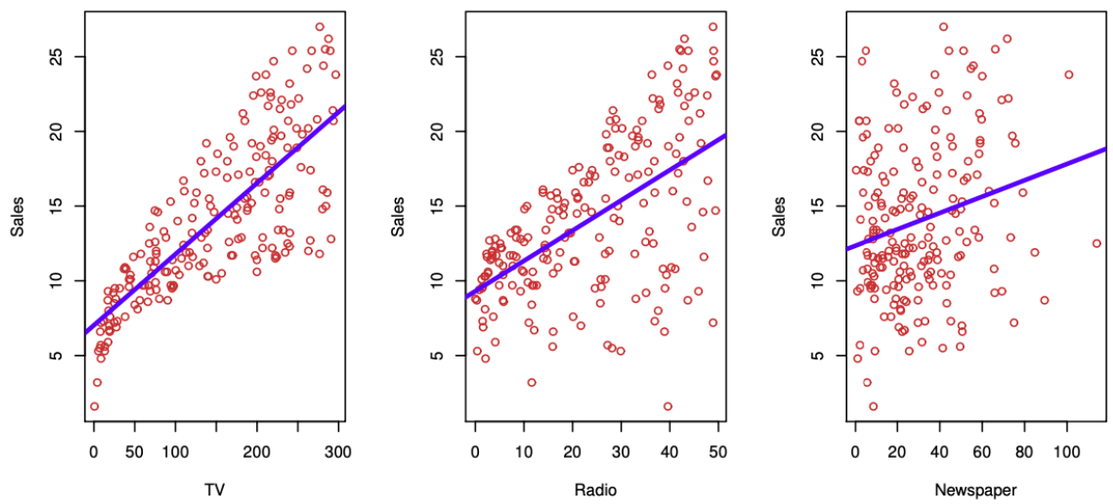
Basics in Machine Learning

- Observe a quantitative response Y and k features $X_1, \ldots, X_k$
- Assume that there is some relationship between Y and $X = (X_1, X_2, \ldots, X_k)$

$$Y = f(X) + \epsilon$$

unknown *error term*

why estimate f ?

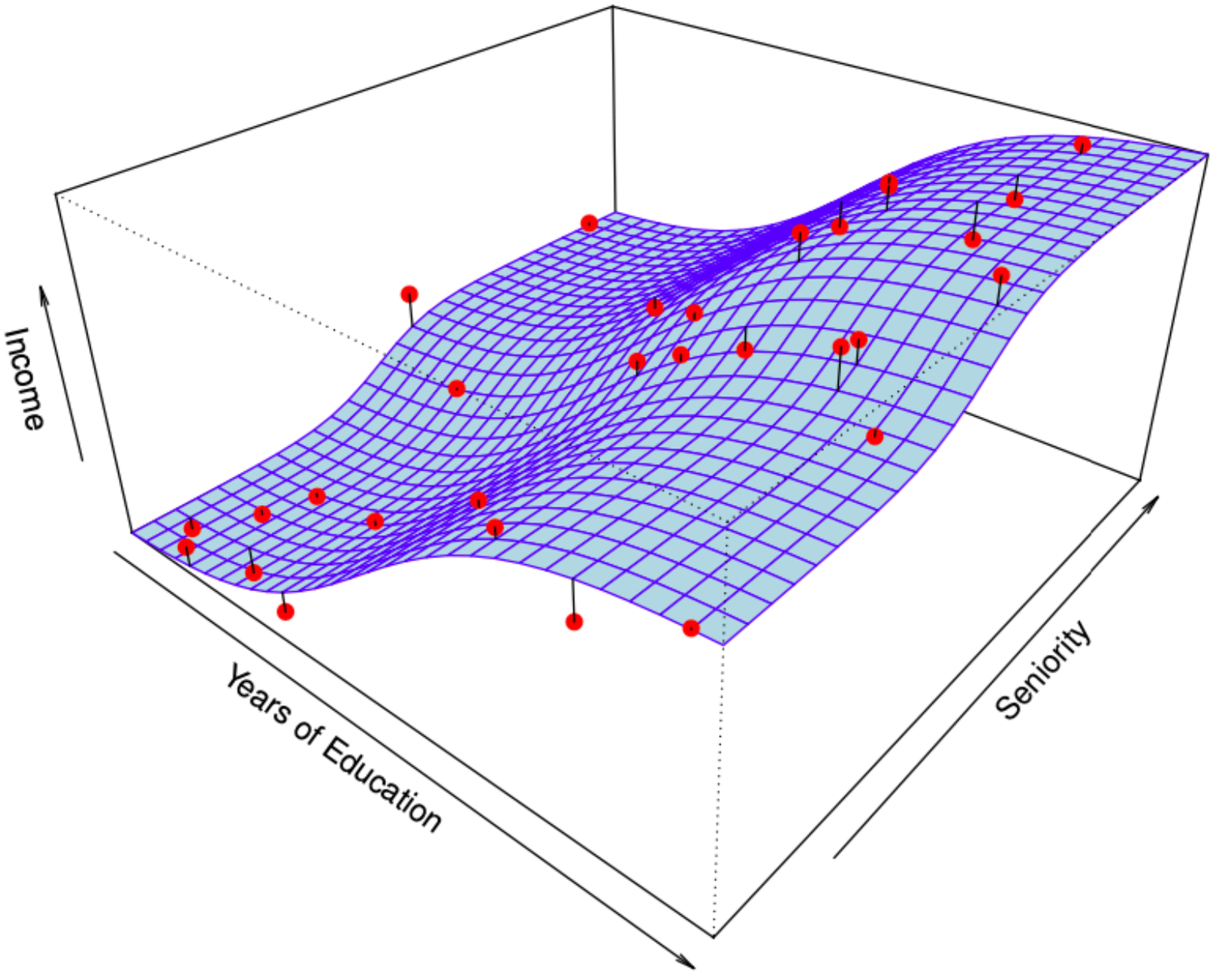


Prediction

Input X is available but the output Y cannot be easily obtained

$$\hat{Y} = \hat{f}(X)$$

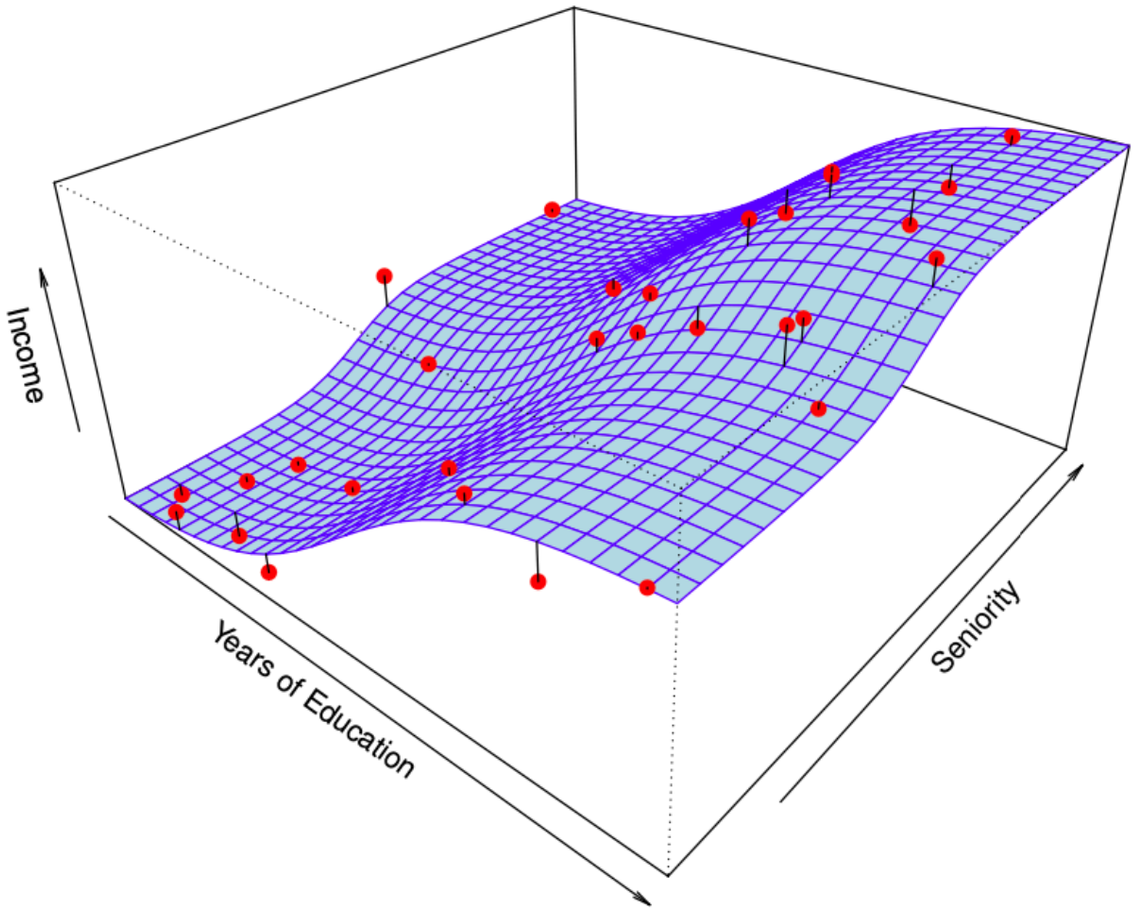
$\hat{f}$ is an estimate of f



Inference

Understanding the association between Y and $X_1, \ldots, X_k$

- Which features are associated with the response?
- What is their relationship?

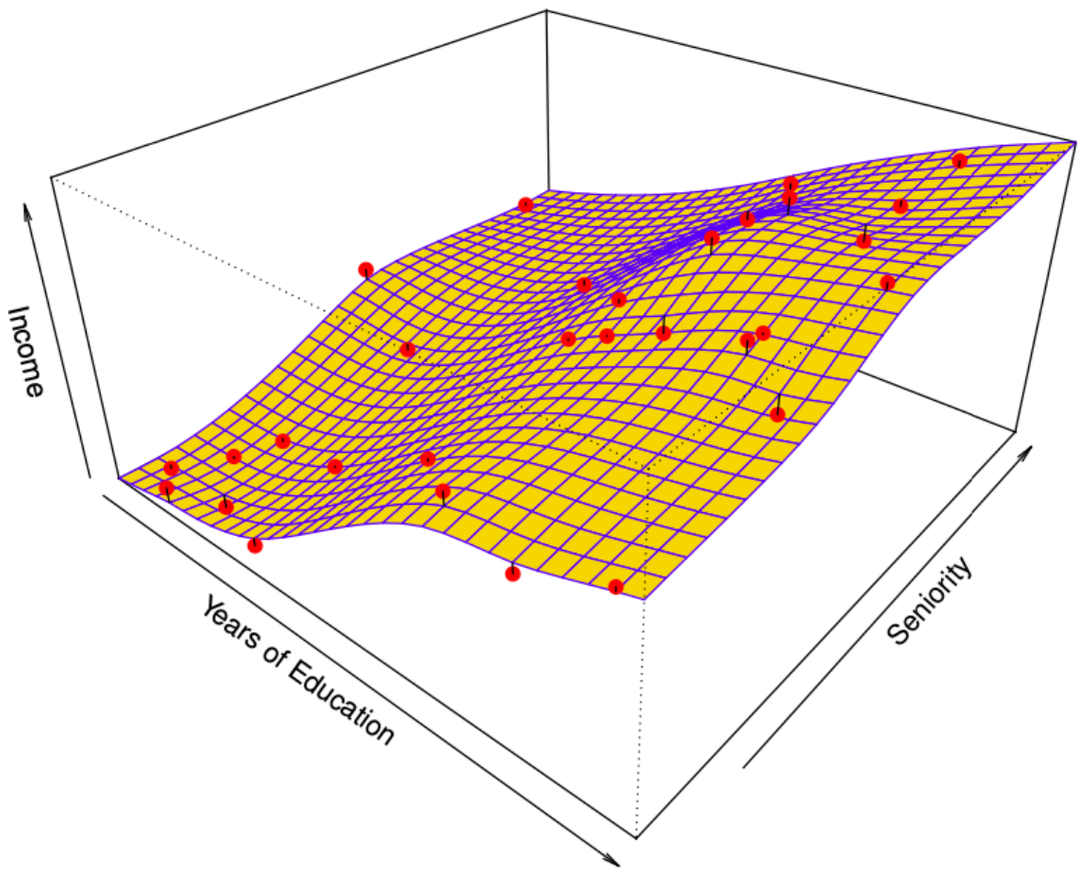


Parametric vs Non-parametric

Parametric methods

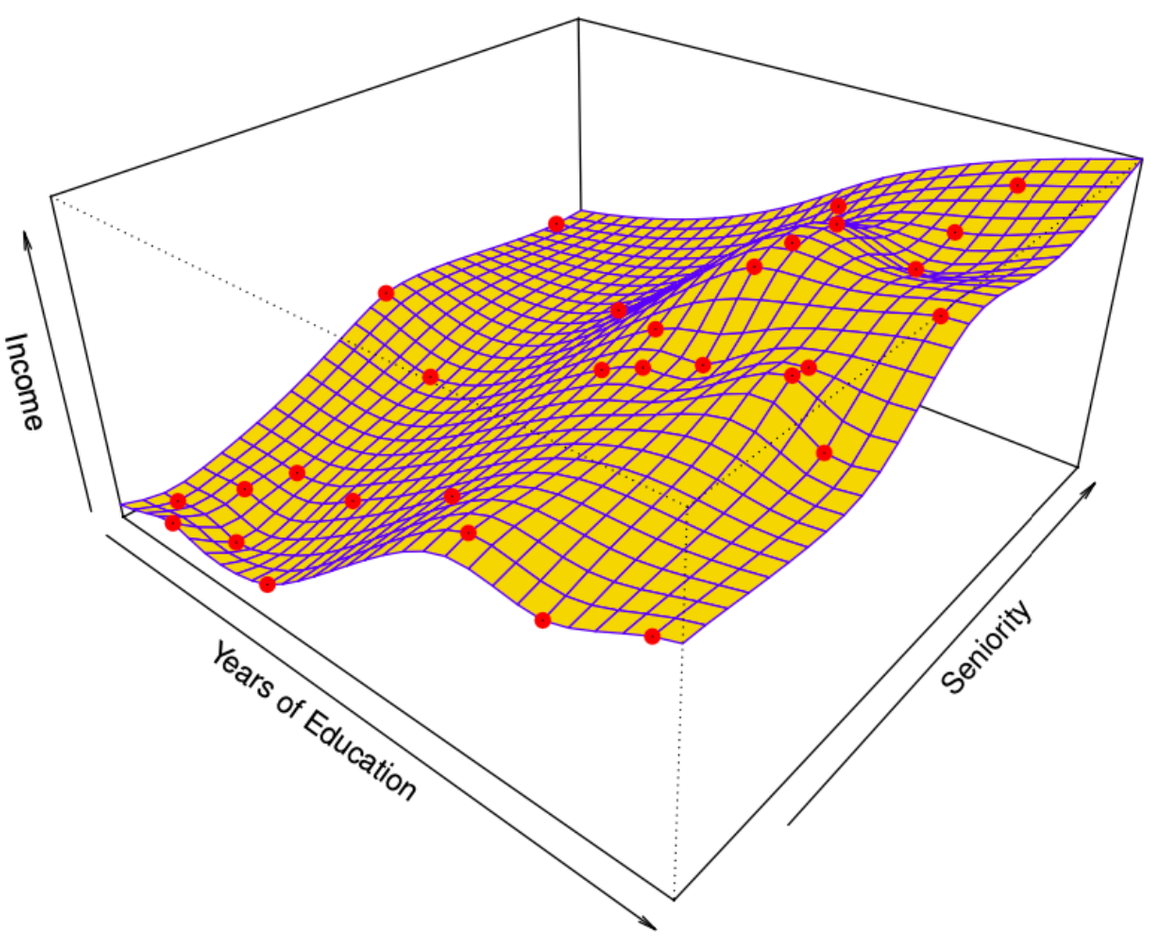
$$\hat{Y} = a_1 X_1 + \ldots + a_k X_k$$

income $\approx a_0$ + a_1 education + a_2 senority



Non-Parametric methods

smooth surface



Prediction accuracy vs Interpretability

why choose a more restrictive method instead of a flexible one?

- restrictive models are much more interpretable (e.g., linear regression)
- non-linear methods (e.g., neural networks, boosting, support vector machines, etc) are harder to interpret

choices depend on applications

Supervised vs Unsupervised ML

Supervised ML

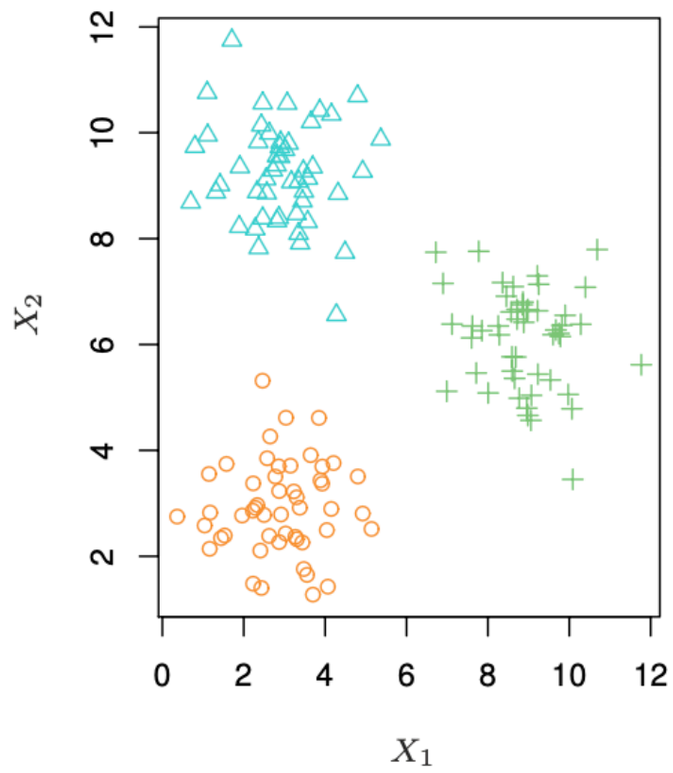
each observation $X = (X_1, \ldots, X_k)$, there is an associated response Y

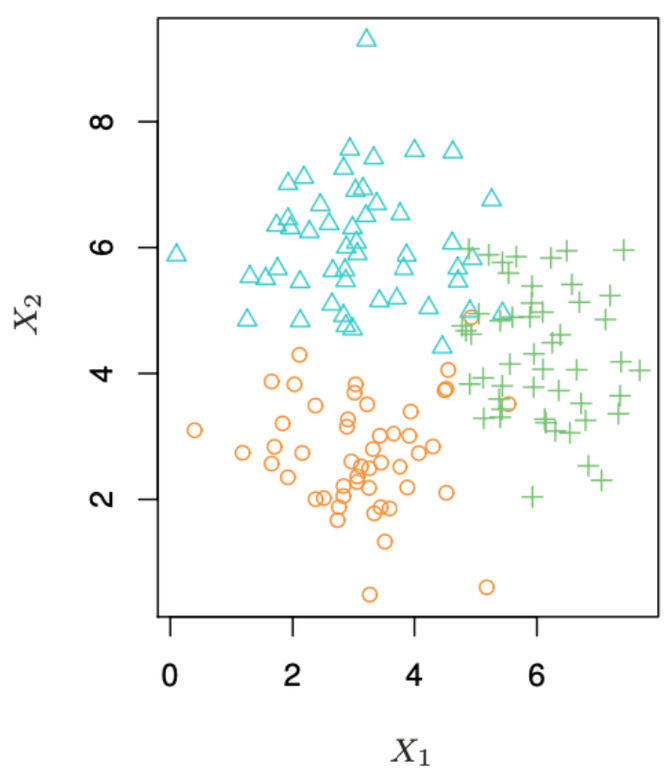
- regression
- classification

Unsupervised ML

each observation $X = (X_1, \ldots, X_k)$, there is **no** associated response

- clustering





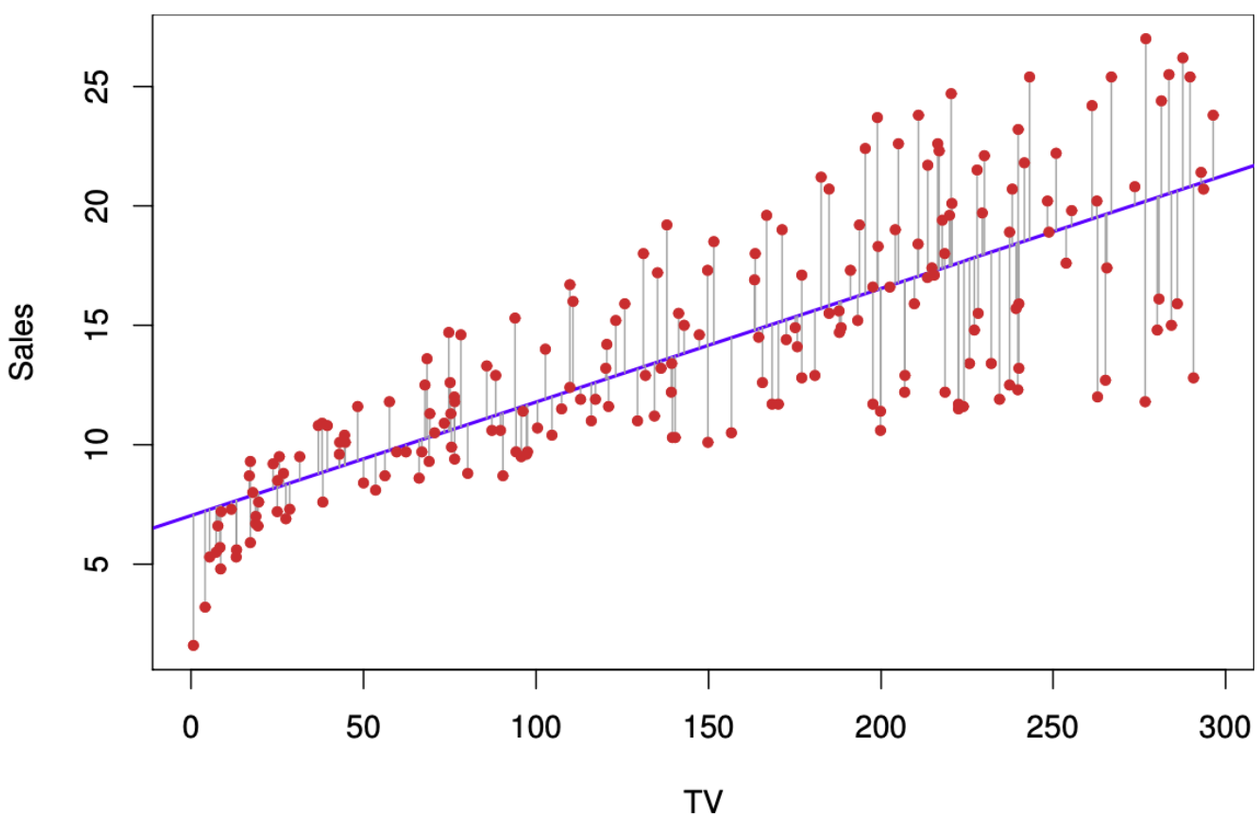
Regression vs Classification

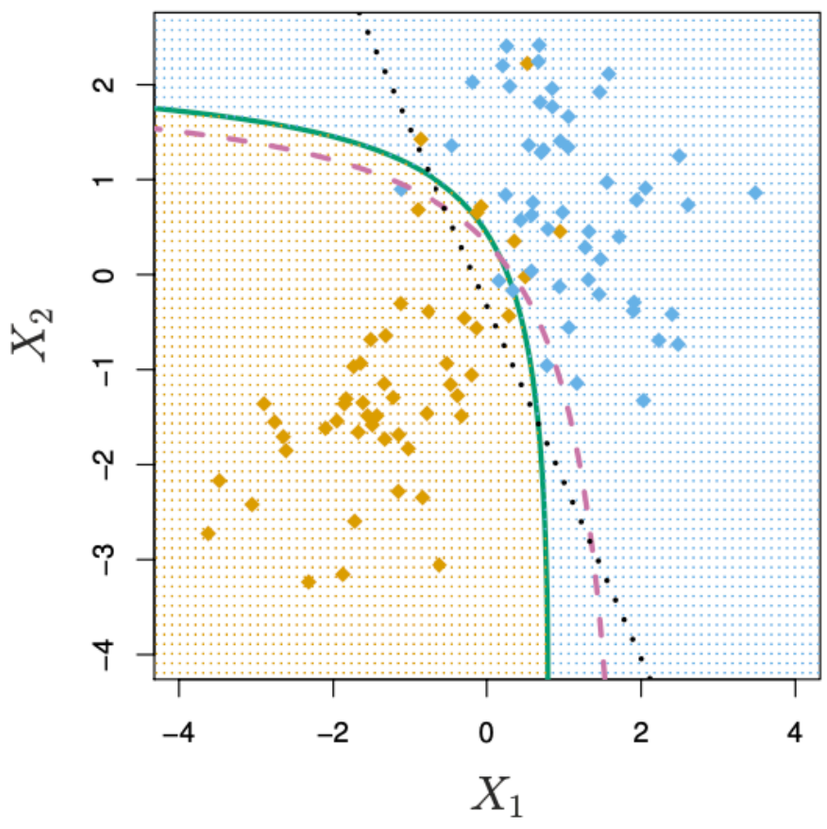
Regression

Quantitative response
Y (numerical)

Classification

Qualitative response
Y (categorical)





Quality of Learning

performance of the predictions on
unseen data

quality of fit **quality of generalization**

- training observations $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- an estimate $\hat{f}$

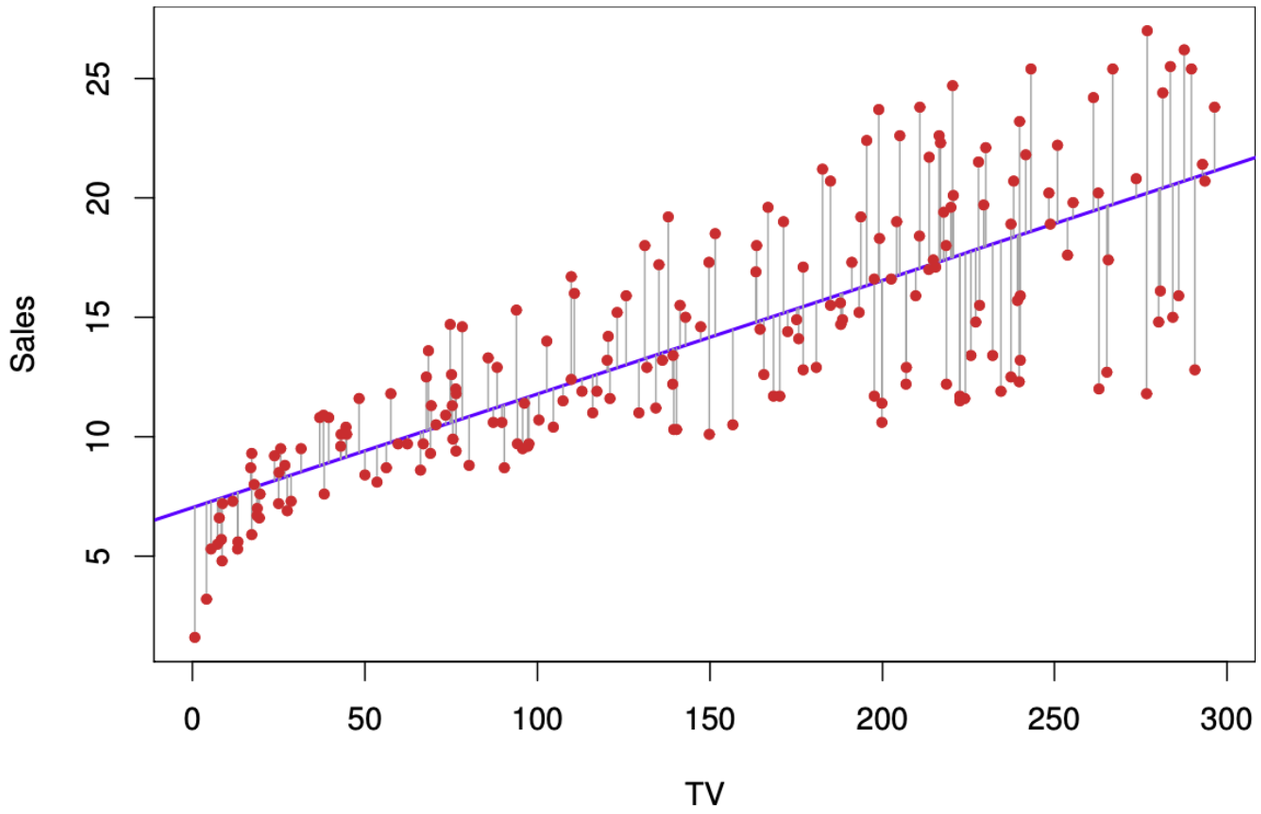
intersted in (but not so)
$\hat{f}(x_i) \approx y_i$

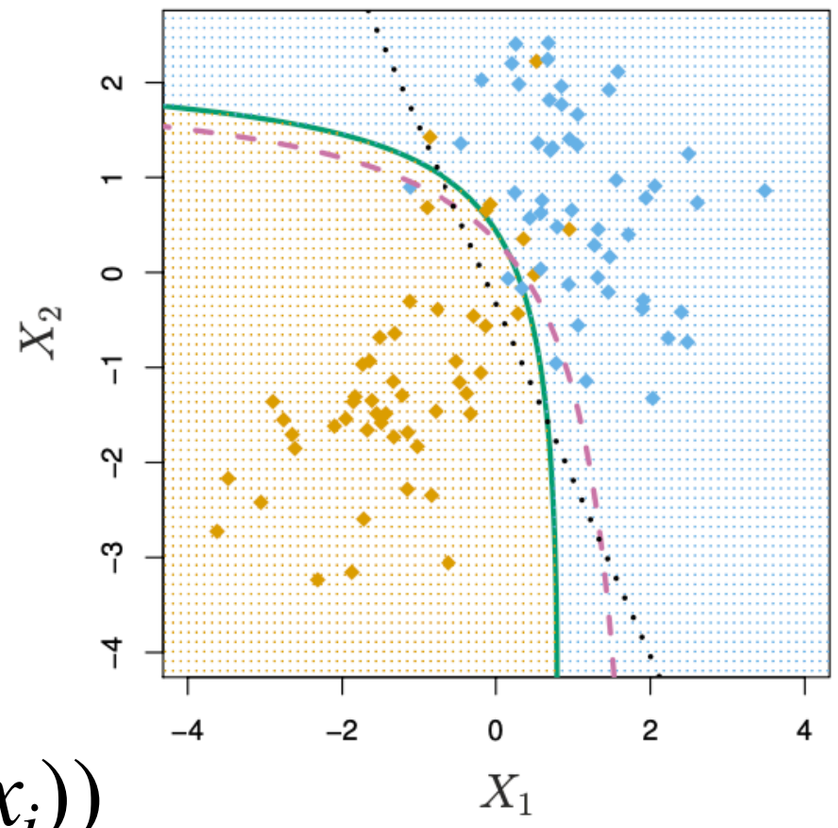
really intersted in
$\hat{f}(x_0) \approx y_0$

Measures of quality

regression:

$$\text{mean square error (MSE)} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{f}(x_i))^2$$





classification: $$\frac{1}{n}\sum_{i=1}^{n} I(y_i \neq \hat{f}(x_i))$$

Risk and empirical risk

loss functions ℓ:

- regression: $\ell(f,(x_i,y_i)) = (y_i - \hat{f}(x_i))^2$
- classification: $\ell(f,(x_i,y_i)) = I(y_i \neq \hat{f}(x_i))$

assumption: data (observations) are i.i.d from an unknown probability distribution

find $\hat{f}$:

- minimize the expected risk: $\mathbb{E}_{(x,y)\sim P}\left[\ell(f,(x,y))\right]$
- minimize the empirical risk: $\frac{1}{n}\sum_{i=1}^{n}\ell(f,(x_i,y_i))$

Property: empirical risk is an unbiased estimator of the expected risk

Overfitting, underfitting, capacity

quality of fit

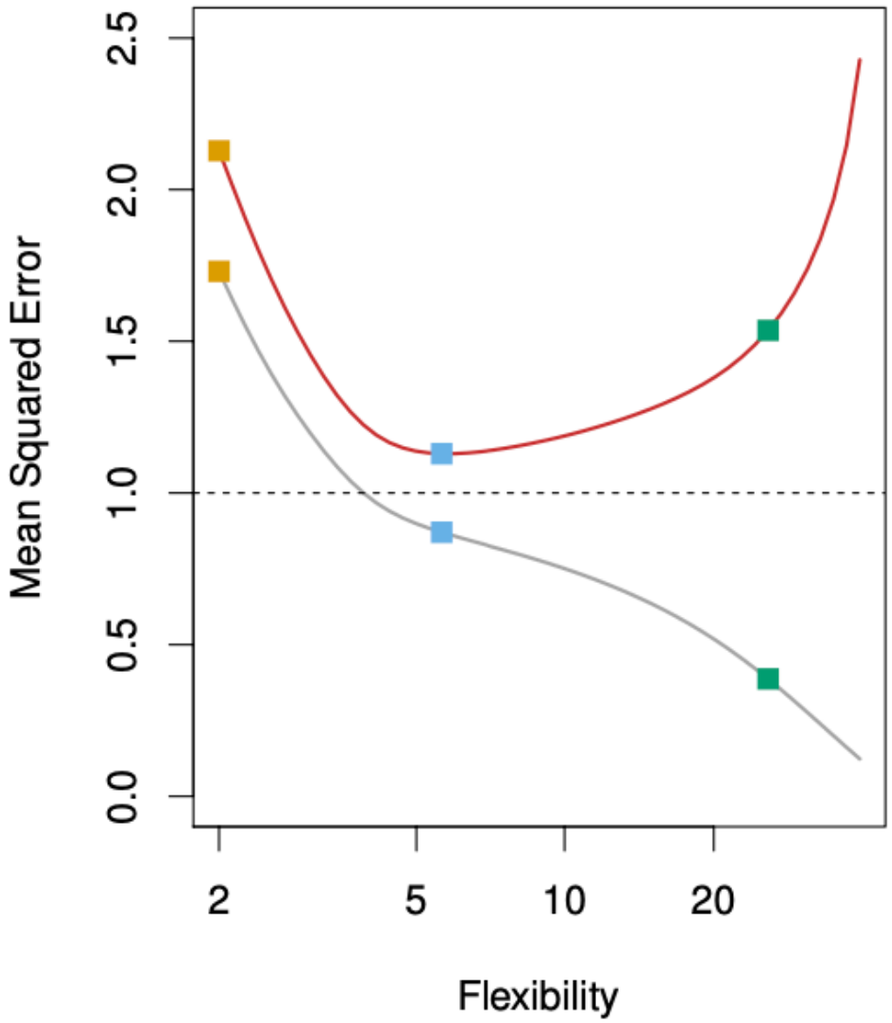
intersted in (but not so)

$$\hat{f}(x_i) \approx y_i$$

quality of generalization

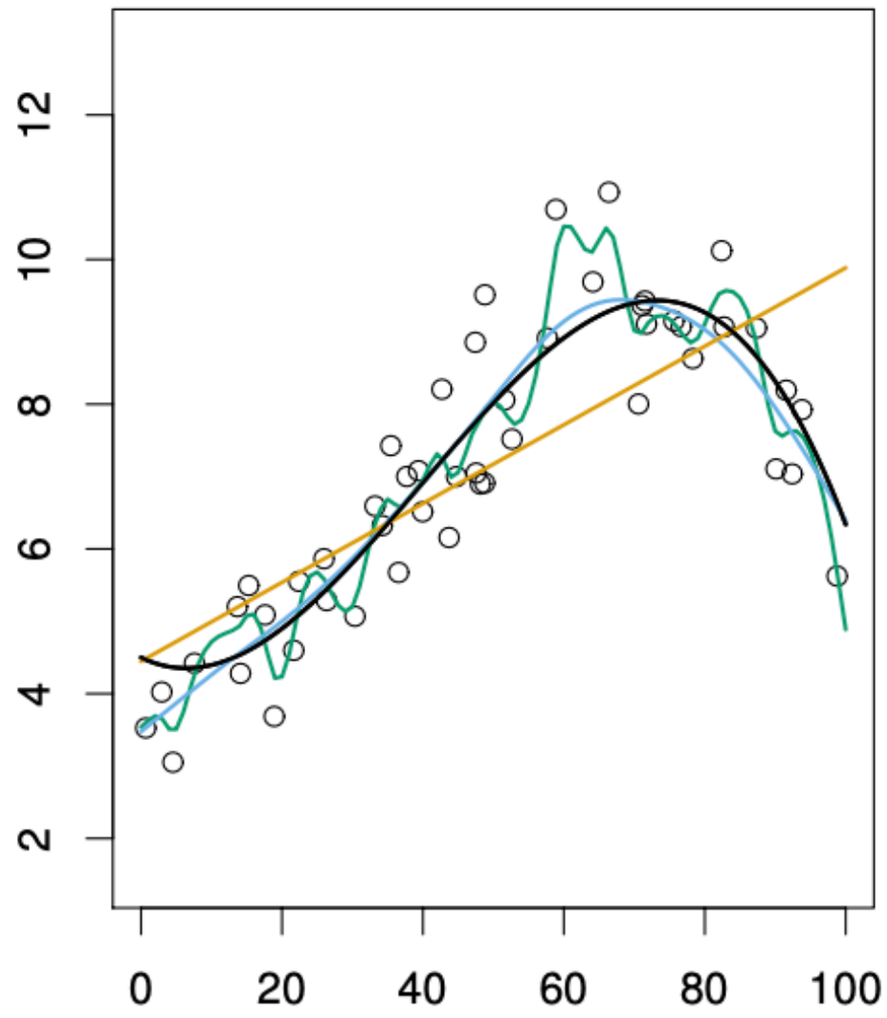
really intersted in

$$\hat{f}(x_0) \approx y_0$$



trying to fit data points

but no guarantee on the generalization



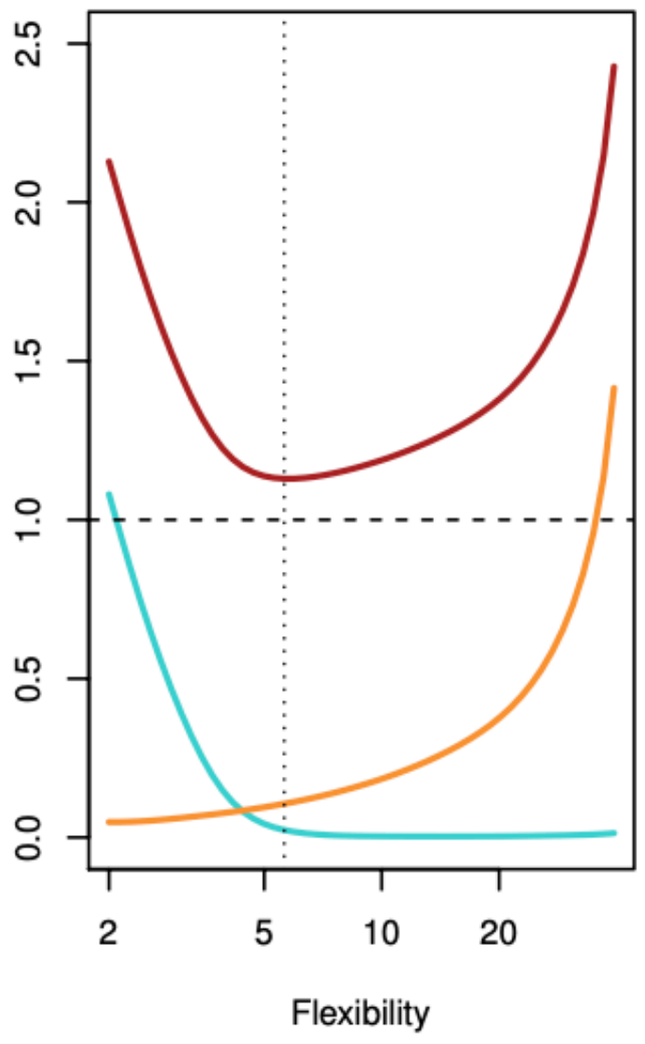
Biais-Variance Tradeoff

$$\mathbb{E}\left[(y_0 - \hat{f}(x_0))^2\right] = \text{Var}(\hat{f}(x_0)) + [\text{Biais}(\hat{f}(x_0))]^2 + \text{Var}(\epsilon)$$

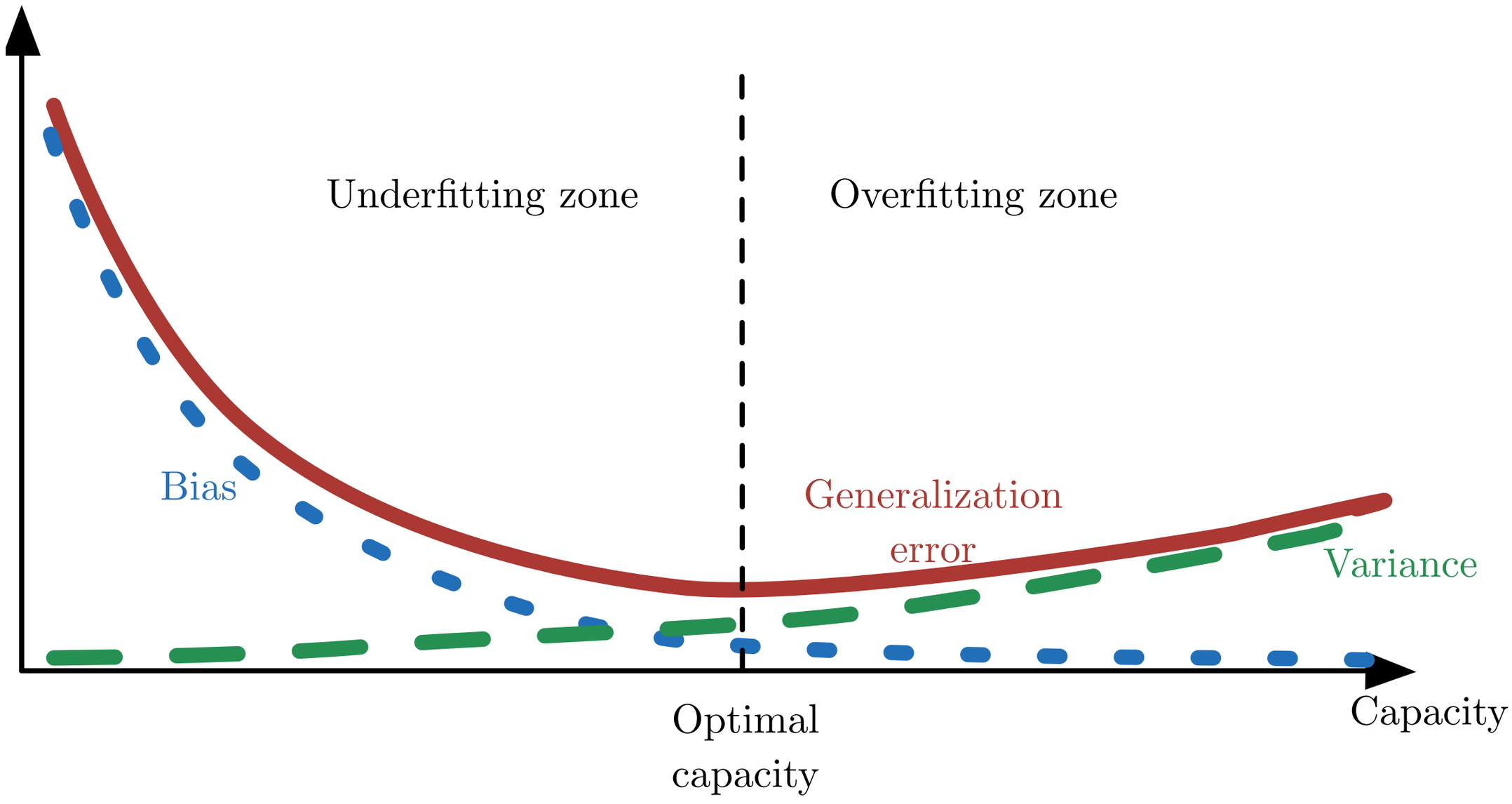
- **Variance**: the amount by which $\hat{f}$ would change if we use different training data sets (how much the model changes accross data sets)

- **Biais**: error by approximating a real-life problem (how much the model fits the data on average)

- Variance and biais are “opposite” (as the functions of model flexibility)

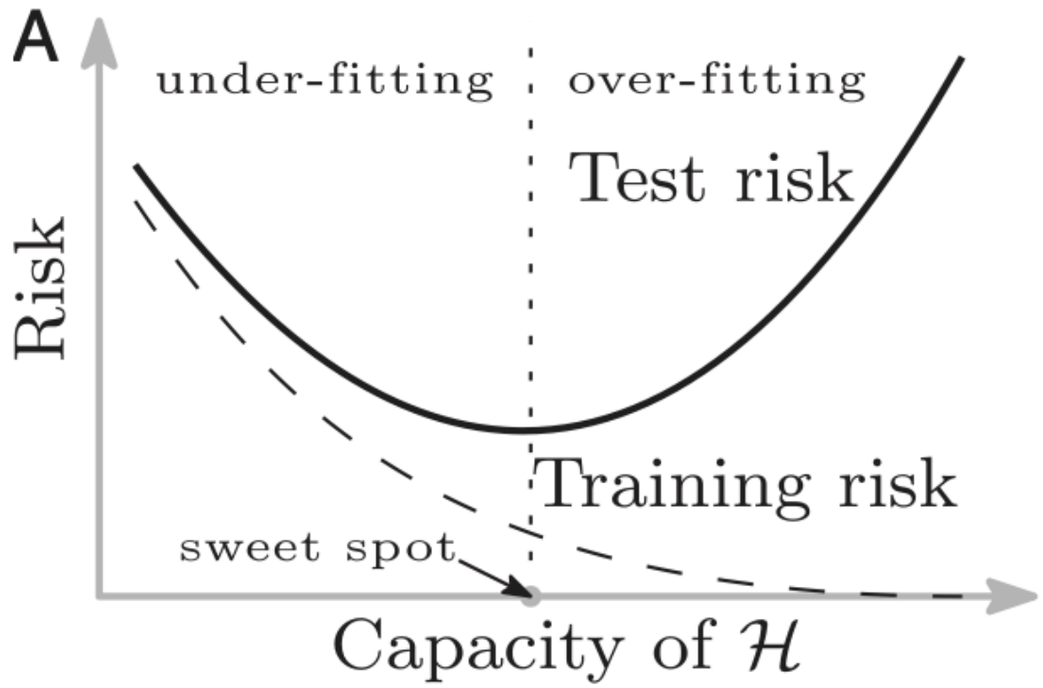


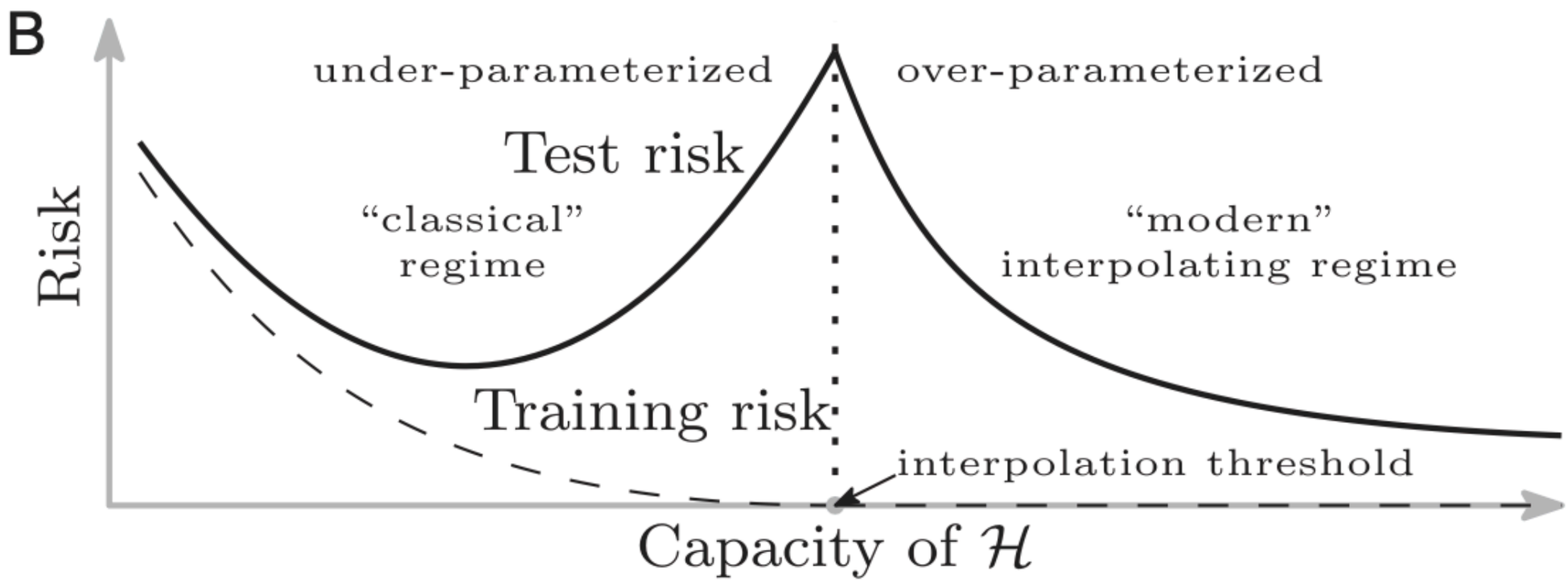
Bias and Variance



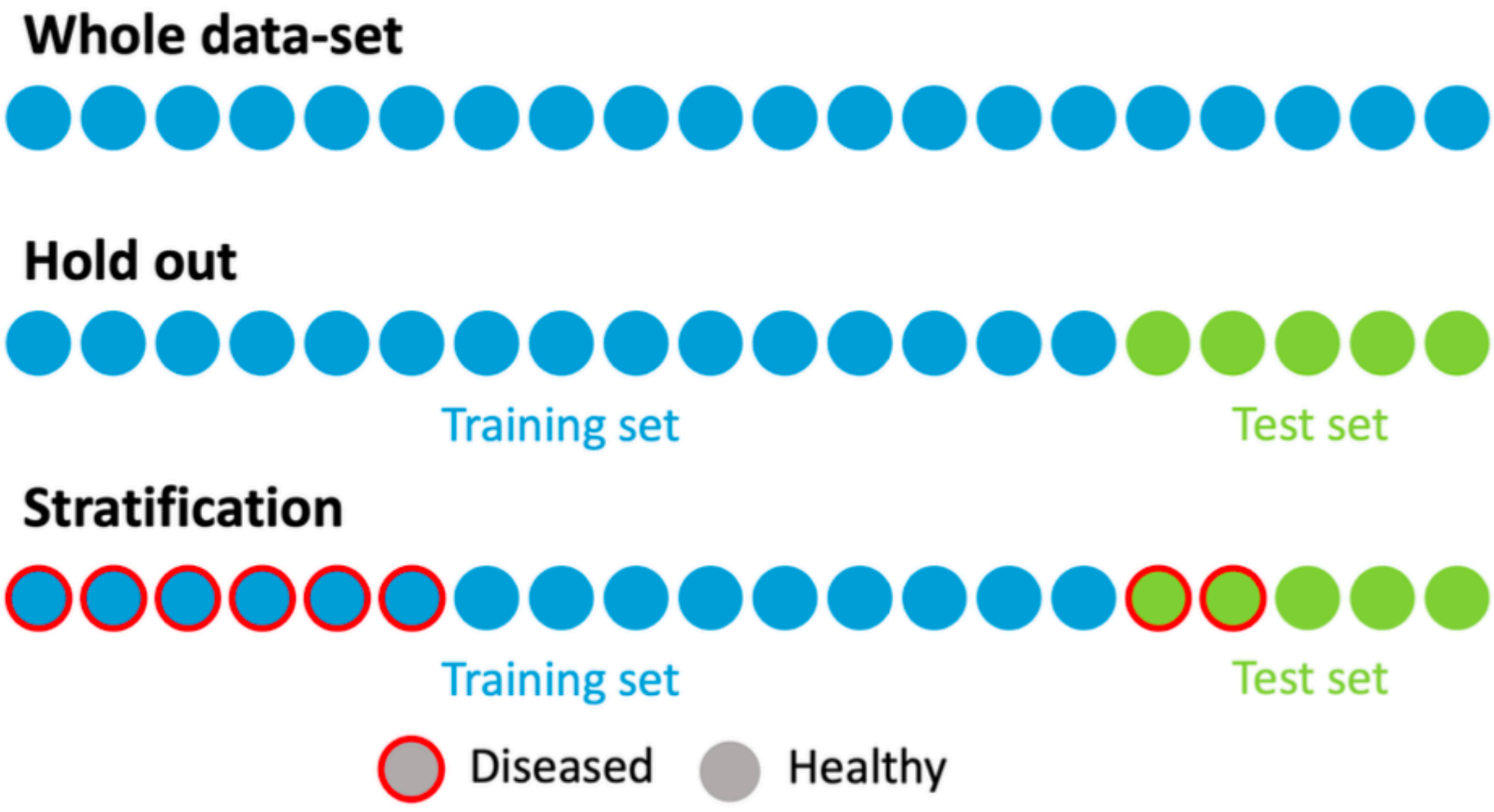
Beyond U-shaped bias-variance

- interpolation
- neural networks, random forest

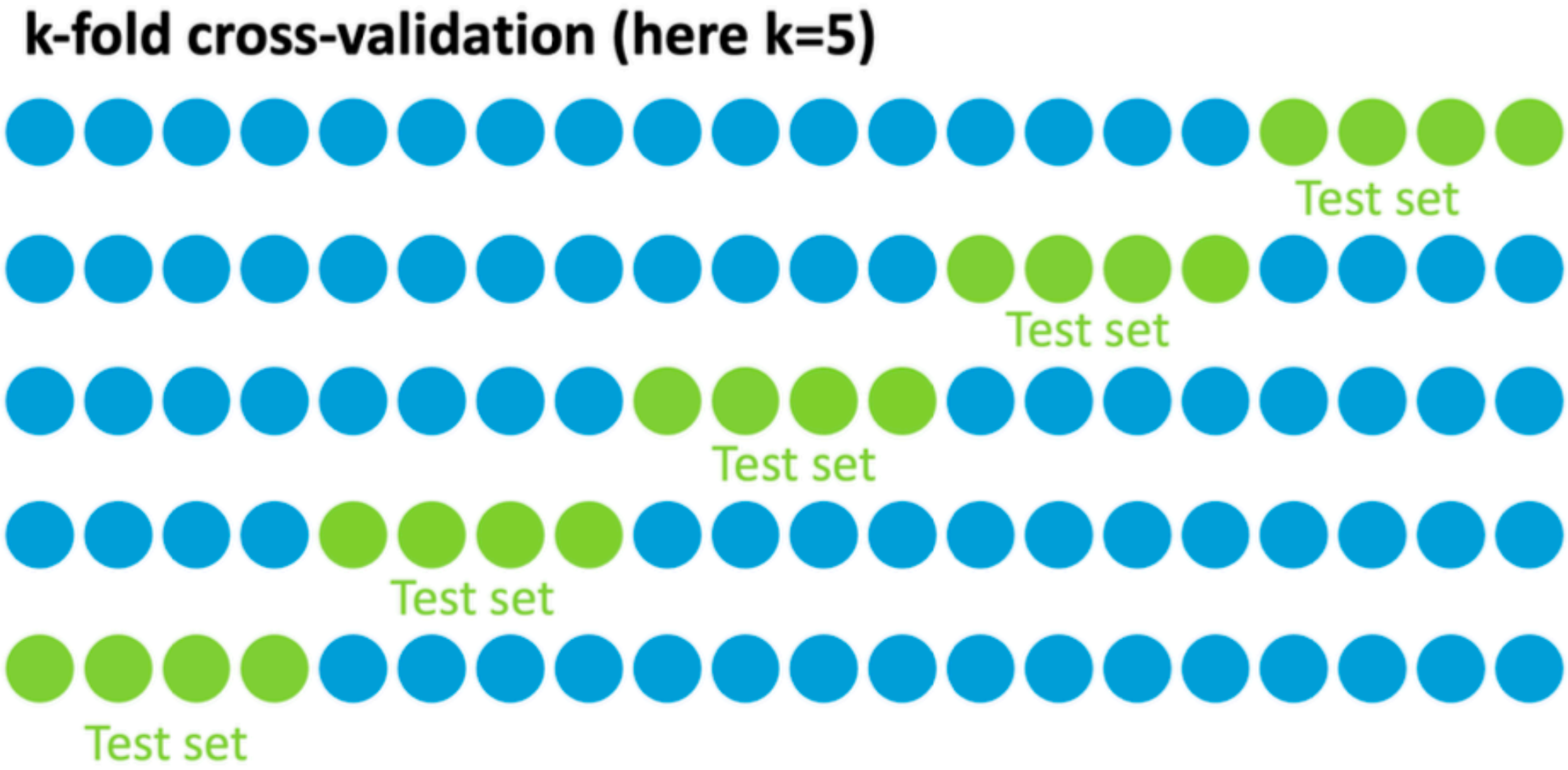




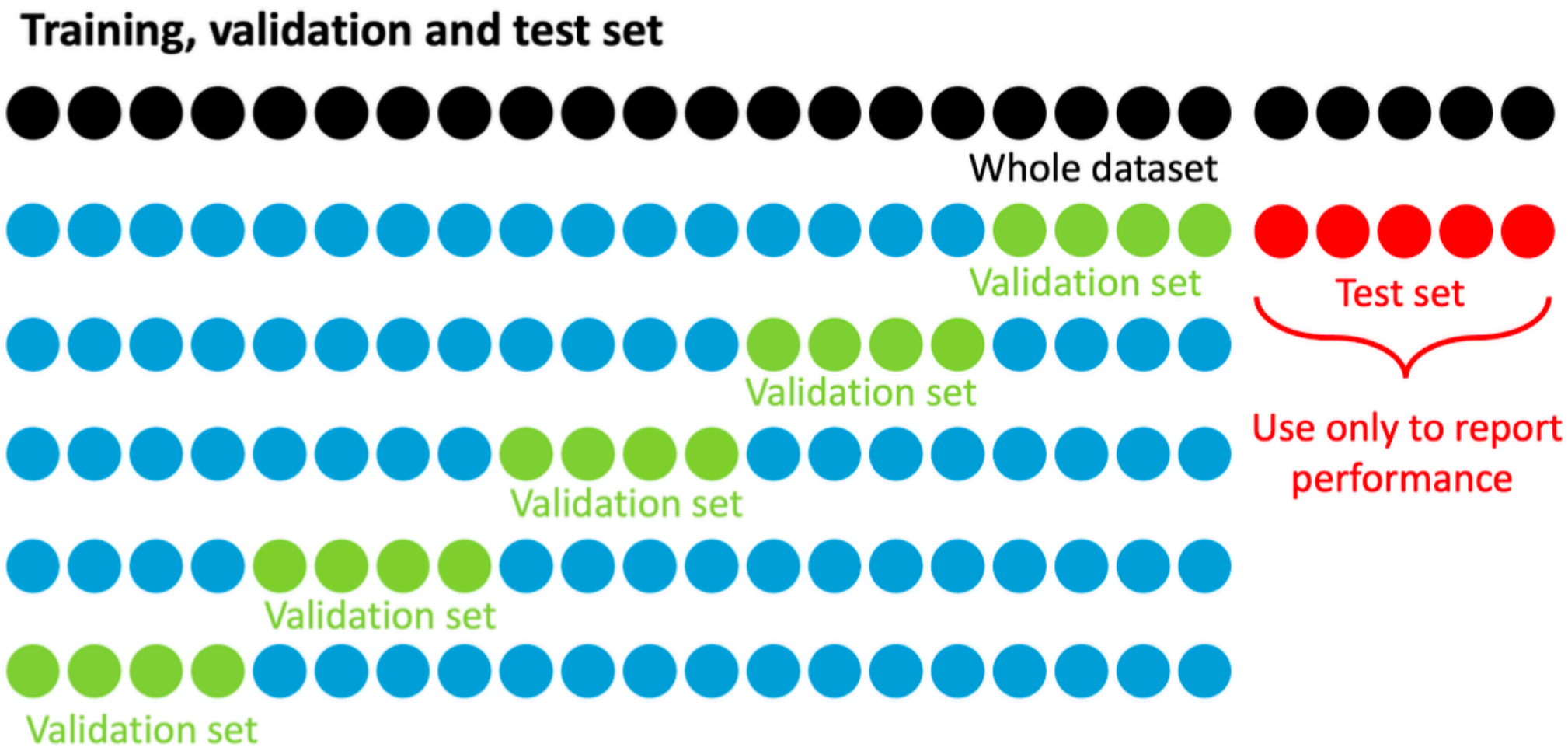
Evaluation strategies



Evaluation strategies



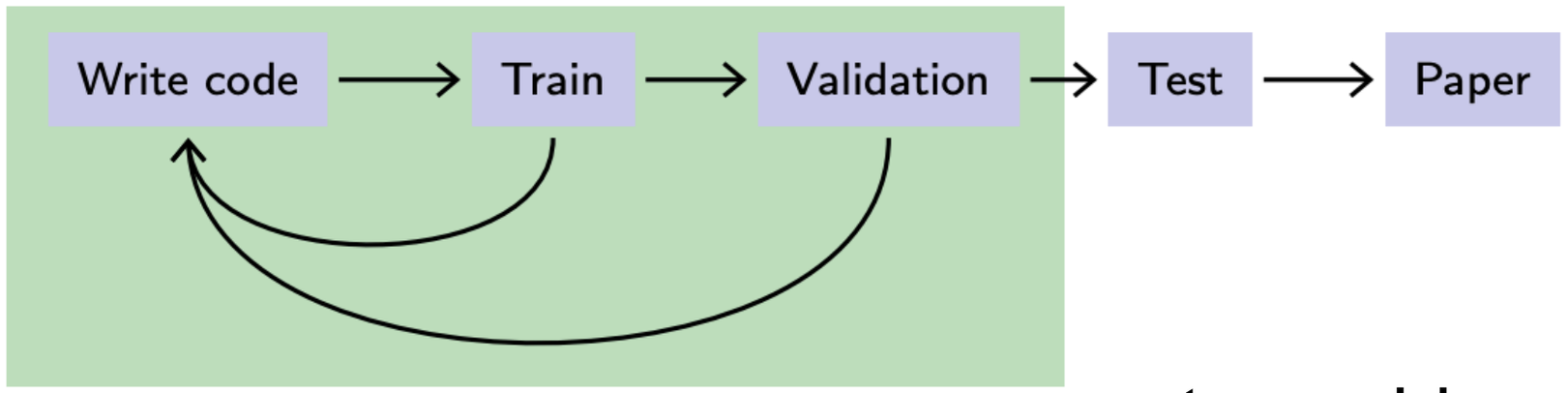
Evaluation strategies



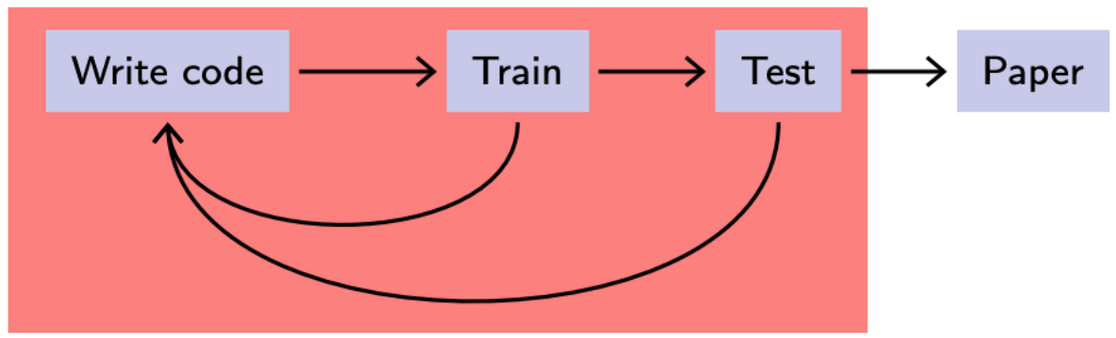
Use to train the model, experiment with different architectures…

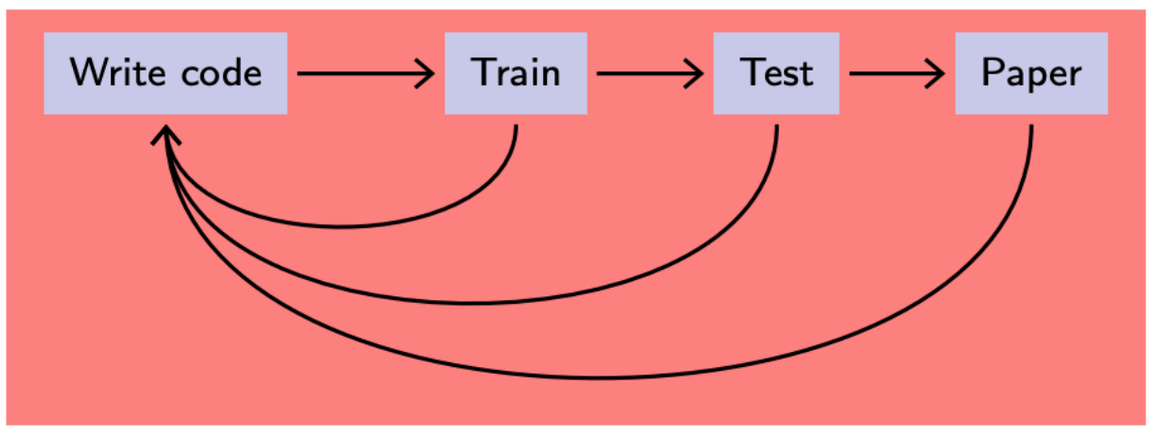
Remarks on practices

ideally



…but often





to avoid:

- early evaluation stopping
- meta-parameter over-tuning
- data-set selection
- seed-selection

References

- An introduction to Statistical Learning. *James, Witten, Hastie, Tibshirani* (Basic notions)
- Reconciling modern machine learning practice and the classical bias-variance tradeoff. *Belkin, Hsu, Ma, Mandal* (Beyond U-shape bias-variance)
- Evaluating machine learning models and their diagnostic value. *Gael Varoquaux, Olivier Colliot* (Evaluation and metrics)