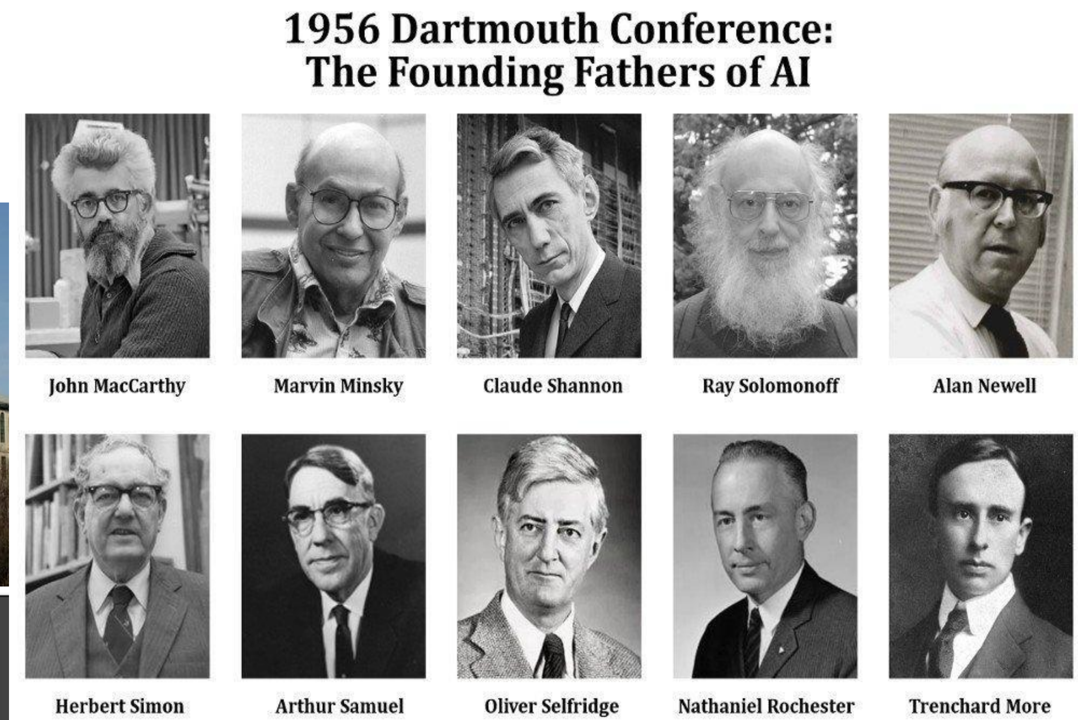
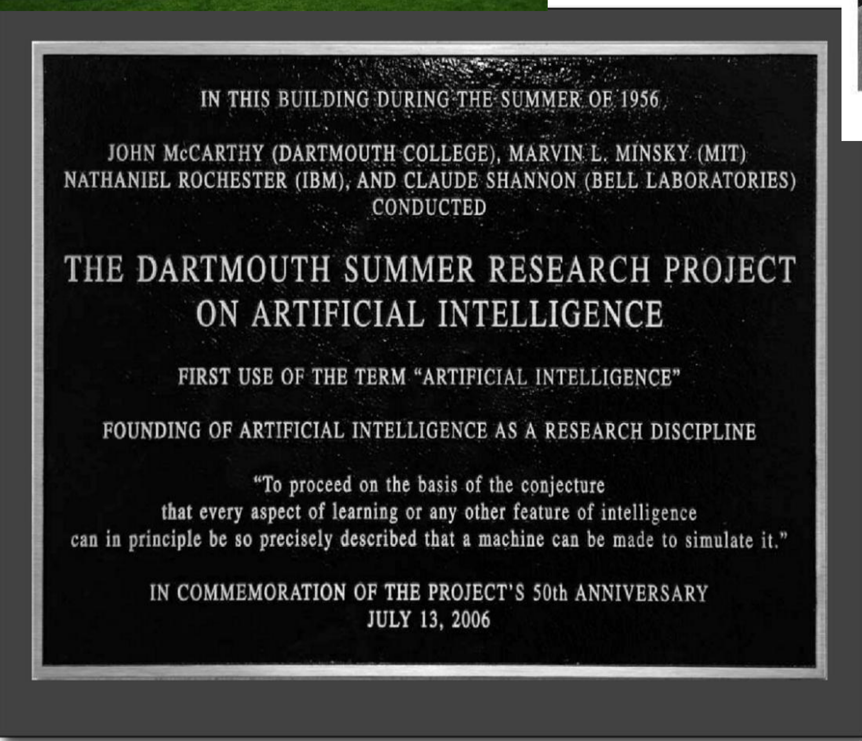
CS60050: Machine Learning

Sourangshu Bhattacharya

CSE, IIT Kharagpur

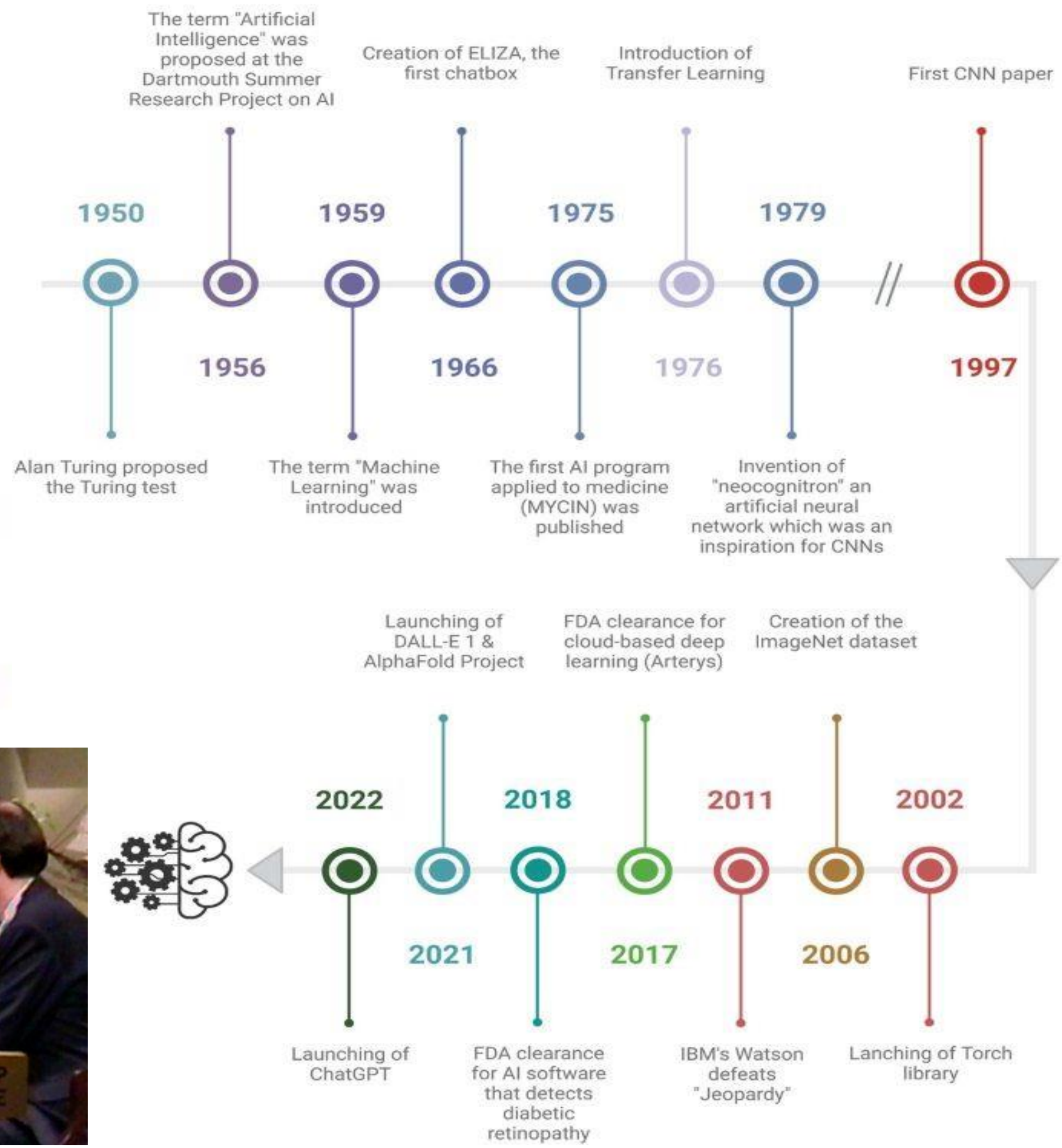
Some slides taken from Prof. Niloy Ganguly, CSE


1956 Dartmouth Conference:
The Founding Fathers of AI
John MacCarthy
Marvin Minsky
Claude Shannon
Ray Solomonoff
Alan Newell
Herbert Simon
Arthur Samuel
Oliver Selfridge
Nathaniel Rochester
Trenchard More


IN THIS BUILDING DURING THE SUMMER OF 1956
JOHN McCARTHY (DARTMOUTH COLLEGE), MARVIN L. MINSKY (MIT)
NATHANIEL ROCHESTER (IBM), AND CLAUDE SHANNON (BELL LABORATORIES)
CONDUCTED
THE DARTMOUTH SUMMER RESEARCH PROJECT
ON ARTIFICIAL INTELLIGENCE
FIRST USE OF THE TERM "ARTIFICIAL INTELLIGENCE"
FOUNDING OF ARTIFICIAL INTELLIGENCE AS A RESEARCH DISCIPLINE
"To proceed on the basis of the conjecture
that every aspect of learning or any other feature of intelligence
can in principle be so precisely described that a machine can be made to simulate it."
IN COMMEMORATION OF THE PROJECT'S 50th ANNIVERSARY
JULY 13, 2006


MASSACHVSETTS INSTITVTE OF TECHNOLOGY

Major Milestones in AI history



Kasparov Proves No Match for Computer

By Rajiv Chandrasekaran
Washington Post Staff Writer
Monday, May 12, 1997; Page A01

Turing Test

- (Human) judge communicates with a human and a machine over text-only channel.
- Both human and machine try to act like a human.
- Judge tries to tell which is which.
- Numerous variants

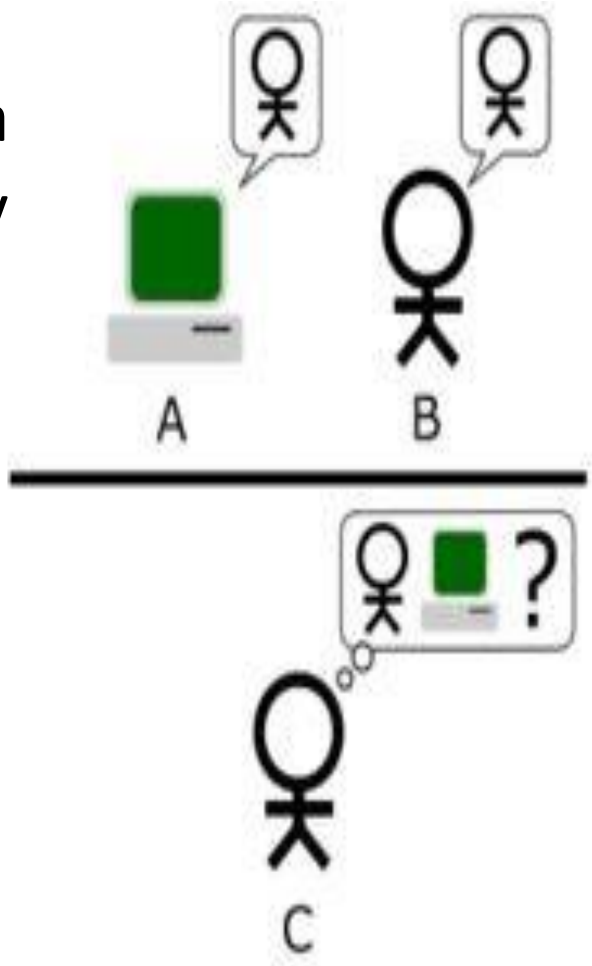


image from http://en.wikipedia.org/wiki/Turing_test

Turing Test on Unsuspecting Judges

- It is possible to (temporarily) fool humans who do not realize they may be talking to a bot

- ELIZA program [Weizenbaum 66] rephrases partner's statements and questions (~psychotherapist)

Turing Test

```
===============================================================================
           EEEEEEEE  L         IIIIIII  ZZZZZZZ      AAA
           E         L            I          Z      A   A
           E         L            I         Z      A     A
           EEEEE     L            I        Z       A     A
           E         L            I       Z        AAAAAAA
           E         L            I      Z         A     A
           EEEEEEEE  LLLLLLLL  IIIIIII  ZZZZZZ     A     A

===============================================================================
ELIZA > (PLEASE TYPE IN ALL CAPS) WHAT'S YOUR NAME DEAR ?
AMIT
ELIZA > HI, I'M ELIZA. WHAT DO YOU WANT TO TALK ABOUT ?
AMIT > CAN I TALK ABOUT MY PROBLEM ?
ELIZA > SURE... !
AMIT > DUE TO ON GOING EXAMS IN COLLEGE, I HAVE A LOT OF STRESS.
ELIZA > PLEASE RELAX A LITTLE. SLEEP WELL.
AMIT > THANKS FOR YOUR ADVICE.
ELIZA > NO MENTION.
AMIT > BYE
ELIZA > BYE AND KEEP IN TOUCH...

===============================================================================
```

Turing Test

```
ELIZA: Is something troubling you ?
YOU:   life is meaningless
ELIZA: Do you feel strongly about discussing such things ?
YOU:   everyone is born, grows and dies
ELIZA: Who, for example?
YOU:   you
ELIZA: You're not really talking about me -- are you ?
YOU:   me
ELIZA: That is interesting. Please continue.
YOU:   what is the purpose of sucess and happiness
ELIZA: Are such questions much on your mind ?
YOU:   all the time
ELIZA: Tell me more about that.
YOU:   why should i study
ELIZA: That's quite interesting.
YOU:   i cant take my knowledge with me
ELIZA: Is it important to you that your knowledge with you ?
YOU:   i cant take anything with me
ELIZA: Have you tried ?
YOU:   i'll have to die for that
ELIZA: Can you elaborate on that ?
```

What is Artificial Intelligence

“[The automation of] activities that we associate with human thinking, activities such as decision making, problem solving, learning” (Bellman 1978)

“The study of mental faculties through the use of computational models” (Charniak & McDermott, 1985)

Systems that think like humans	Systems that think rationally
Systems that act like humans	Systems that act rationally

Good Old AI Days

Representing Knowledge

- Logic
- Rules
- Semantic Graphs/Nets

A Few Statements

- All people who are graduating are happy.
- All happy people smile.
- Someone is graduating.
- Is someone smiling? (Conclusion)

Predicates

1. For all (x) graduating (x) $\rightarrow$ happy (x)
2. For all (x) happy (x) $\rightarrow$ smiling (x)
3. There exists (x) graduating (x)

Rule Based Inference Example

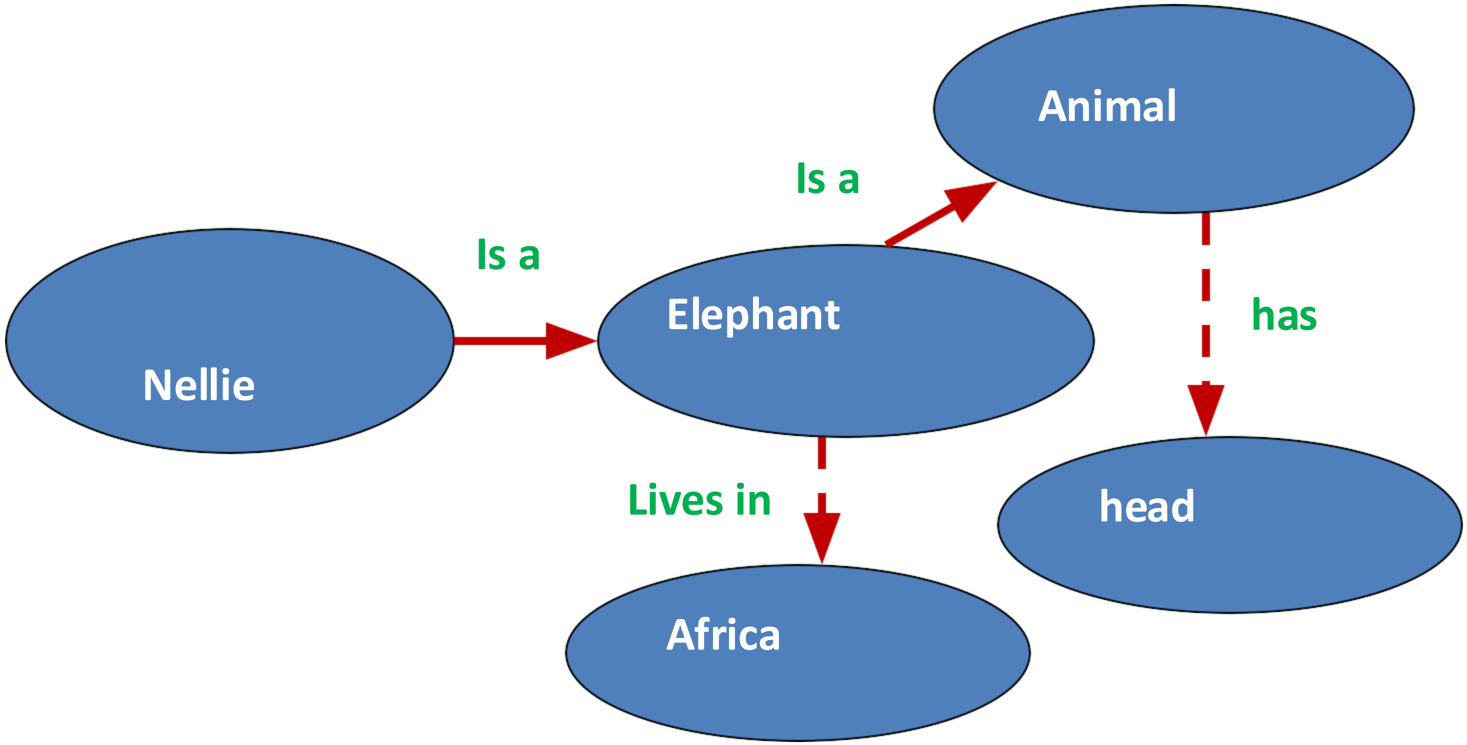
(R1) if gas_in_engine and does not start, then problem(spark_plugs).

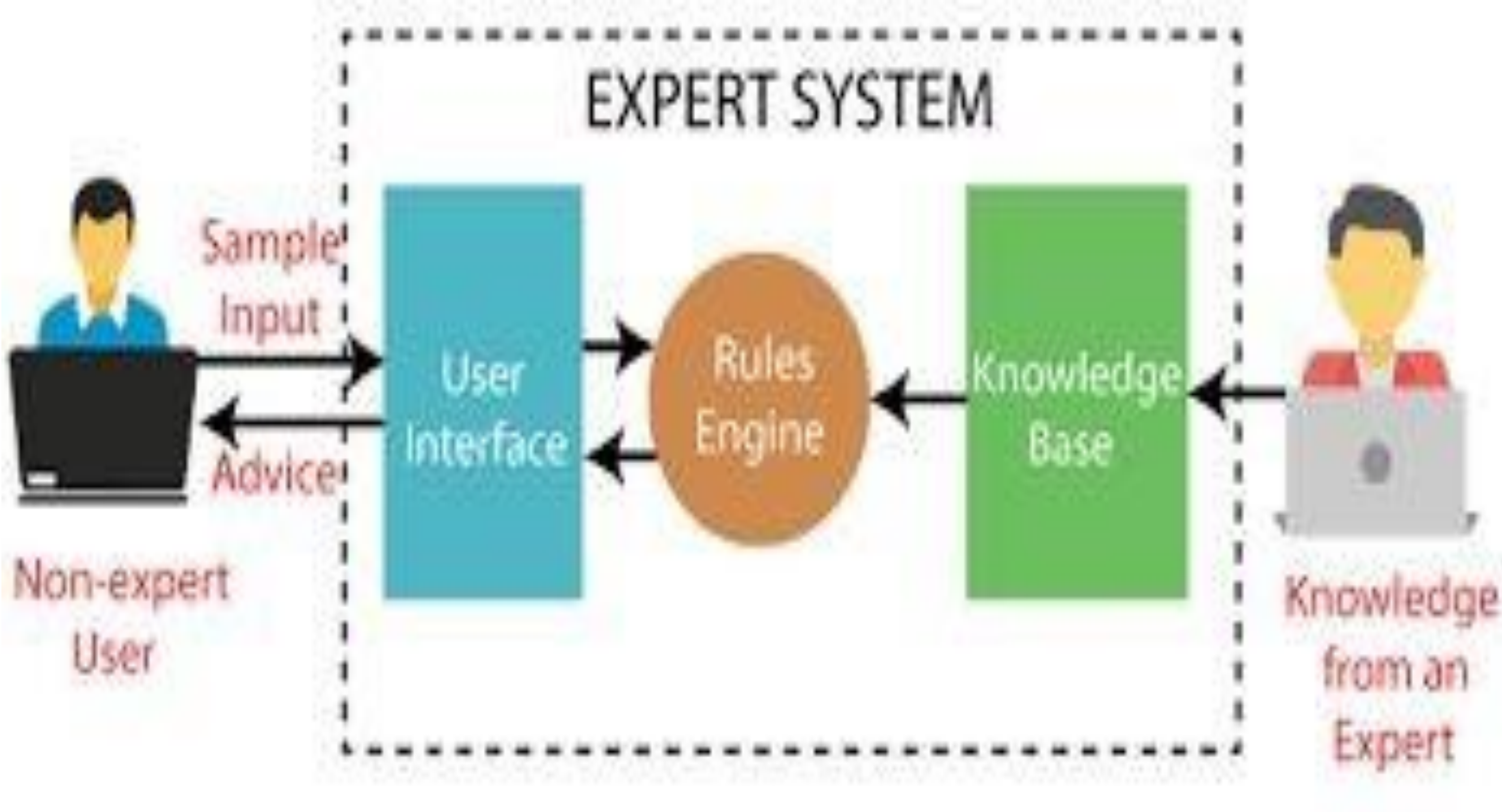
(R2) if not (does not start) and not (lights_on), then problem(battery).

(R3) if not(turns_over) and light_on, then problem(starter).

(R4) if gas_in_tank and gas_in_carb, then gas_in_engine

Semantic Nets




EXPERT SYSTEM
Sample Input
Advice
User Interface
Rules Engine
Knowledge Base
Non-expert User
Knowledge from an Expert

Hit the Wall

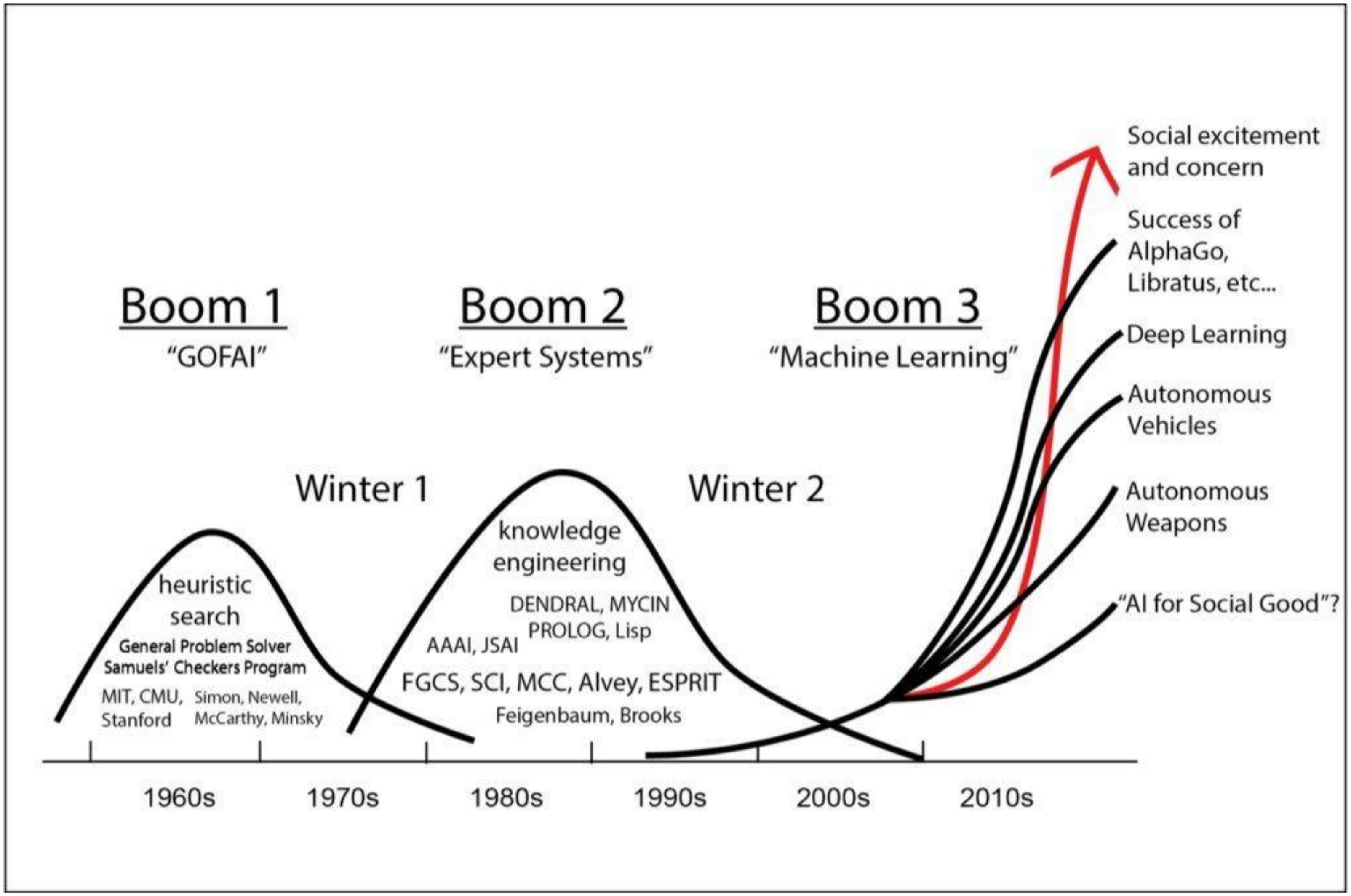
- **Ambiguity**: highly funded translation programs (Russian to English) were good at syntactic manipulation but bad at disambiguation

  "The spirit is willing but the flesh is weak" becomes "The vodka is good but the meat is rotten"

- **Scalability/complexity**: early examples were very small, programs could not scale to bigger instances

- Limitations of **representations** used

AI Winter



Boom 1
"GOFAI"
Boom 2
"Expert Systems"
Boom 3
"Machine Learning"
Winter 1
Winter 2
heuristic search
General Problem Solver
Samuels' Checkers Program
MIT, CMU, Stanford
Simon, Newell, McCarthy, Minsky
knowledge engineering
DENDRAL, MYCIN
PROLOG, Lisp
AAAI, JSAI
FGCS, SCI, MCC, Alvey, ESPRIT
Feigenbaum, Brooks
Social excitement and concern
Success of AlphaGo, Libratus, etc...
Deep Learning
Autonomous Vehicles
Autonomous Weapons
"AI for Social Good"?
1960s
1970s
1980s
1990s
2000s
2010s

Machine Learning

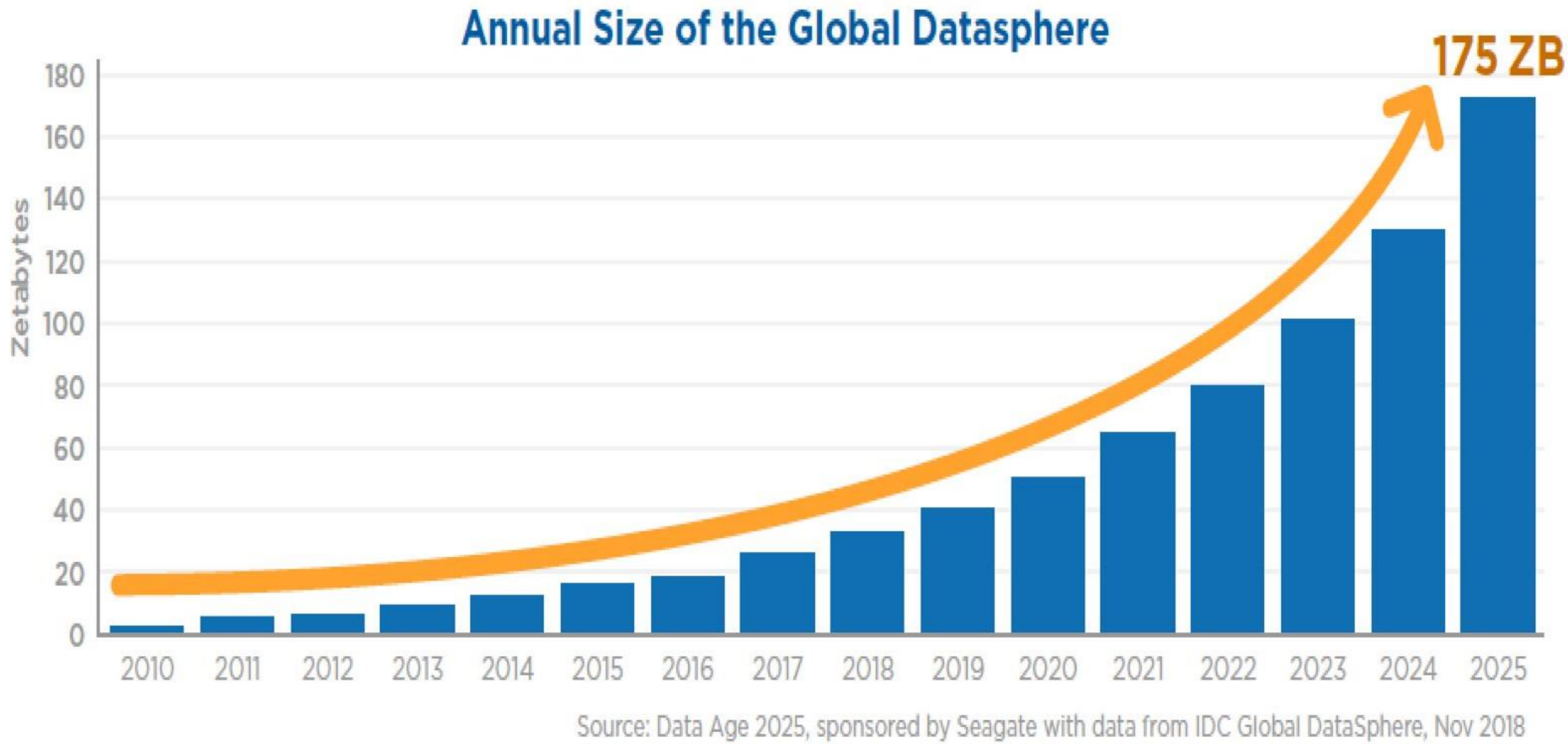
Data

Data

- Collection of measurements or observations that can be used to train a model.
- Can be categorical (cats,dogs,lion,etc), ordinal (tall, medium, short), continuous (10-15,15-20,20-25,...).

Type of Media	Amount per Minute	Amount per Day
Emails sent	231.4 million	333.22 billion
Crypto purchased	90.2 million	129.89 billion
Texts sent	16 million	24.04 billion
Google searches	5.9 million	8.5 billion
Snaps shared on Snapchat	2.43 million	3.5 billion
Pieces of content shared on Facebook	1.7 million	2.45 billion
Swipes on Tinder	1.1 million	1.58 billion
Hours streamed	1 million	1.44 billion
USD spent on Amazon	443,000	637.92 million
USD sent on Venmo	437,600	630.14 million
Tweets shared on Twitter	347,200	499.97 million
Hours spent in Zoom meetings	104,600	150.62 million
USD spent on DoorDash	76,400	110.02 million

Data



Annual Size of the Global Datasphere
175 ZB
Zetabytes
180
160
140
120
100
80
60
40
20
0
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
Source: Data Age 2025, sponsored by Seagate with data from IDC Global DataSphere, Nov 2018

Machine Learning

— a field of **study** in **artificial intelligence** concerned with the development and study of **statistical algorithms** that can **learn** from data (seen data) and **generalize** to unseen data and thus perform tasks without explicit instructions.

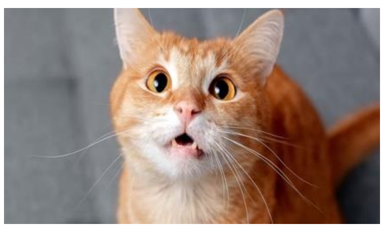
— "a computer program is said to learn from **experience E** with respect to some **task T** and some **performance measure P**, if its performance on T, as measured by P, **improves** with experience E." - Tom Mitchell, 1997

Emergence of Supervised Learning

Supervised Learning
Human Perspective

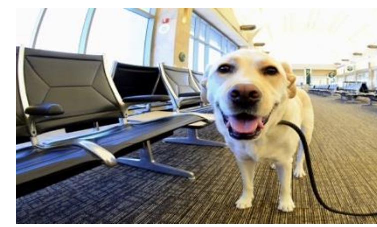
- **Scenario [How a child learns]**
  - Below are the images of cats and dogs

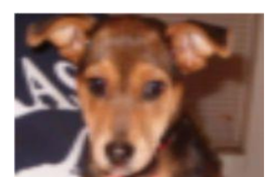
Cats





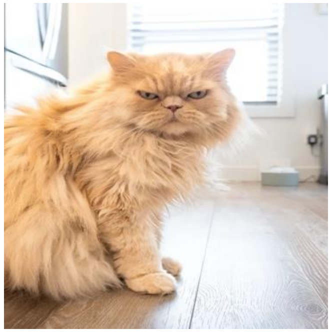
Dogs

Supervised Learning
Human Perspective

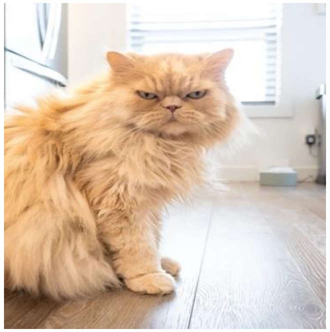
- **Scenario [How a child learns]**
  - Now, we provide the child with some images and ask the child which is what?



Supervised Learning
Human Perspective

- **Scenario [How a child learns]**
  - Child guesses the images are cats and dogs as shown below.

Cats

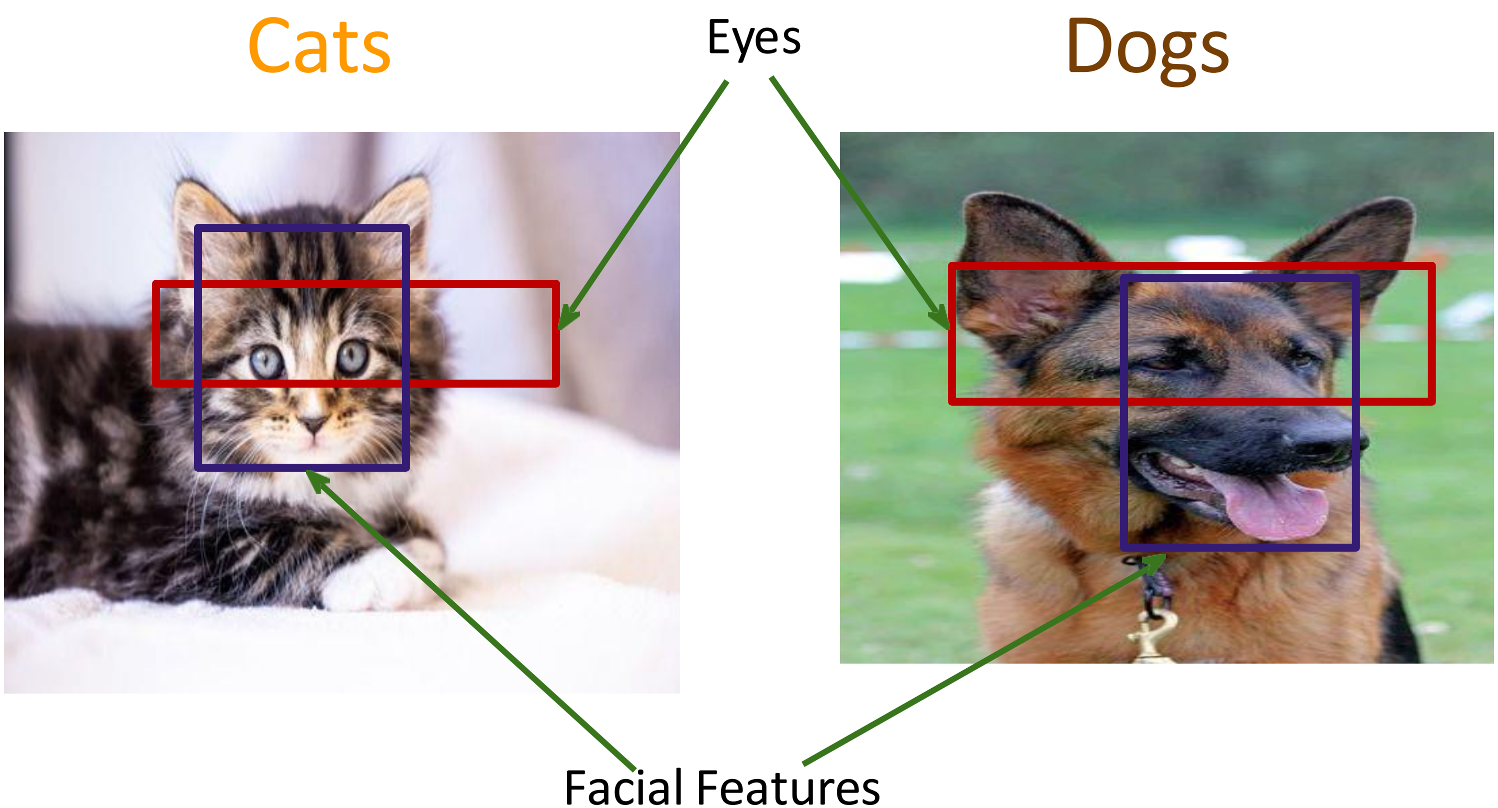


Dogs

How did a child know?

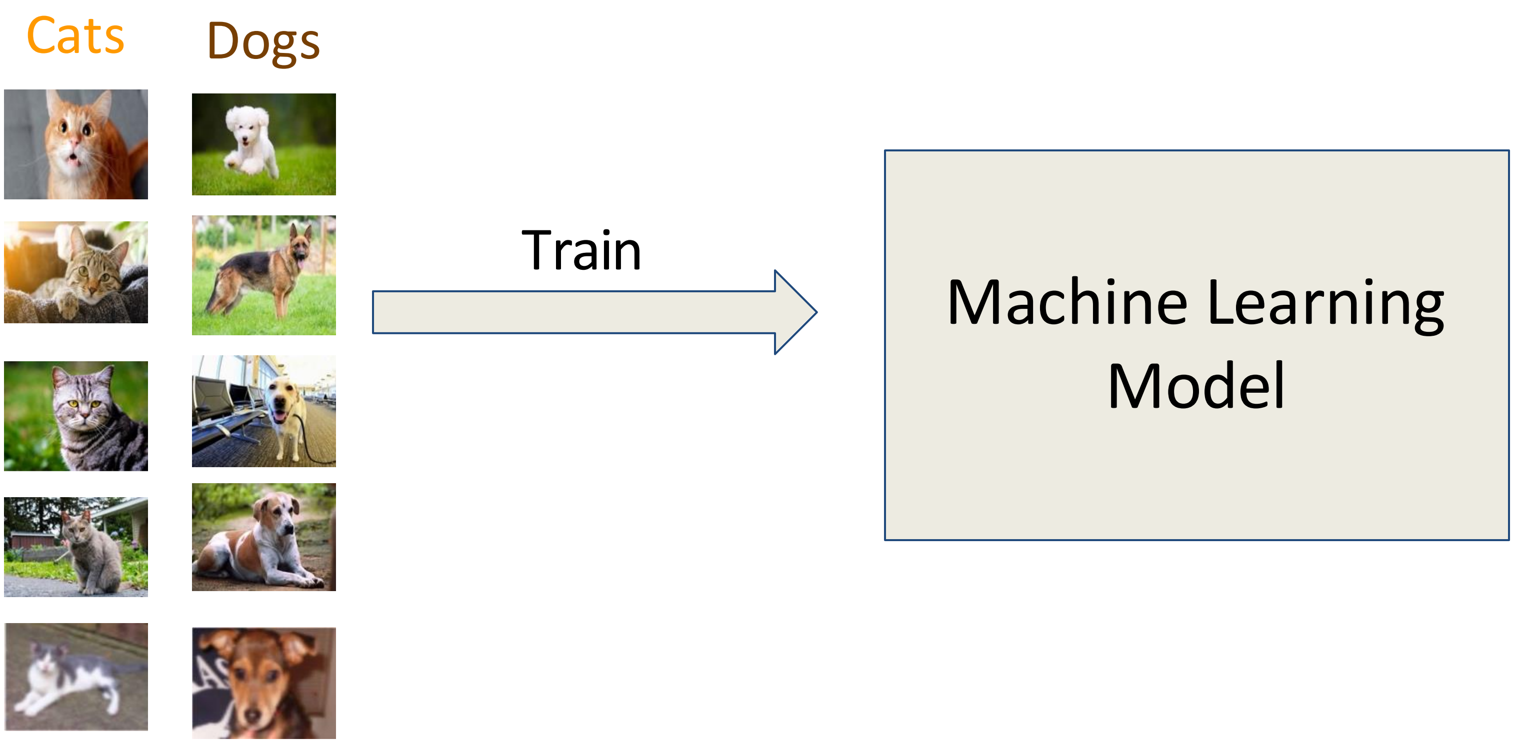
Supervised Learning
Human Perspective

– **Scenario [How a child learns]**
  - The child understood the features which distinguishes dogs from cats!



Supervised Learning
Model Perspective

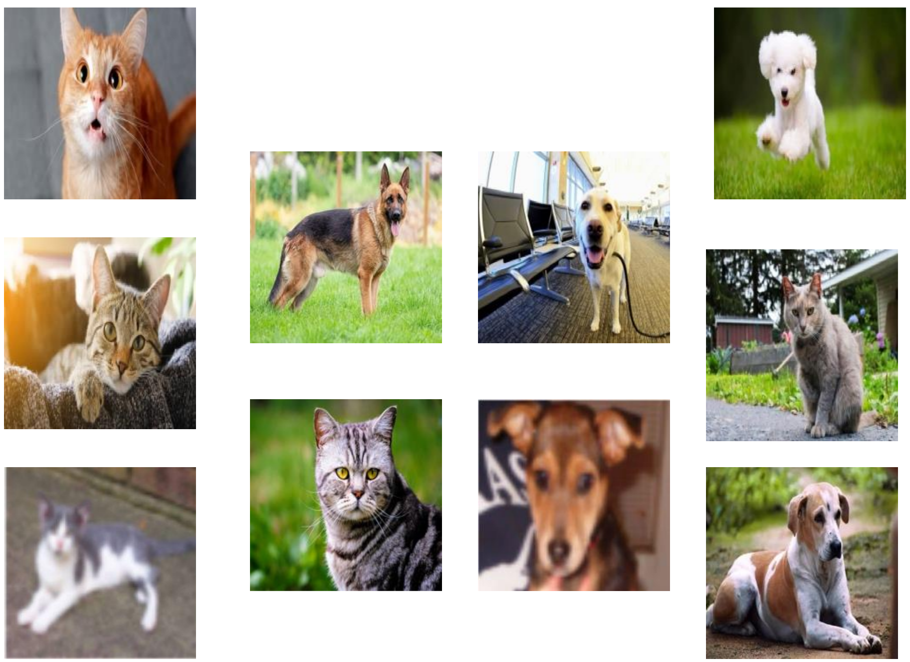
- **Model Perspective**
  - If we are to use a machine learning model which tells us how to decide/classify which is a cat or dog, we train the model on the images as shown below



Supervised Learning
Model Perspective

– **Procedure :**
  - Divide Dataset into Train and Test

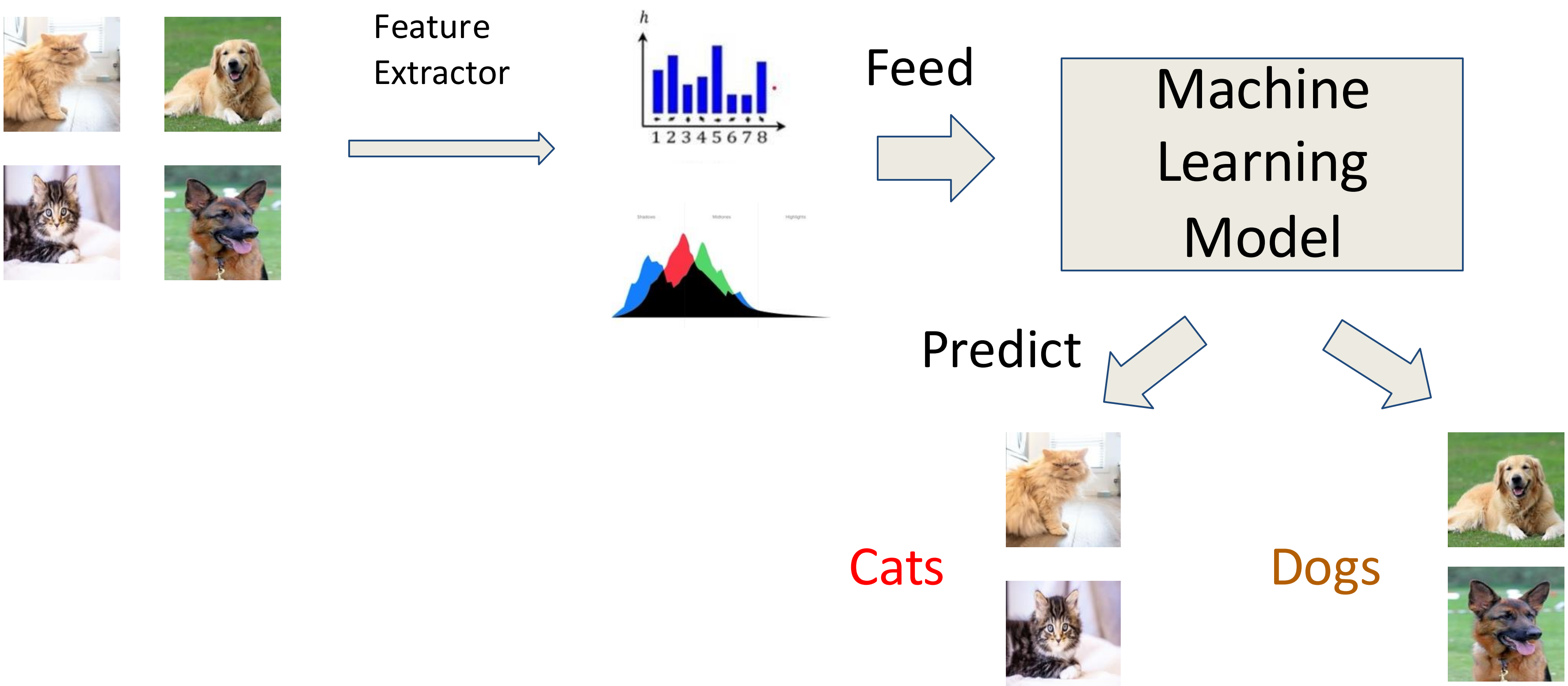
Training Dataset



Testing Dataset

Supervised Learning
Model Perspective

– **Procedure**
  - Predict the classes from the Model on the images in the similar manner



Problems with Data Annotation

- **Supervised Learning requires labelled data.**
- Large amounts of data available (in Zettabytes).
- Almost all are unlabelled.
- Cannot label each and every data be it image, text, audio, video, molecules etc. **Labour Intensive**
- What can we do?

Unsupervised Learning

When data is unlabelled

Unsupervised Learning

- **Definition :**
  Type of learning where data is unlabelled/unknown and models learn these type of data for hidden patterns or data groupings.

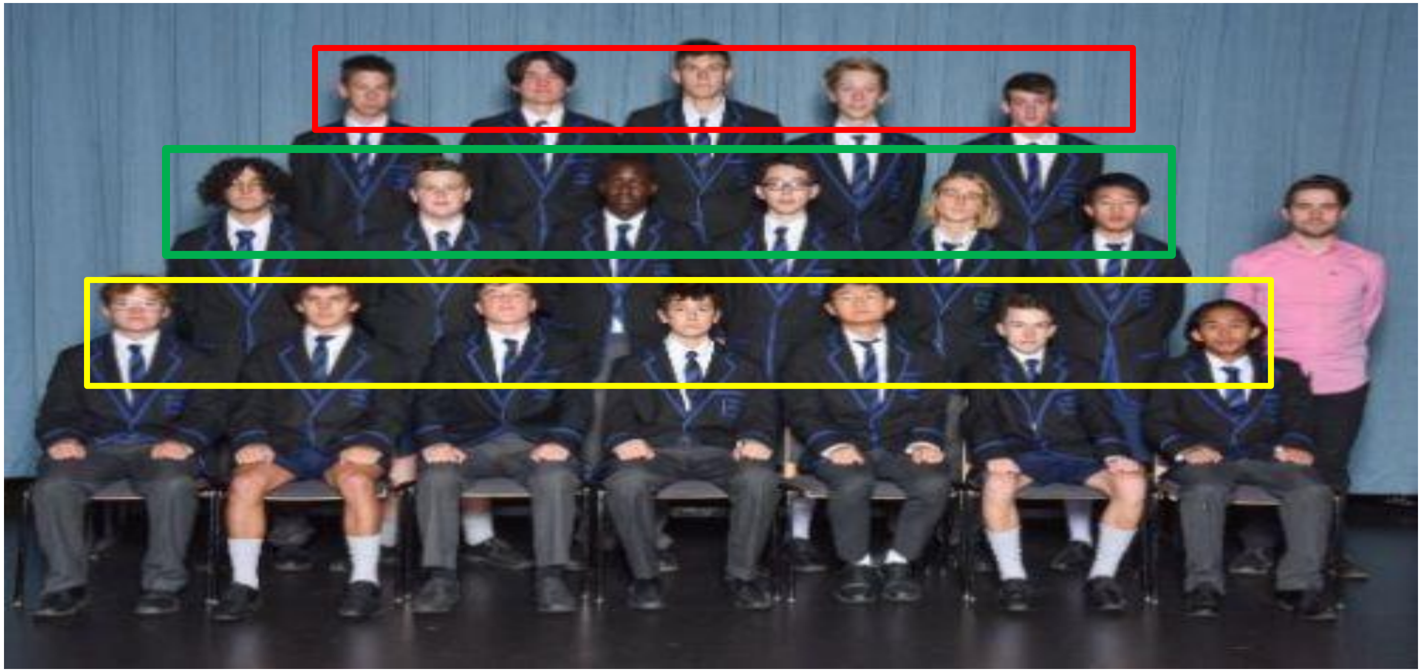
- **Types :**
  - **Clustering :** Discover groupings from unknown data. Example : Spam Emails
  - **Association :** Find out rules to express your unknown data. Example : Recommendation System

Unsupervised Learning

- **Scenario [Human Behavior]**
  - Suppose your class is being ready for a group photograph and taking positions in order of heights.
  - You do not possess any prior knowledge of the heights of your classmates.

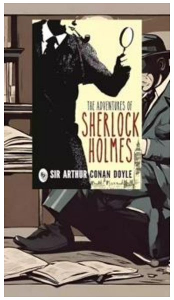
Unsupervised Learning

- **Scenario [Human Behavior]**
  - You try to get into a position according to the height.
  - You figure it out without being told as to where to stand.



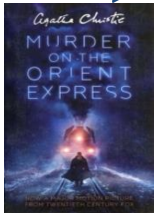
Unsupervised Learning

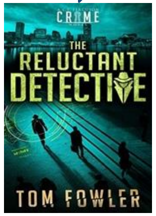
Those who bought this

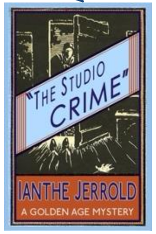


Cluster of Detective Novels

also bought these







Recommender Systems
Example of Unsupervised Learning

Semi Supervised Learning

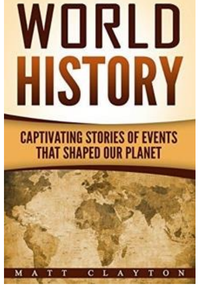
- Definition :
  - Supervised $\bigcap$ Unsupervised
  - Using labelled and unlabelled data for classification and regression tasks.
  - Number of labelled data is usually much less than that of unlabelled data.
  - Primarily dealing with unlabelled data.
- One such example is pseudo labelling.

Semi Supervised Learning

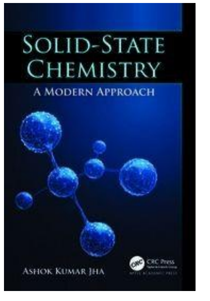
Books are present, know labels of some and not of others

Semi Supervised Learning

We know the labels of these 2 books among other books



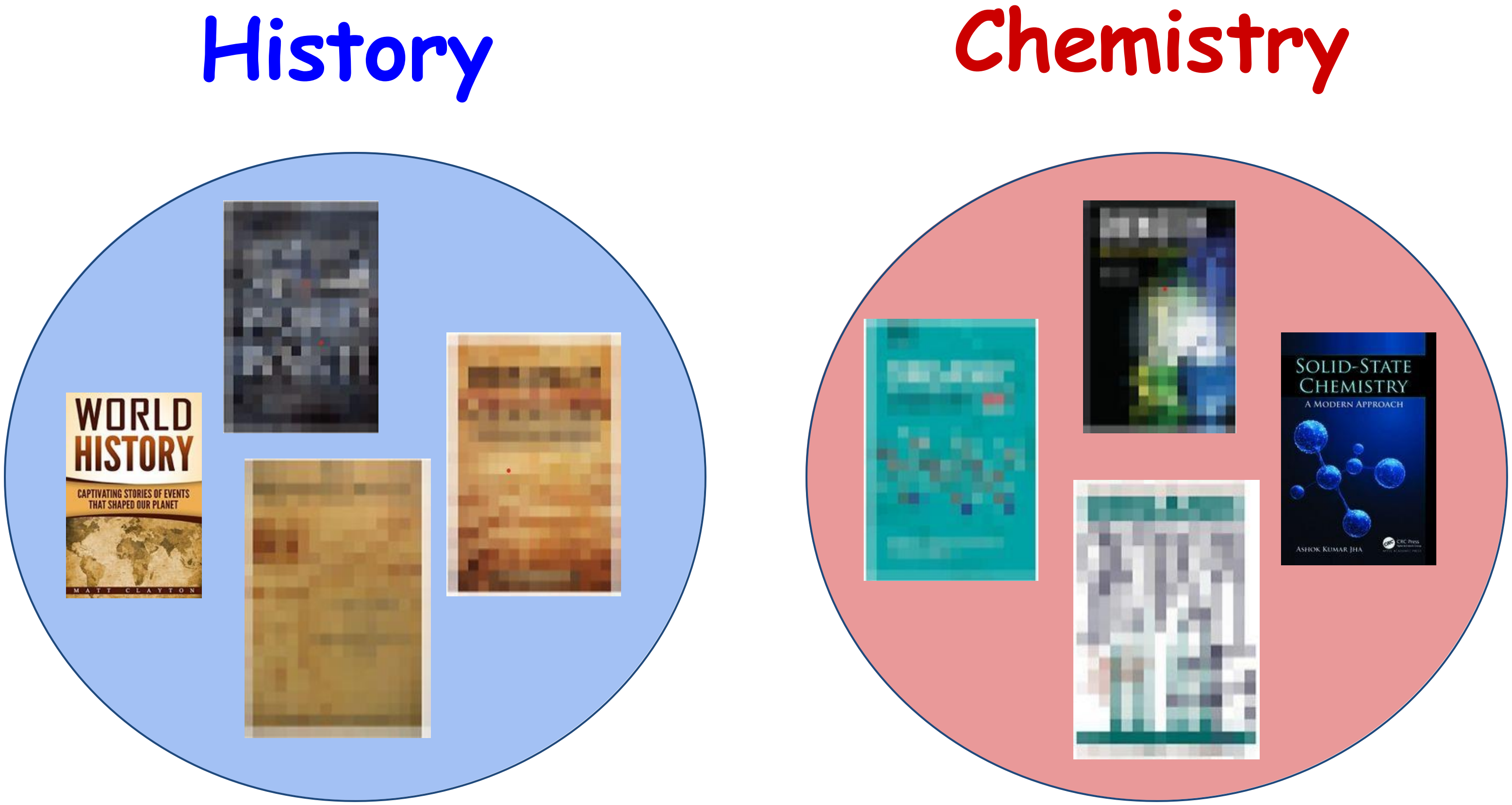
History



Chemistry

How can we get to know about the class of other books?

Semi Supervised Learning



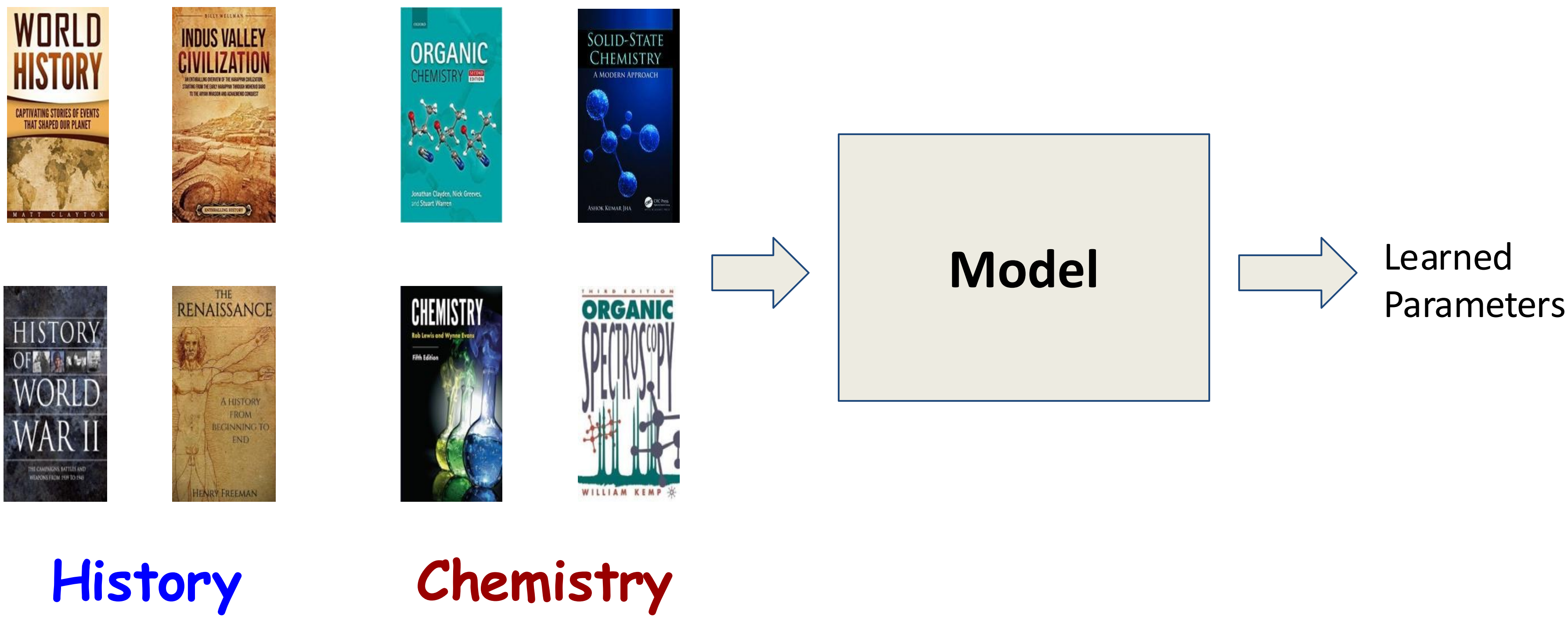
We use Clustering Algorithm (Unsupervised Learning) to cluster similar books (based on content) and label them!

Semi Supervised Learning

History | **Chemistry**

We use Clustering Algorithm (Unsupervised Learning) to cluster similar books (based on content) and label them!

Semi Supervised Learning



We then use these labelled data and train our Model in a Supervised Learning Manner!

Types of Learning

Supervised Learning
Unsupervised Learning
Semi-Supervised Learning.

Ensemble Learning [combining models together]

Self Supervised Learning [how we learn language]

Reinforcement Learning [how robot learns]

Discriminative And Generative Tasks

- **Discriminative Tasks : Classification**

  - Identify whether Cat or Dog
  - Identify whether the next word would be noun or pronoun.

- **Generative Tasks :**
  - Generates sentences based on instruction.

ChatGPT

Generative Task is a Discriminative Task!

Consider the well known sentence containing all the English Alphabets -

“The quick brown fox jumps over a lazy dog”

Suppose Given

“The quick brown

The model needs to generate the entire sentence

Generative Task is a Discriminative Task!

Task given “The quick **brown**” classify which of the following words will be the next word.
[**fox, ox, tiger, ant, duck**] — the model l classify “**fox**” as 1 and the rest as 0.

Recursively Given “The quick **brown fox**”, classify which word will be the next word — the model classify “**jumps**”

Given .“The quick **brown fox jumps**”, classify which word will be the next word — the model will classify “**over**” and so on….

We observe that if a series of such **Prediction** (or) **Classification** task (or) **Discriminative** Task are done, and each word predicted is appended with the phrase and run again for Prediction, we get a sentence.

Generative Task is a Discriminative Task!

Et Voila! We find that Generative Task is a Sequence of Discriminative Task!

Concept Learning Example

Version Spaces

Concept learning

Example / Instance: an atomic (real life) situation / object over which we want to learn.

Instance space: Set of all possible instances.

Attributes: observable quantities which describe a situation.

Concept: a Boolean valued function over set of examples.

Hypothesis space: subset of all Boolean valued functions over instance space.

Concept Learning - example

Attributes: Sky, Air temp, Humidity, Wind, Weather, Forecast.

Instance space X. What is the size ?

Hypothesis space: conjunction of literals (which are conditions over attributes).
Conditions are of the form: (attr=val) or (attr=?) or (attr=ϕ)

What is the size of hypothesis space ?

Concept Learning - example

Training Examples for EnjoySport

Sky	Temp	Humid	Wind	Water	Forecst	EnjoySpt
Sunny	Warm	Normal	Strong	Warm	Same	Yes
Sunny	Warm	High	Strong	Warm	Same	Yes
Rainy	Cold	High	Strong	Warm	Change	No
Sunny	Warm	High	Strong	Cool	Change	Yes

What is the general concept?

Inductive learning problem

Training examples: $D=\{ (x_1,c(x_1) , \ldots , (x_n,(c(x_n)) \}$

Problem: Given D, learn $h \epsilon H$, such that for all $x \epsilon X$, $h(x)=c(x)$.

Inductive learning assumption:

Any hypothesis found to approximate target concept well over sufficiently large training set, will also approximate it well over unseen examples.

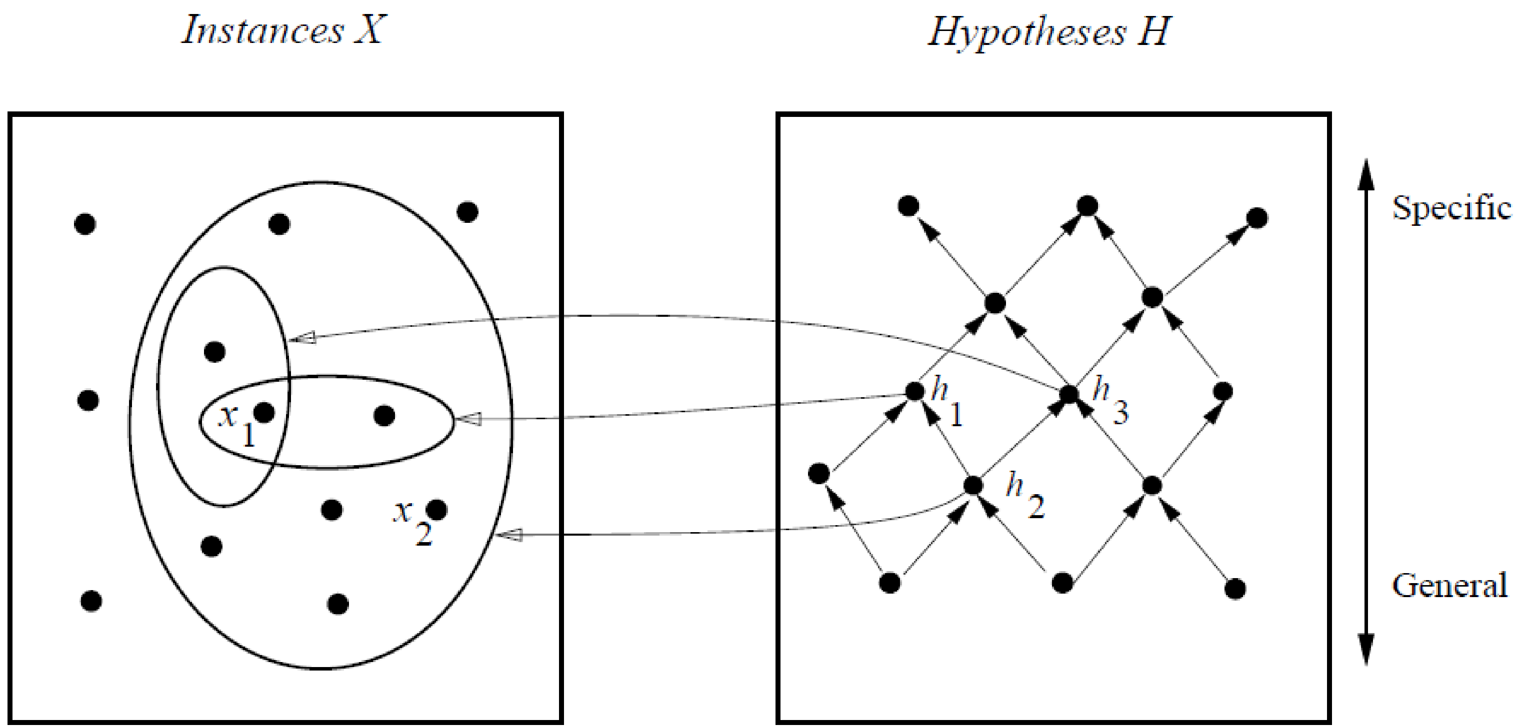
General to specific ordering

Example x is said to be positive if $c(x) = 1$, else negative.

Hypothesis h "satisfies" x, if $h(x) = 1$.

Hypothesis h_2 is said to be "more general or equal to" h_1 if
for all x: $h_1(x) = 1$ implies $h_2(x) = 1$

General to specific ordering



x_1 = <*Sunny, Warm, High, Strong, Cool, Same*>
x_2 = <*Sunny, Warm, High, Light, Warm, Same*>

h_1 = <*Sunny, ?, ?, Strong, ?, ?*>
h_2 = <*Sunny, ?, ?, ?, ?, ?*>
h_3 = <*Sunny, ?, ?, ?, Cool, ?*>

Find - S

Finding maximally specific hypothesis

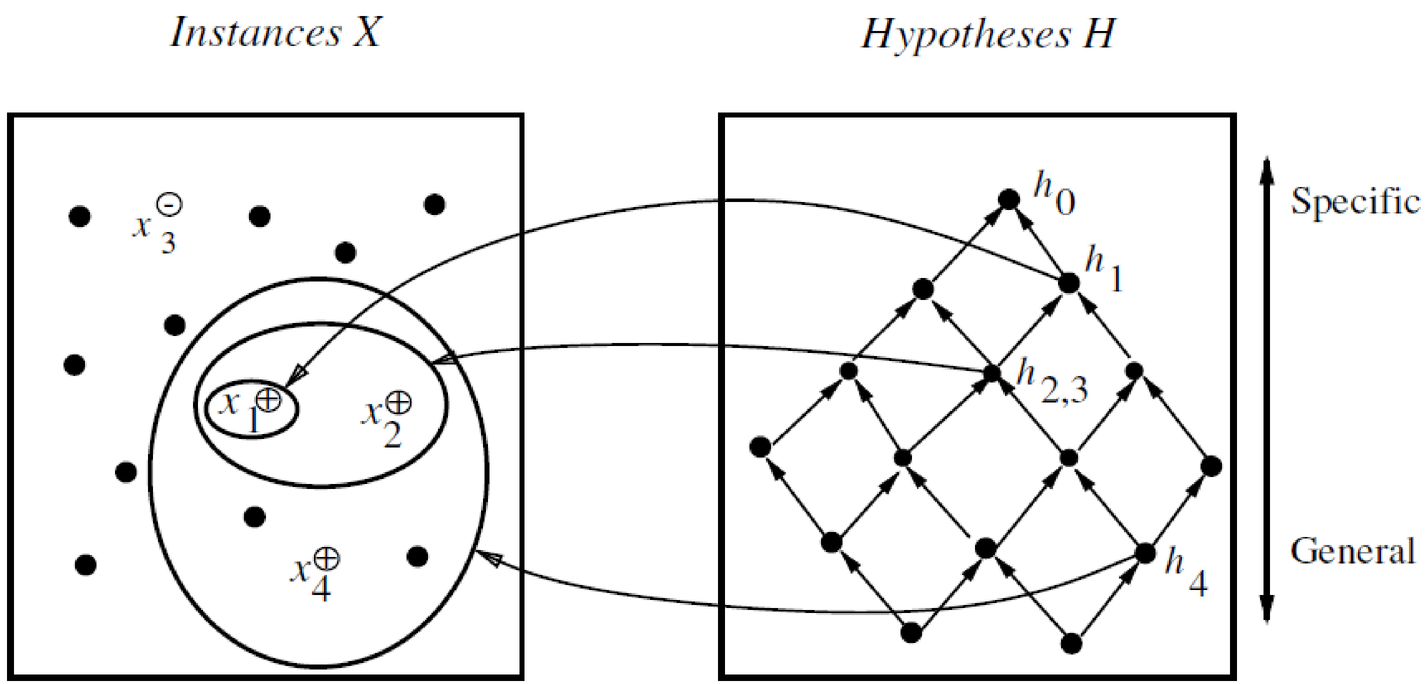
1. Initialize h to the most specific hypothesis in H
2. For each positive training instance x
   - For each attribute constraint a_i in h

     If the constraint a_i in h is satisfied by x

     Then do nothing

     Else replace a_i in h by the next more general constraint that is satisfied by x
3. Output hypothesis h

Find – S Example



h_0 = <∅, ∅, ∅, ∅, ∅, ∅>

x_1 = <*Sunny Warm Normal Strong Warm Same*>, +

x_2 = <*Sunny Warm High Strong Warm Same*>, +

x_3 = <*Rainy Cold High Strong Warm Change*>, -

x_4 = <*Sunny Warm High Strong Cool Change*>, +

h_1 = <*Sunny Warm Normal Strong Warm San*

h_2 = <*Sunny Warm ? Strong Warm Same*>

h_3 = <*Sunny Warm ? Strong Warm Same*>

h_4 = <*Sunny Warm ? Strong ? ?* >

Find – S Problems

Can’t tell whether it has learned the concept

Can’t tell whether the data is inconsistent

Picks maximally specific hypothesis

There might be several maximally specific hypothesis.

Version Space

A hypothesis h is **consistent** with a set of training examples D of target concept c if and only if $h(x) = c(x)$ for each training example $\langle x, c(x) \rangle$ in D.

$$Consistent(h, D) \equiv (\forall \langle x, c(x) \rangle \in D)\ h(x) = c(x)$$

The **version space**, $VS_{H,D}$, with respect to hypothesis space H and training examples D, is the subset of hypotheses from H consistent with all training examples in D.

$$VS_{H,D} \equiv \{h \in H | Consistent(h, D)\}$$

Version space representation

The **General boundary**, G, of version space $VS_{H,D}$ is the set of its maximally general members

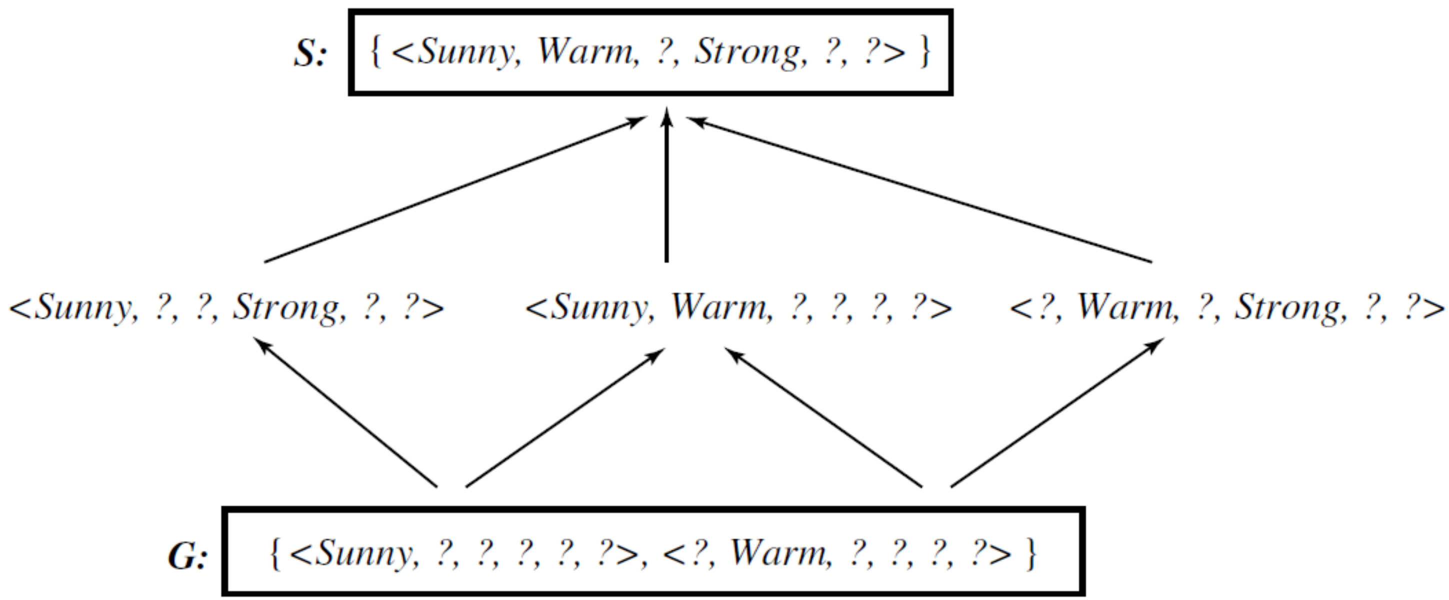
The **Specific boundary**, S, of version space $VS_{H,D}$ is the set of its maximally specific members

Every member of the version space lies between these boundaries

$$VS_{H,D} = \{h \in H | (\exists s \in S)(\exists g \in G)(g \geq h \geq s)\}$$

where $x \geq y$ means x is more general or equal to y

Version space



Candidate Elimination

$G \leftarrow$ maximally general hypotheses in H
$S \leftarrow$ maximally specific hypotheses in H
For each training example d, do

- If d is a positive example
  - Remove from G any hypothesis inconsistent with d
  - For each hypothesis s in S that is not consistent with d
    * Remove s from S
    * Add to S all minimal generalizations h of s such that
      1. h is consistent with d, and
      2. some member of G is more general than h
    * Remove from S any hypothesis that is more

Candidate Elimination

If *d* is a negative example:

- Remove from S any hypothesis inconsistent with d
- For each hypothesis g in G that is not consistent with d
  - $*$ Remove g from G
  - $*$ Add to G all minimal specializations h of g such that
    1. h is consistent with d, and
    2. some member of S is more specific than h
  - $*$ Remove from G any hypothesis that is less general than another hypothesis in G

Example Problem

Training Examples for EnjoySport

Sky	Temp	Humid	Wind	Water	Forecst	EnjoySpt
Sunny	Warm	Normal	Strong	Warm	Same	Yes
Sunny	Warm	High	Strong	Warm	Same	Yes
Rainy	Cold	High	Strong	Warm	Change	No
Sunny	Warm	High	Strong	Cool	Change	Yes

What is the general concept?

Example

Workout …

Convergence

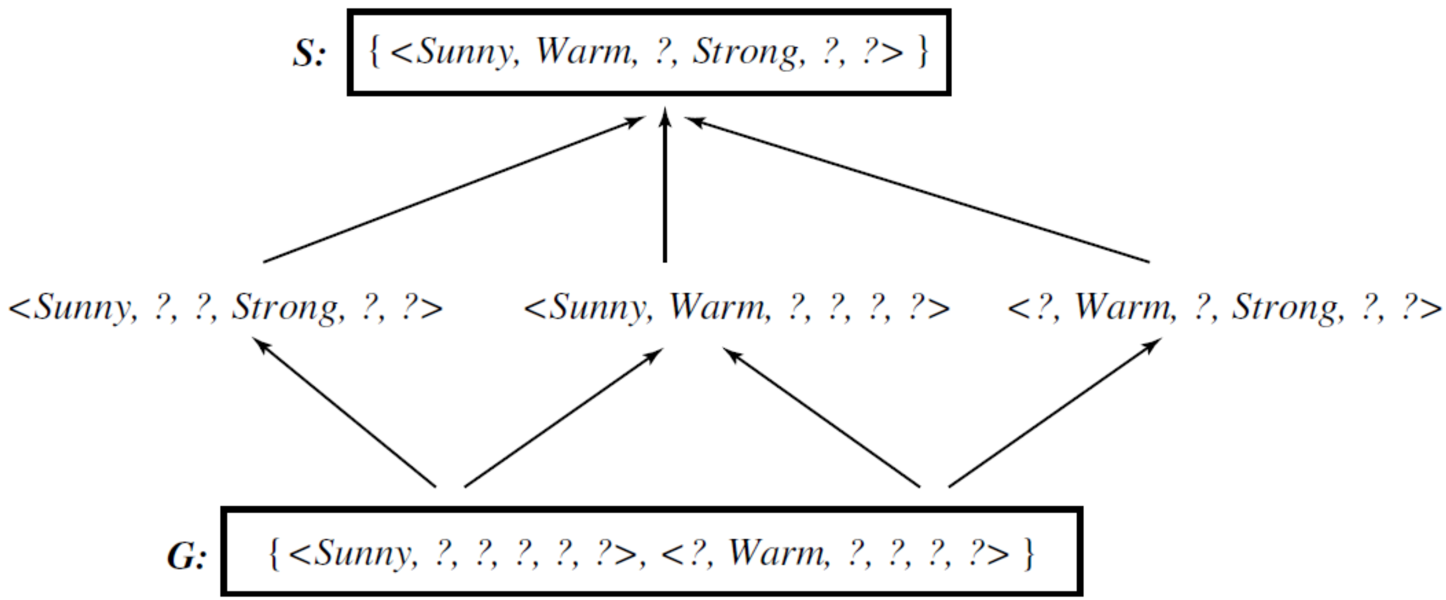
Candidate elimination will converge to the target concept if:

- Training data doesn't have errors.
- Target concept lies in the hypothesis space.

Otherwise

- G and S sets become null.

Partially learned concept

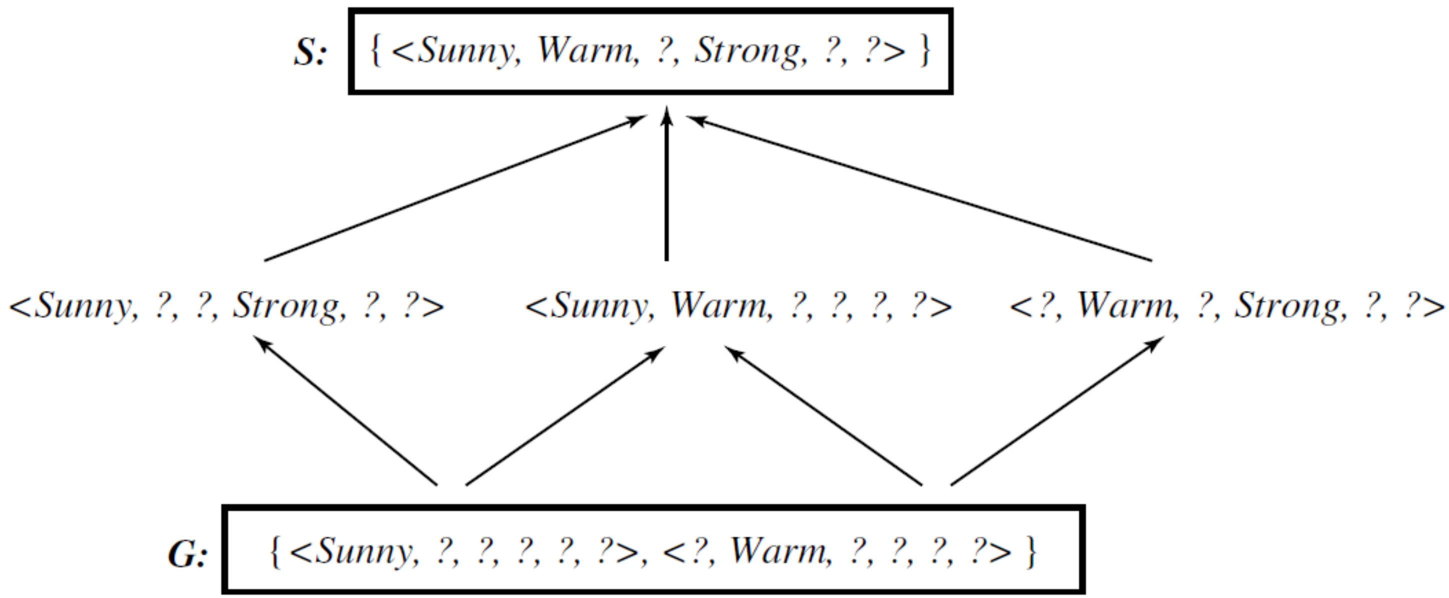


$$\langle Sunny\ Warm\ Normal\ Strong\ Cool\ Change \rangle$$

$$\langle Rainy\ Cool\ Normal\ Light\ Warm\ Same \rangle$$

$$\langle Sunny\ Warm\ Normal\ Light\ Warm\ Same \rangle$$

What next training example ?



<Sunny, Warm, Normal, Light, Warm, Same>

Observations

The hypothesis space is biased.

- Example: XOR concept cannot be expressed.

Unbiased learner – disjunction of conjunctions.

Learned Version space:

- S set: all positive examples
- G set: compliment of all negative examples

Can we use the partially learned concept from above ?

- There is perfect ambiguity for all examples not in training set.

Unbiased learning

Learning in an unbiased hypothesis space is futile as it cannot generalize to examples other than training examples.

End of Slides