Machine Learning Algorithms

Ms. Sandhya Harikumar,

Department of Computer Science & Engineering,

Amrita Vishwa Vidyapeetham, Amritapuri

Outline

➢Ill posed versus well posed problems

➢Inductive learning

➢ML algorithm versus Conventional algorithms

Machine Learning Definition

Definition 1 : Arthur Samuel(1959)

Field of study that gives computers the ability to learn without being explicitly programmed.

Definition 2 : Tom Mitchell (1998)

A computer program is said to learn from Experience E with respect to some Task T and some Performance measure P if performance on T, as measured by P, improves with experience E.

Ill posed versus Well posed Learning problems

- ➢ Ill posed problems :
   - No clear goal/Has simultaneous goals
   - No clear solution paths
   - No unique solution

   "Today is Sunday. Shreyas would like to watch a movie"

- ➢ Ill posed problems should be converted into a Well-Posed problem in terms of Task, Experience, and Performance

- ➢ Well posed problem
   - Solution exists
   - The solution is unique
   - The solution depends on the data/experience

   "Task : Classify whether Krish movie will be liked by Shreyas"

Conventional versus ML algorithms

- Enable computers to learn from data without being explicitly programmed.
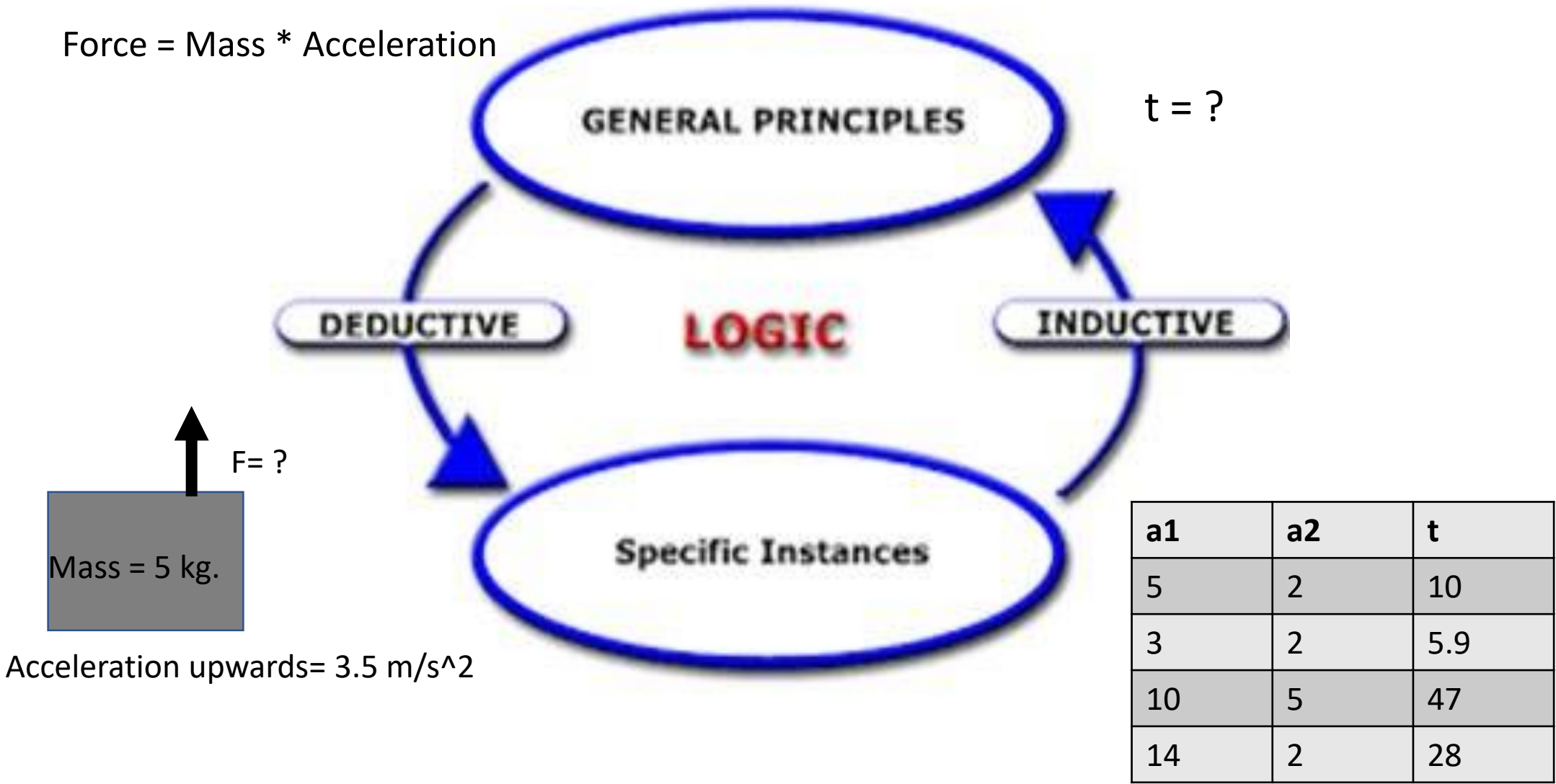
Traditional Programming

2,12,3,1,14

Sorting

Data → Computer → Output   1,2,3,12,14

Program →

Machine Learning

2,10,3,1,30,20

1,2,3,10,20,30

6,40,4,90,60,9
4,6,9,10,20,40

Data → Computer → Program

Output →

Generate Model

ai<aj; if i<j

$a_i < a_j$; if $i<=n/2$
$a_i = a_{(i-n/2)} * 10$; if $i>n/2$

Inductive versus Deductive Learning

Force = Mass * Acceleration

t = ?

GENERAL PRINCIPLES

DEDUCTIVE

LOGIC

INDUCTIVE

F= ?

Mass = 5 kg.

Acceleration upwards= 3.5 m/s^2

Specific Instances

a1	a2	t
5	2	10
3	2	5.9
10	5	47
14	2	28

Machine Learning Tasks

➢Regression   : Generalize to predict continuous valued output

➢ What will be share price of SBI bank after a month ?

➢Classification : Generalize to predict discrete valued output

➢ Are the documents given of sports or of technology

➢Clustering   : Generalize to form groups

➢ Given keywords of various documents, group the documents

Experience for Machine Learning : Data

➢Data : Qualitative or quantitative variables in organized/unorganized form
   ➢Images, Documents, Videos, Customers transactions

DATA-> Information-> Knowledge

➢Information : Meaningful data
   ➢Customers buying grocery items
   ➢The documents are related to sports…

➢Knowledge : Learned/Analyzed information
   ➢If milk is bought, then bread is also bought
   ➢Indian Cricketers are Sachin, Sehwag, Ganguly, Srinath, ….

Performance measure

➤ Accuracy : How many of the predictions made by the model are correct

➤ Precision : Fraction of relevant instances among the retrieved instances

➤ Recall : Fraction of the total amount of relevant instances that were retrieved actually

…………

precision: TP/cancer diagnoses

Diagnosis

	No cancer	Cancer
No cancer	TN	FP
Cancer	FN	TP

True state

recall: TP/cancer true states

Source : quora.com

Desired characteristics

- Learn and Generate models of high quality

- Generalize patterns with less error

- Learn from noisy and vague data

- Efficient and Scalable

- Deterministic

Takeaways

➢ Ill posed versus well posed problems

➢ ML intends to solve optimization problems that are well posed in the form of T,E,P

➢ ML is inductive learning

➢ How is ML different from Conventional algorithms

➢ Understanding of Tasks, Experience, Performance

➢ Desired characteristics of ML

References

- https://www.statisticshowto.com/well-posedness/
- https://www.houseofbots.com/news-detail/11973-1-clarifying-differences-between-data-analysis-data-mining-data-science-machine-learning,-and-big-data
- https://www.quora.com/What-is-the-best-way-to-understand-the-terms-precision-and-recall