INTRODUCTION TO MACHINE LEARNING

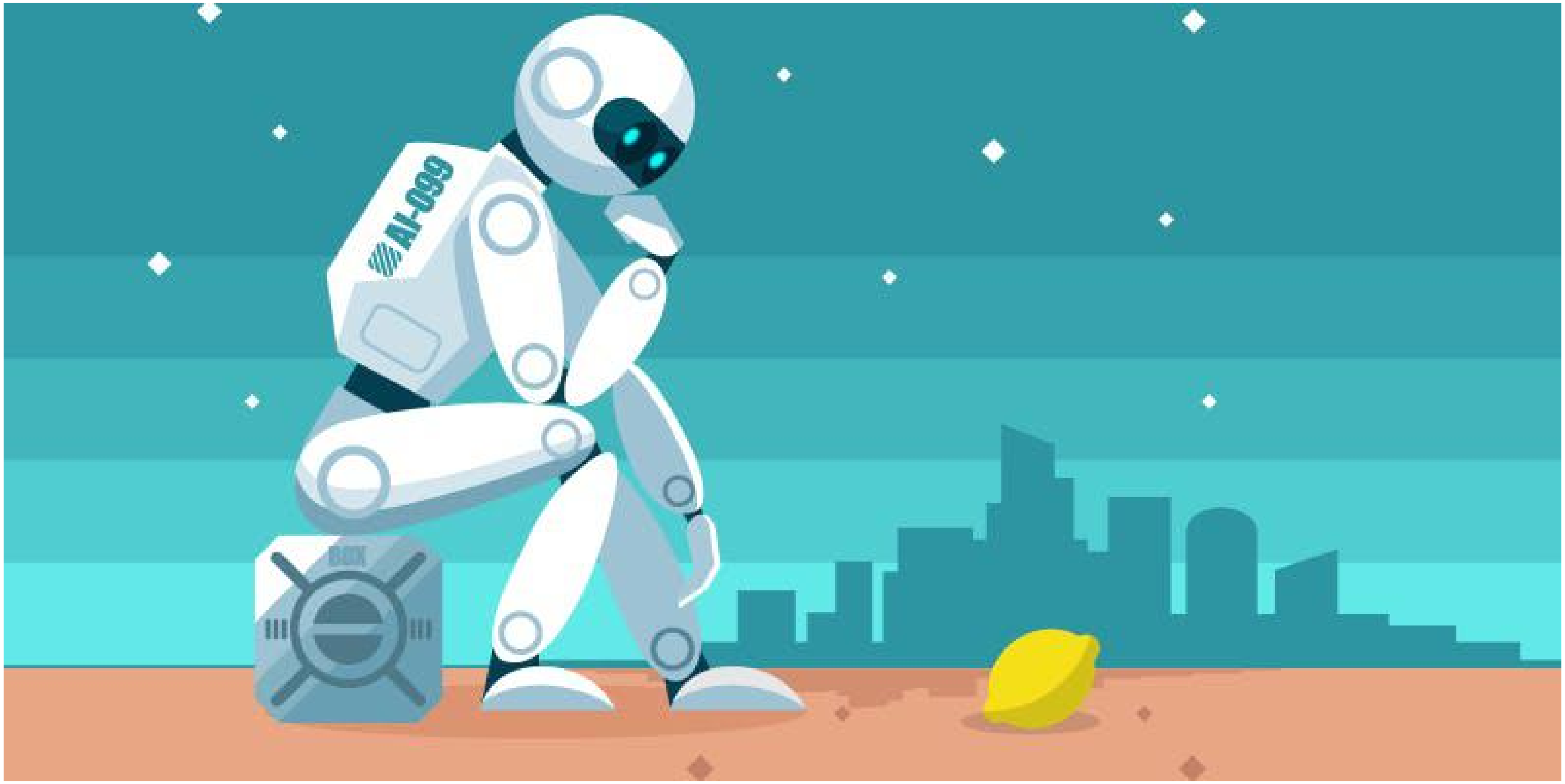
Ms. Krishna Modi, DCS

Fields of Data Science

- Data Mining
- Machine Learning
- Artificial Intelligence

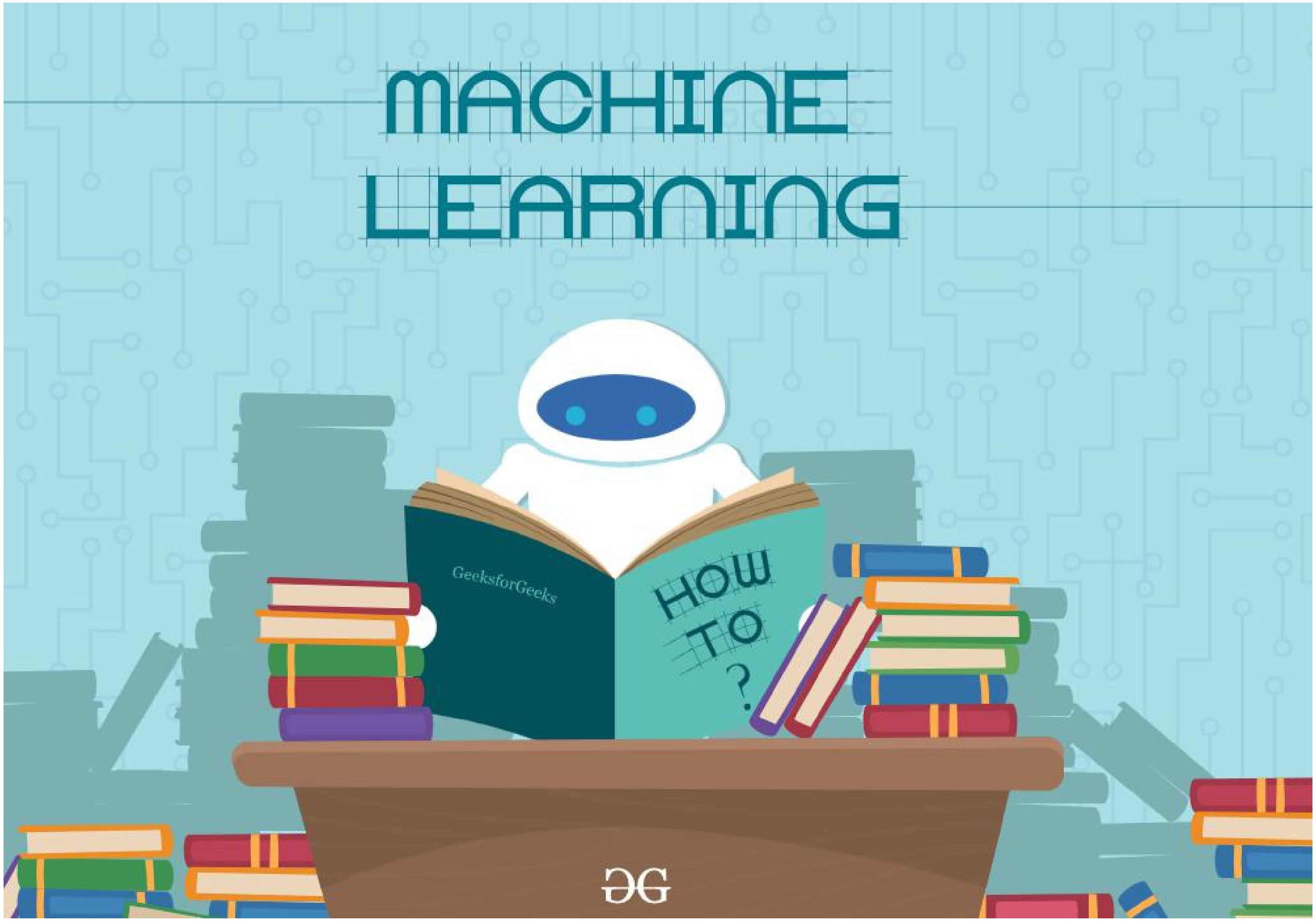
	Description	Computer Programming Skill
Statistics	◦ Quantify data ▪ Statistics is just about the numbers, and quantifying the data.	◦ Pure mathematics
Data Mining	◦ Find patterns, explain phenomenon ▪ Using Statistics as well as other programming methods to find patterns hidden in the data so that you can explain some phenomenon.	◦ More towards math than programming
Machine Learning	◦ Build models to predict future ▪ Using Data Mining techniques and other learning algorithms to build models of what is happening behind some data so that it can predict future outcomes.	◦ More towards programming
Artificial Intelligence	◦ Reason about the world to have intelligent behavior ▪ Using models built by Machine Learning and other ways to reason about the world and give rise to intelligent behavior	◦ Very programming based

Artificial Intelligence
(When Machine Starts Thinking)



Ms. Krishna Modi, DCS

Machine Learning
(When Machine Starts Learning)



Ms. Krishna Modi, DCS

What is learning?

Using past experiences to improve future performance.

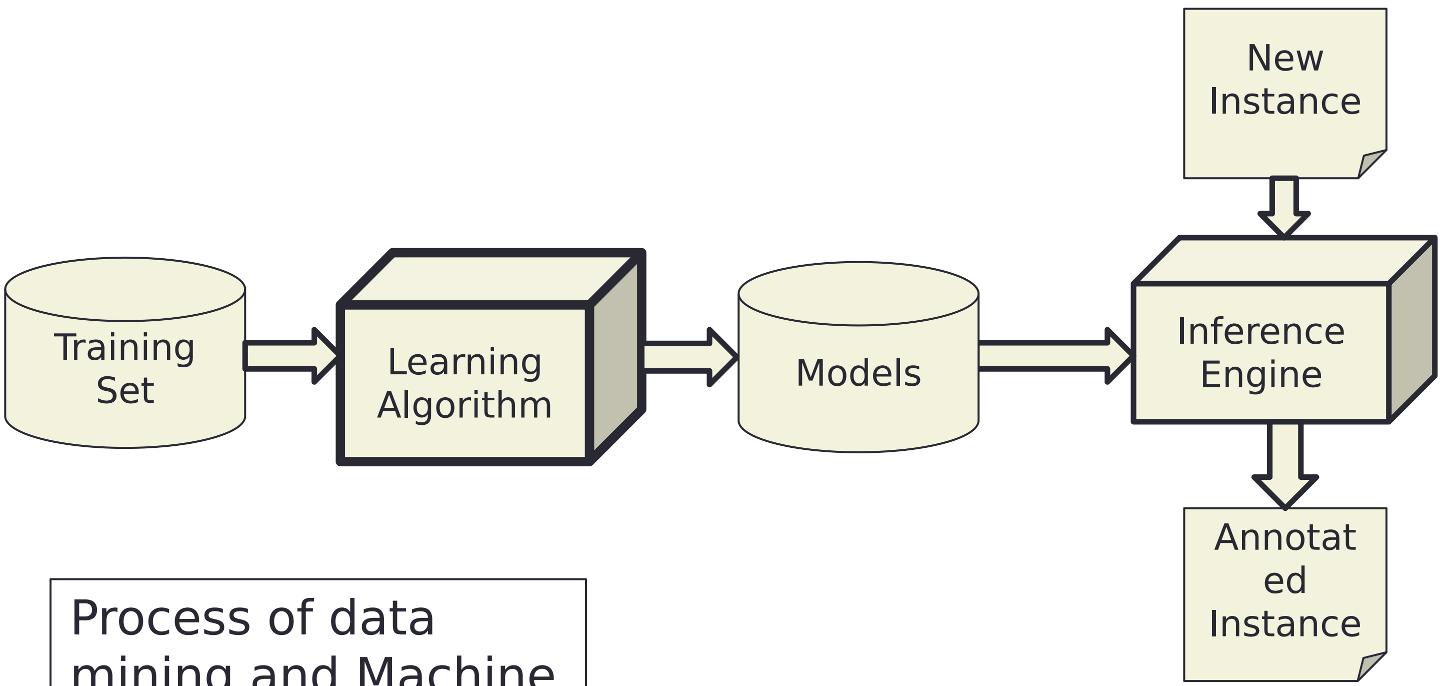
For a machine, experiences come in the form of data.

Machine Learning

- Machine Learning is the sub-field of data science that focus on designing algorithms that can learn from and make prediction on data.

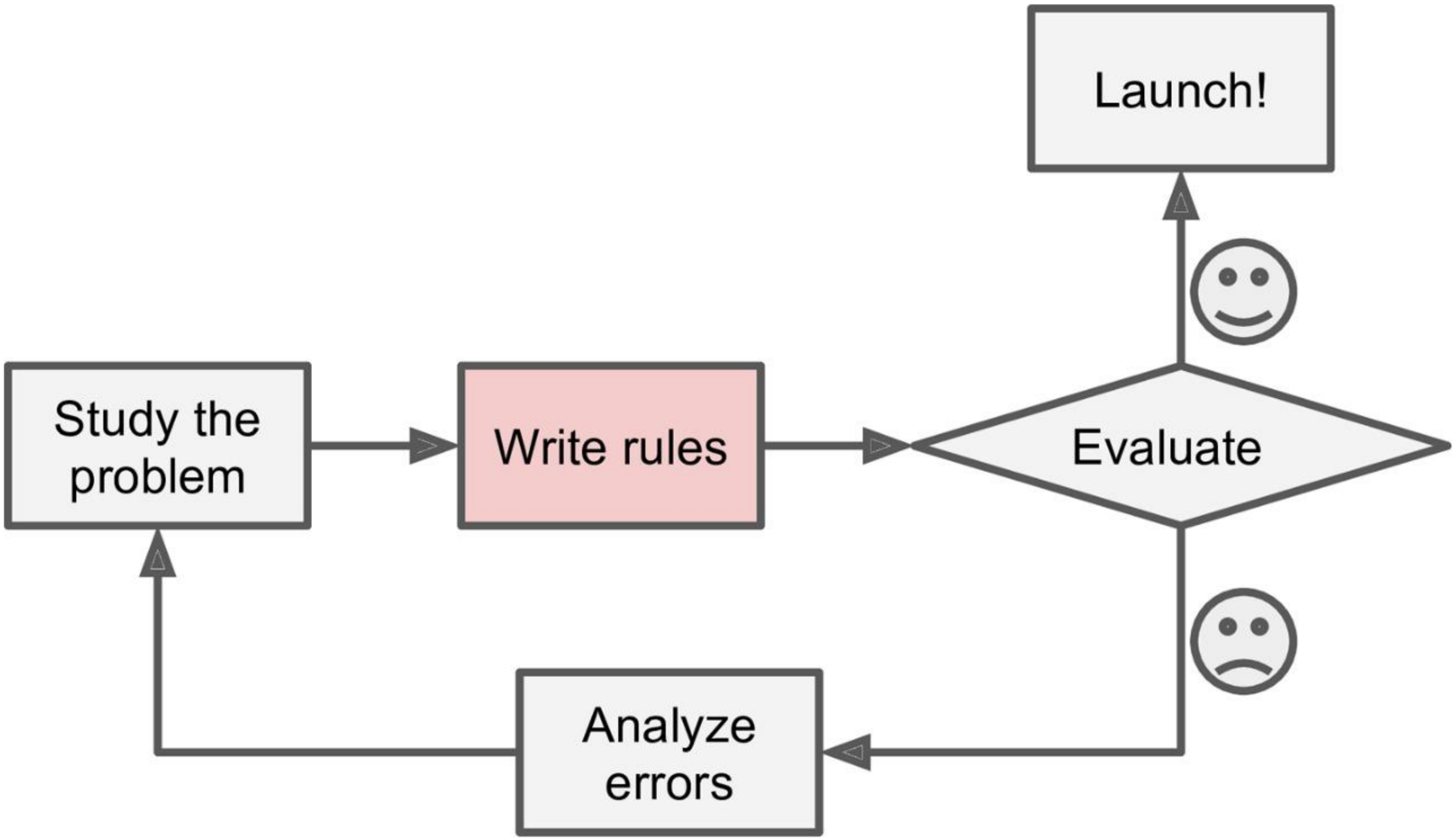
- field of study that gives computers the ability to learn without being explicitly programmed.

Introduction of Machine Learning

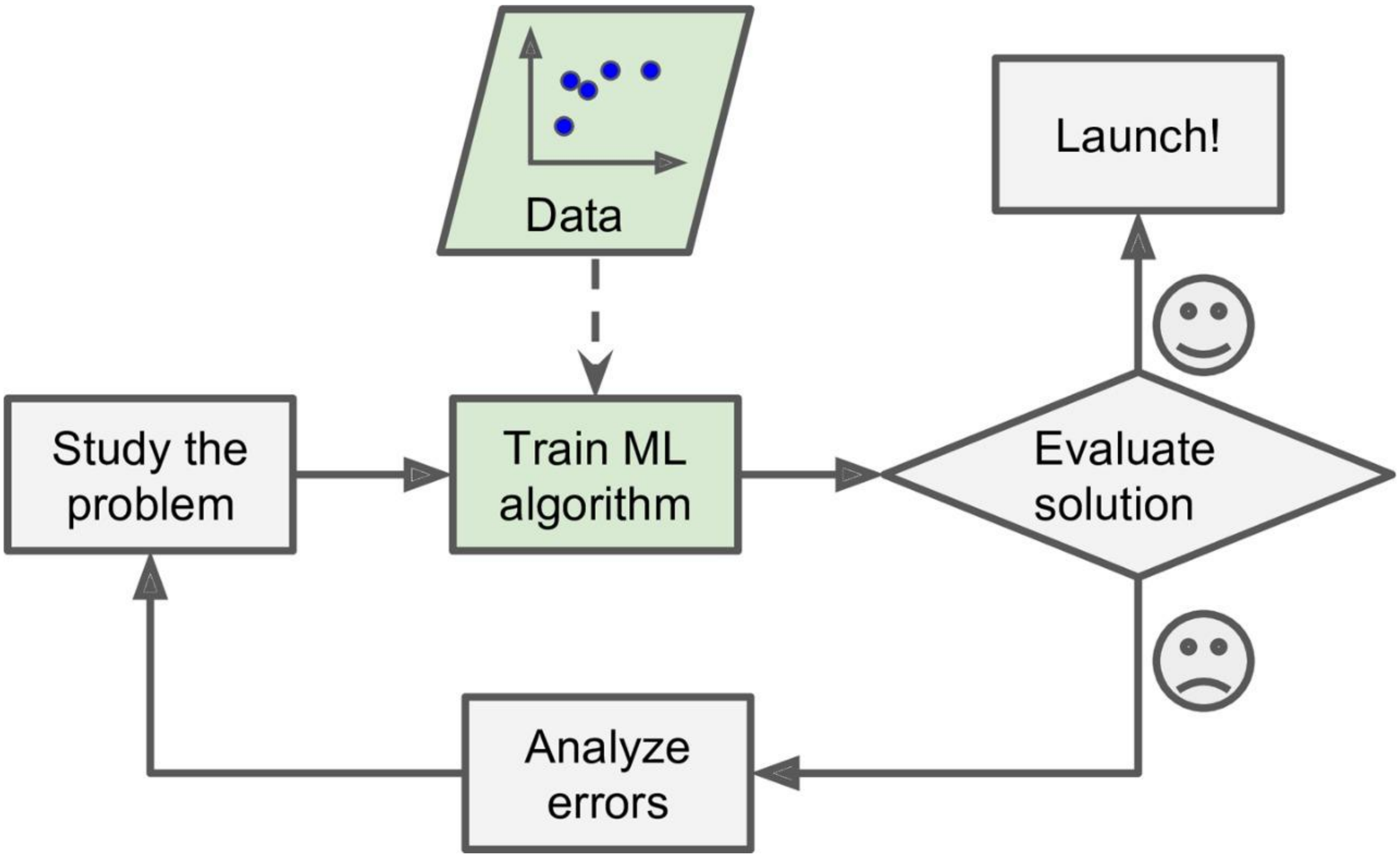


Process of data mining and Machine Learning

Traditional Approach



Machine Learning Approach



Why Machine Learning?

- We need computers to make informed decisions on new, unseen data.
- Often it is too difficult to design a set of rules “by hand”.
- Machine learning is about automatically extracting relevant information from data and applying it to analyze new data.

Videos

- [Video 1]
- [Video 2]

APPLICATIONS AND GAMES

Ms. Krishna Modi, DCS

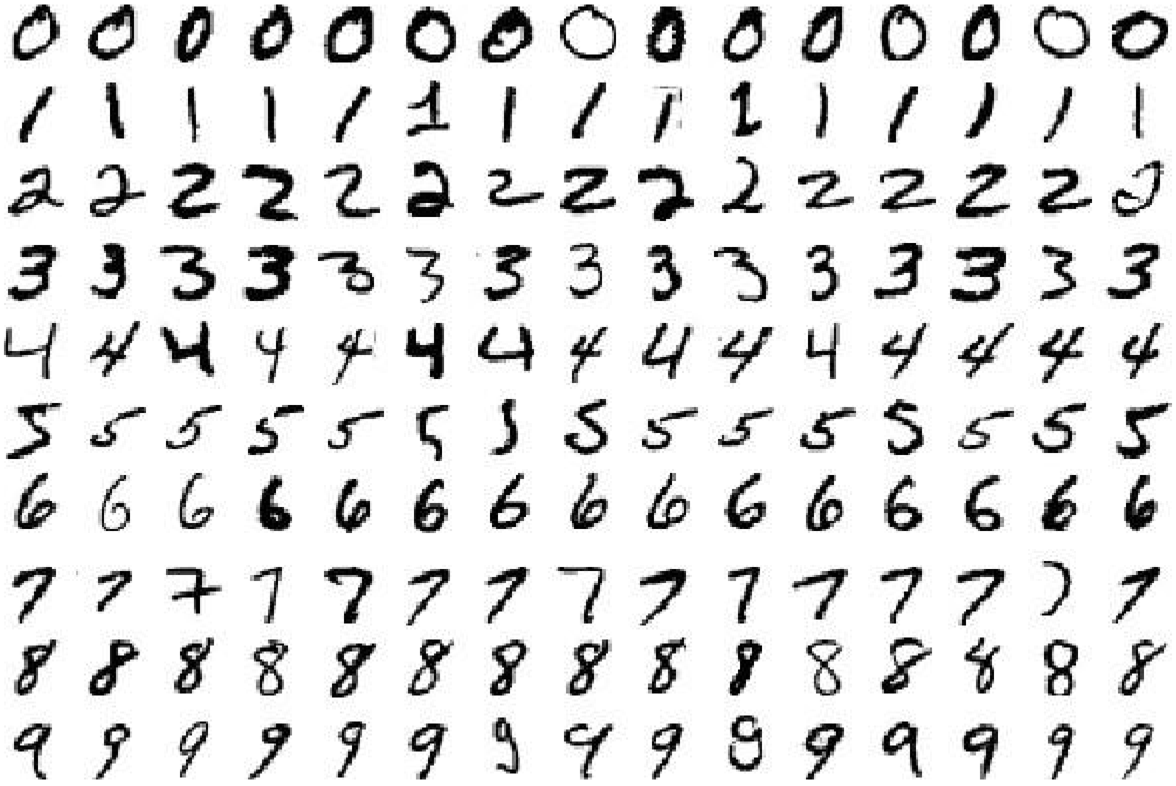
Smart Speakers

Amazon Echo

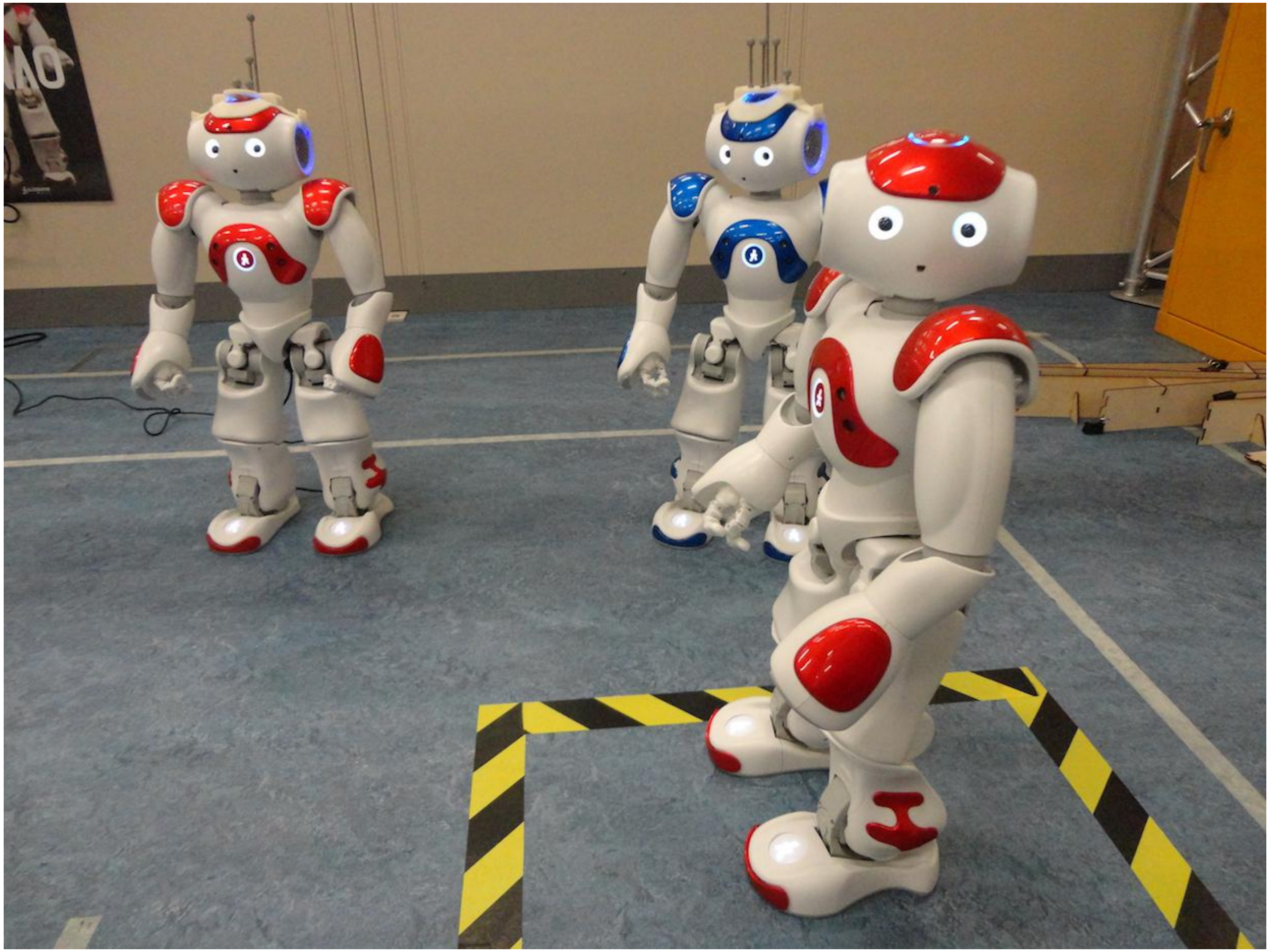
Google Home Mini

Ms. Krishna Modi, DCS

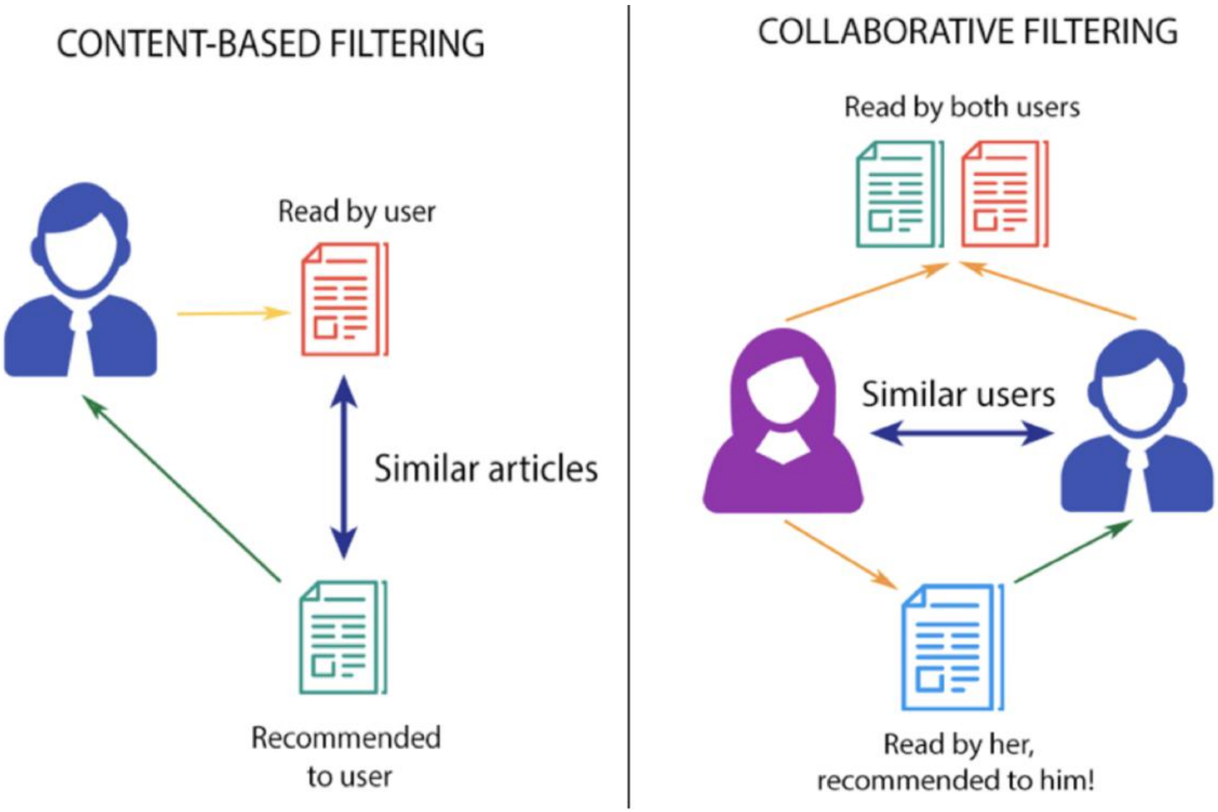
Text Recognition (Image Recognition)



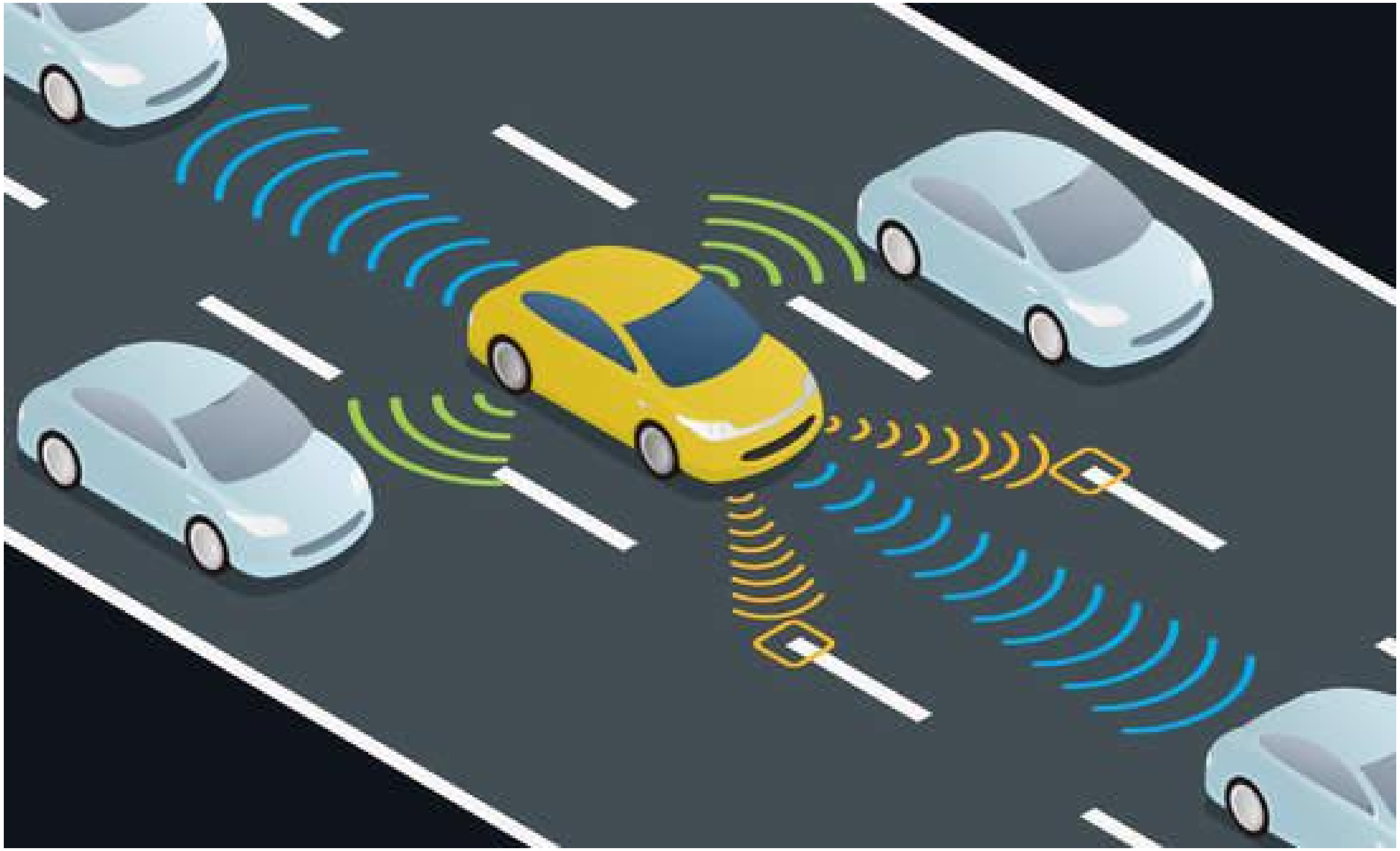
Intelligent Robot



Recommendation System

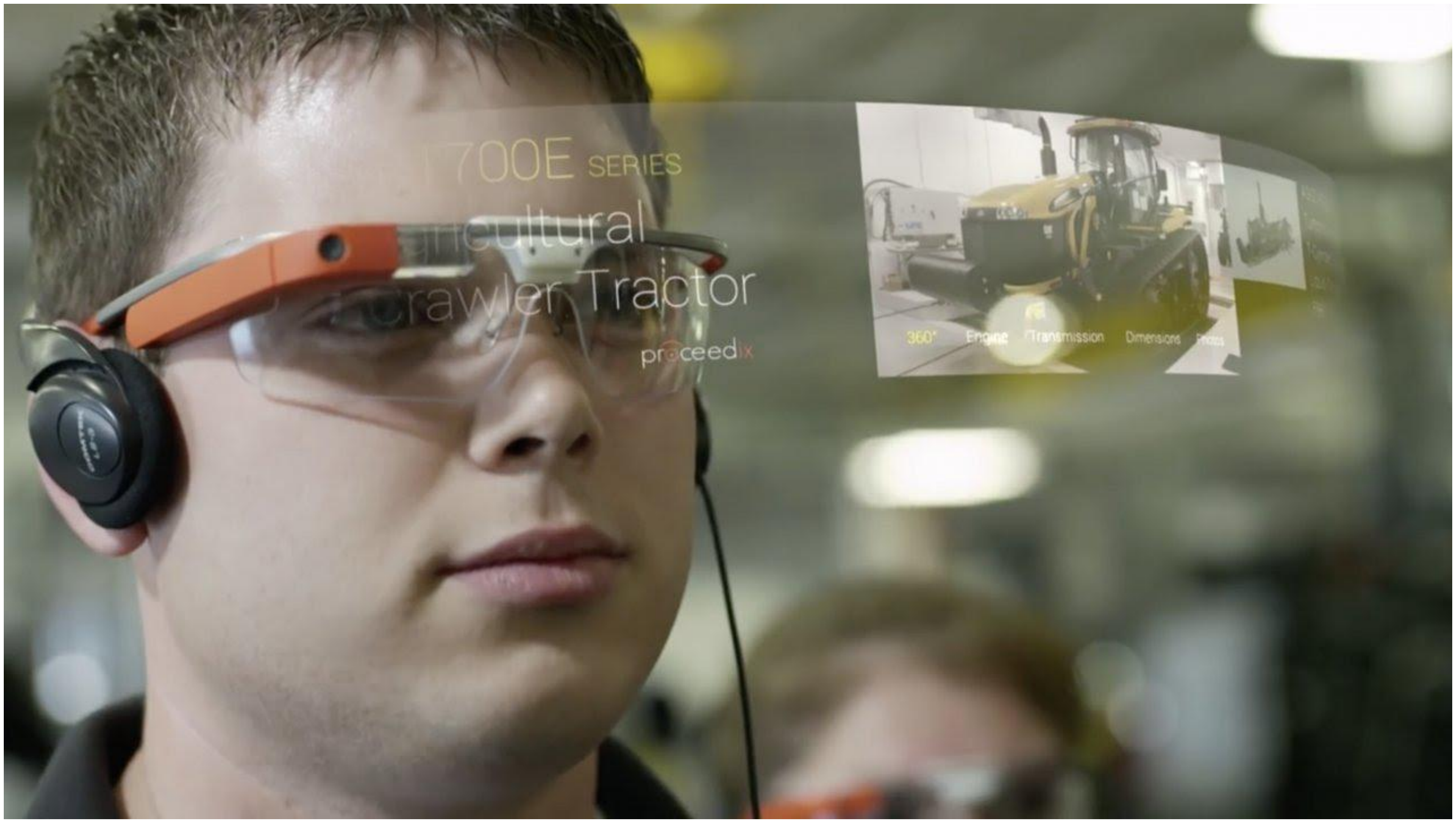


Self-Driving Car

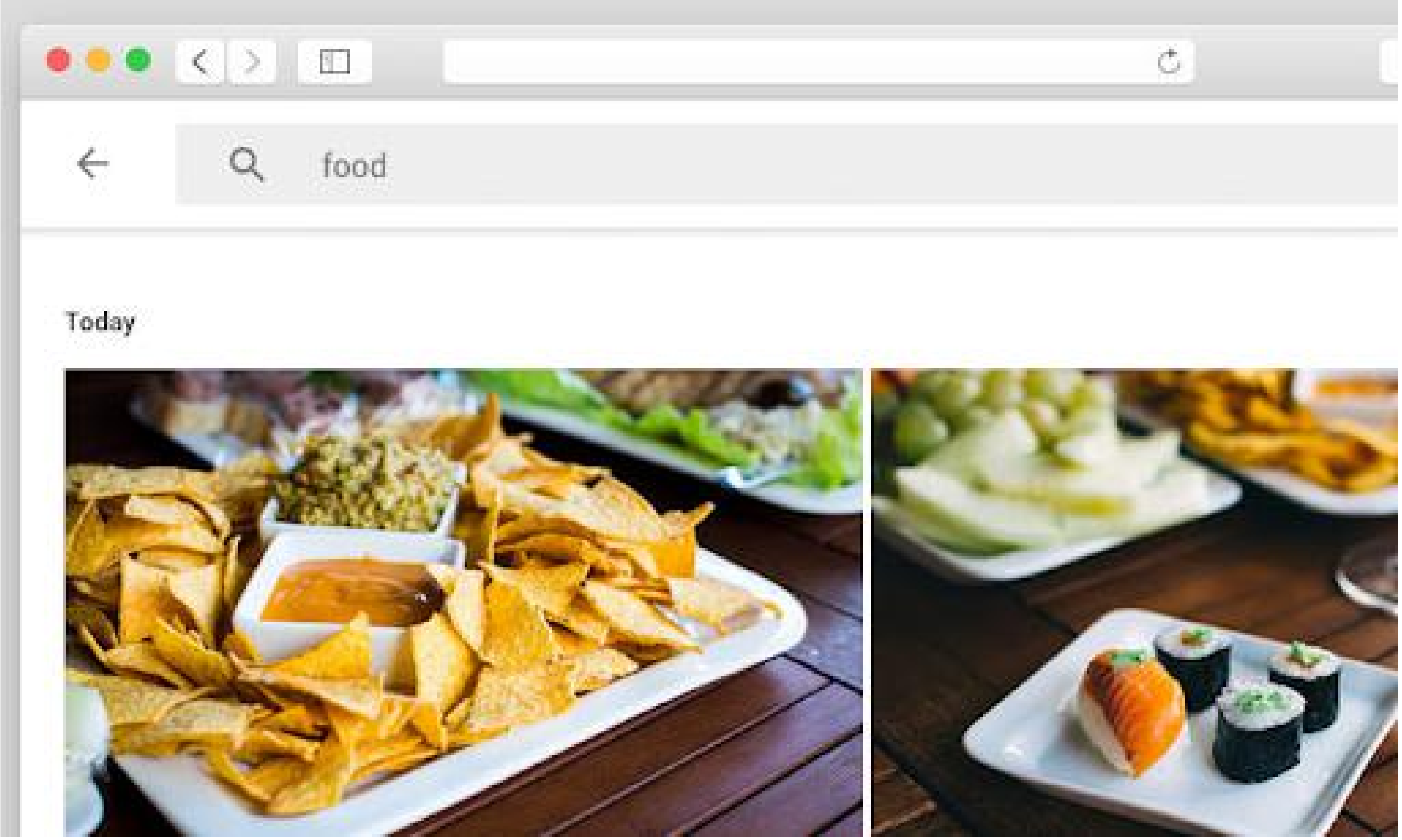


Ms. Krishna Modi, DCS

Smart Glass



Google Photos (Face, Image, Place Recognition)



Applications of ML

Computer vision and robotics:
– detection, recognition and categorization of objects
– face recognition
– tracking objects (rigid and articulated) in video
– modeling visual attention

- **Speech recognition**
- **Biology and medicine**:

– drug discovery
– computational genomics (analysis and design)
– medical imaging and diagnosis

- **Financial industry:**

– Fraud detection
– Credit approval
– Price and market prediction

- **Information retrieval, Web search, Google ads...**

What is Learning problem?

- A computer program is said to learn from experience **E** with respect to some class of tasks ***T*** and performance measure **P**, if its performance at tasks in **T**, as measured by **P**, improves with experience **E**.

What is Learning problem?

Learning = Improving with experience at some task

– Improve over task T

– with respect to performance measure P

– based on experience E

Example : Object Categorization

A handwriting recognition learning problem:

Task T: recognizing and classifying handwritten words within images.

Performance measure *P:* percent of words correctly classified.

Training experience E: a database of handwritten words with given classifications.

A robot driving learning problem:

- **Task T**: driving on public four-lane highways using vision sensors

- **Performance measure P**: average distance traveled before an error (as judged by human overseer)

- **Training experience E**: a sequence of images and steering commands recorded while observing a human driver

Design a learning System

- **Data Acquisition**
- **Data Exploration**
- **Modeling**
- **Testing / Evaluation**

DATA ACQUISITION

Ms. Krishna Modi, DCS

Data Acquisition

1. Collection of relevant data.
   - From data warehouses
   - From sensors

2. Data Transformation

3. Data Cleaning

– Get rid of errors, noise, Removal of redundancies.

4. Missing value treatment.

Sensors

Device, module or subsystem whose purpose is to detect events or changes in its environment and send the information.

- Light sensor
- Sound sensor
- Temperature Sensor
- Contact Sensor
- Proximity Sensor (Range sensor)
- Pressure sensor
- Biometric sensor

Missing Value Treatment

- Missing data in the training data set can reduce the power / fit of a model or can lead to a biased model.
- It can lead to wrong prediction or classification.

Methods to treat missing value

- Ignoring the tuple (Deletion)
- Fill in missing value manually
- Use global constant to fill in the missing value
- Averaging Technique

Ignoring the tuple (Deletion)

User	Device	OS	Transactions
~~A~~	~~Mobile~~	~~NA~~	~~5~~
B	Mobile	Android	3
~~C~~	~~NA~~	~~iOS~~	~~2~~
D	Tablet	Android	1
E	Mobile	iOS	4

Averaging Technique

OS	Revenue
Android	1,804
iOS	3,027
iOS	8,788
Android	NA
Android	3,735
Android	1,056
iOS	9,319
Android	6,199
Android	2,235
iOS	NA
Android	1,146

OS	Global Mean	Group Mean
Android	1,804	1,804
iOS	3,027	3,027
iOS	8,788	8,788
Android	4,145	2,696
Android	3,735	3,735
Android	1,056	1,056
iOS	9,319	9,319
Android	6,199	6,199
Android	2,235	2,235
iOS	4,145	7,045
Android	1,146	1,146

Data Exploration

Ms. Krishna Modi, DCS

WHY DATA EXPLORATION?

Data Quality – accuracy, consistency and completeness

Data Exploration

Steps involved in data Exploration

- Variable Identification
- Outlier Analysis
- Variable creation

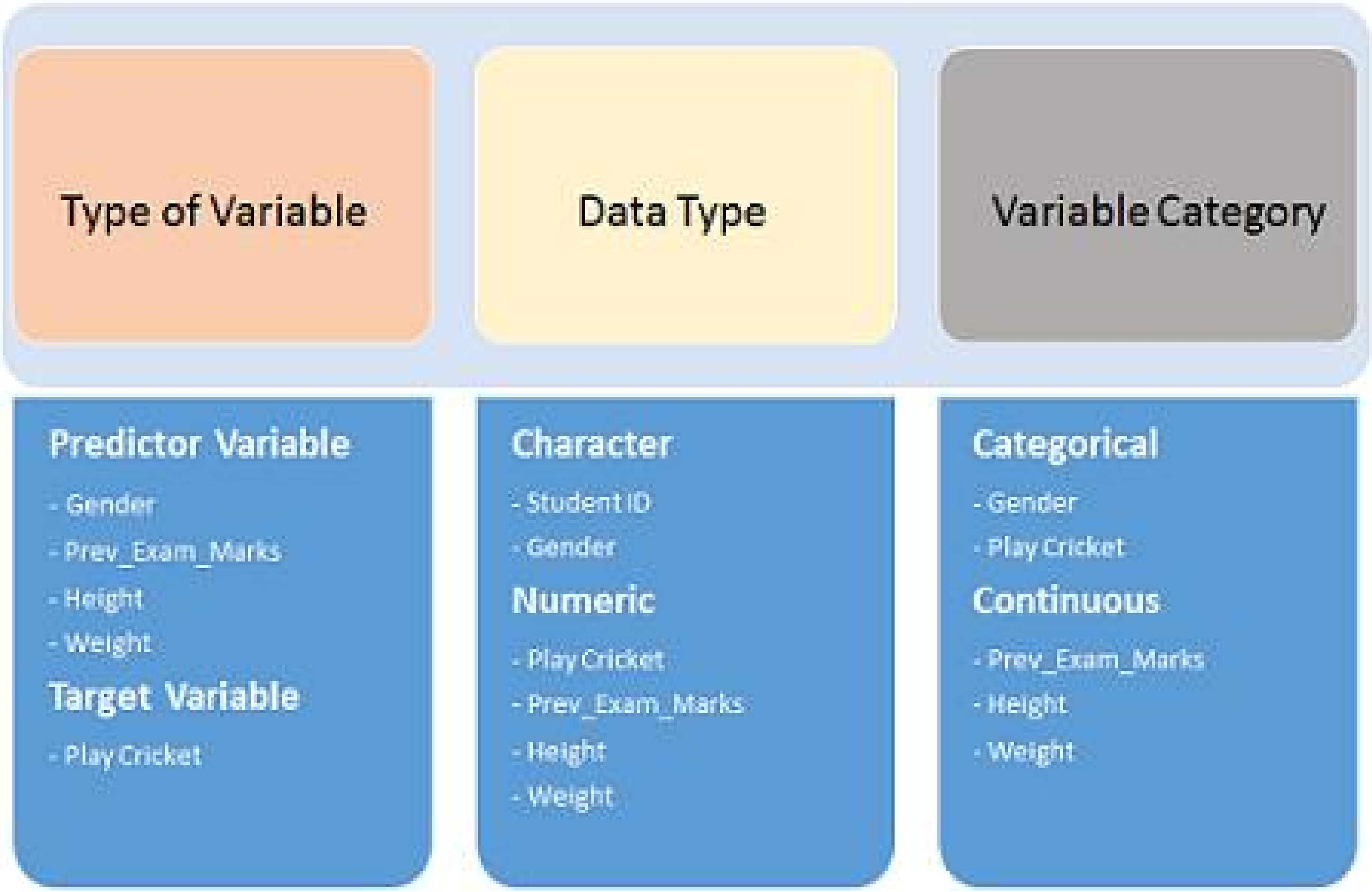
Variable Identification

- First, identify **Predictor** (Input) and **Target** (output) variables.
- Next, identify the data type and category of the variables.

Example

Suppose, we want to predict, whether the students will play cricket or not (refer below data set). Here you need to identify predictor variables, target variable, data type of variables and category of variables.

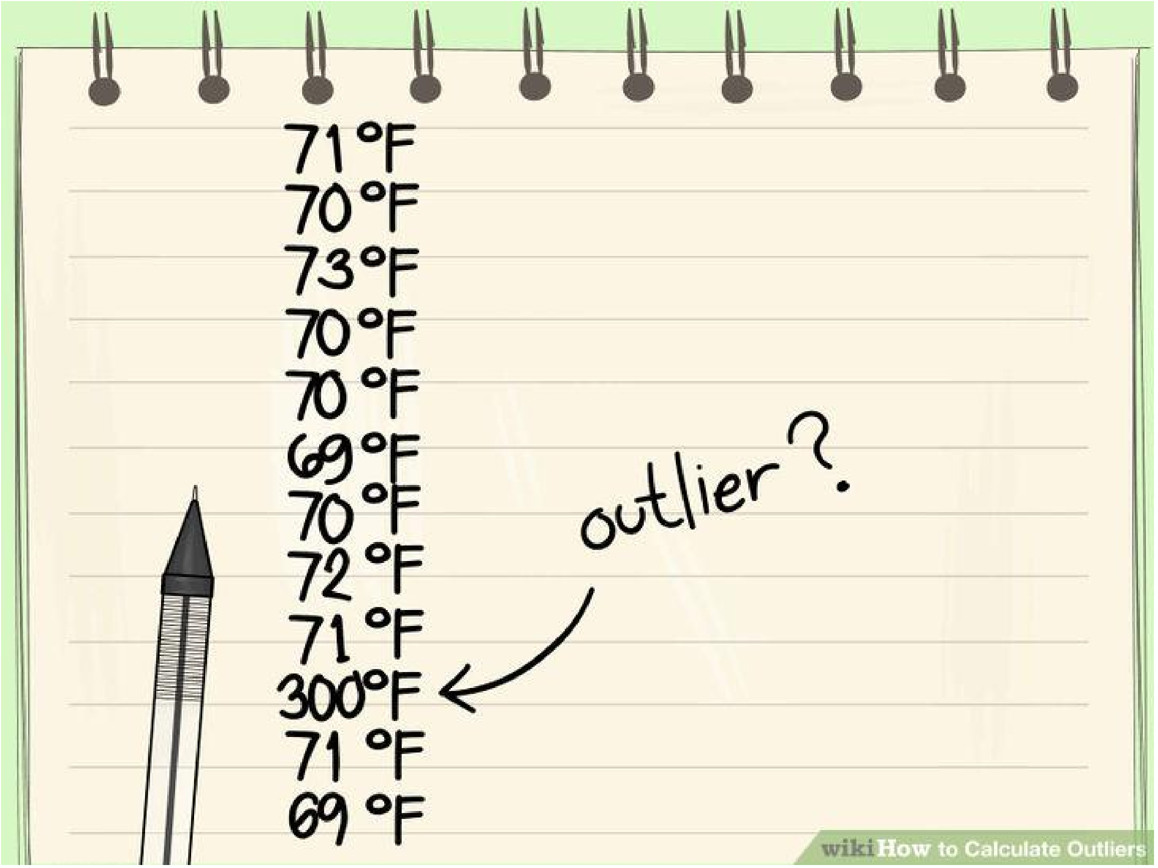
Student_ID	Gender	Prev_Exam_Marks	Height (cm)	Weight Caregory (kgs)	Play Cricket
S001	M	65	178	61	1
S002	F	75	174	56	0
S003	M	45	163	62	1
S004	M	57	175	70	0
S005	F	59	162	67	0


Type of Variable
Data Type
Variable Category
Predictor Variable
- Gender
- Prev_Exam_Marks
- Height
- Weight
Target Variable
- Play Cricket
Character
- Student ID
- Gender
Numeric
- Play Cricket
- Prev_Exam_Marks
- Height
- Weight
Categorical
- Gender
- Play Cricket
Continuous
- Prev_Exam_Marks
- Height
- Weight

OUTLIER ANALYSIS

Outlier is defined as an object that deviates from other objects.

Outlier Analysis



Medical Diagnosis

Outlier in ECG data (representing second degree heart block)

Various types of outlier

- Data Entry Error
- Measurement Error
- Natural Outlier

Impact of outlier on dataset

Without outlier

4 4 5 5 5 5 6 6 6 7 7

With Outlier

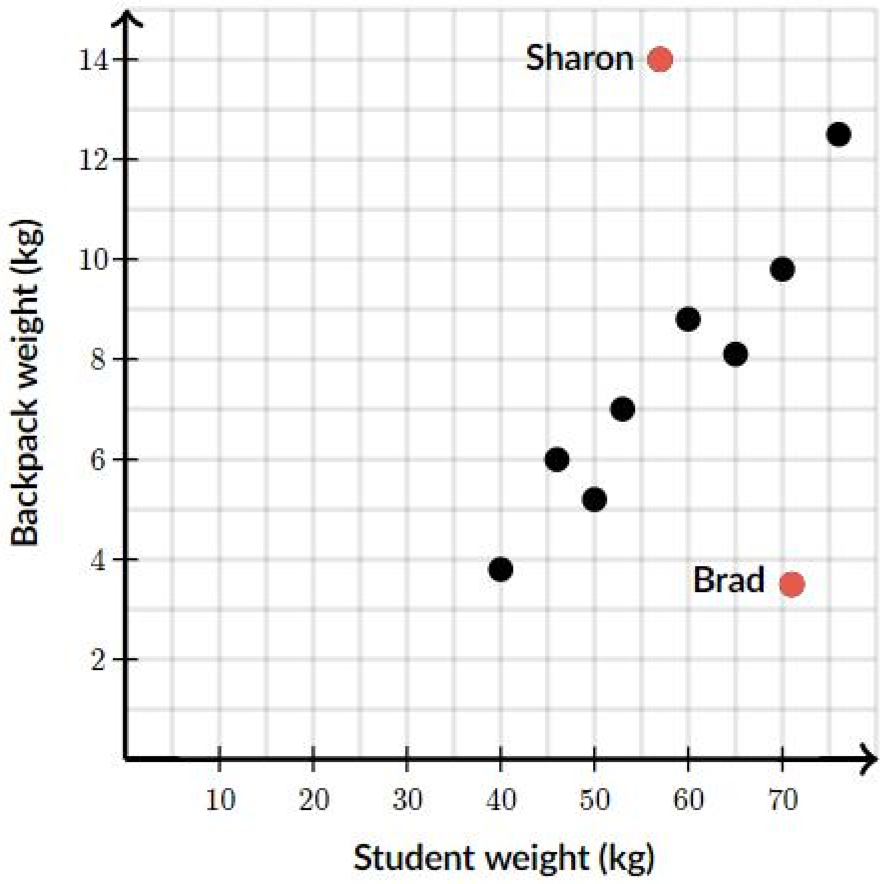
4 4 5 5 5 5 6 6 6 7 7 300

Find mean, Median and mode.

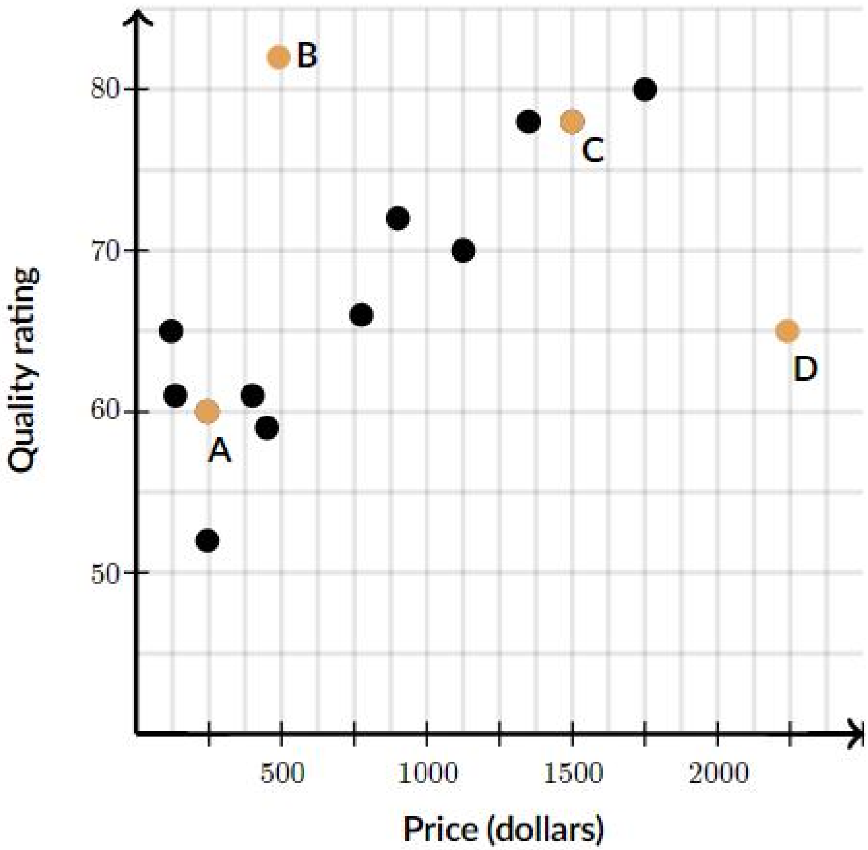
Impact of Outliers on a dataset

Without Outlier	With Outlier
4, 4, 5, 5, 5, 5, 6, 6, 6, 7, 7	4, 4, 5, 5, 5, 5, 6, 6, 6, 7, 7,300
Mean = 5.45	Mean = 30.00
Median = 5.00	Median = 5.50
Mode = 5.00	Mode = 5.00

Scatter Plot



Example 1 : Computer Shopping



Variable Creation

Creating Derived Variable

Emp_Code	Gender	Date	New_Day	New_Month	New_Year
A001	Male	21-Sep-11	21	9	2011
A002	Female	27-Feb-13	27	2	2013
A003	Female	14-Nov-12	14	11	2012
A004	Male	07-Apr-13	7	4	2013
A005	Female	21-Jan-11	21	1	2011
A006	Male	26-Apr-13	26	4	2013
A007	Male	15-Mar-12	15	3	2012

Variable Creation

Categorical variable into numeric value

Emp_Code	Gender	Var_Male	Var_Female
A001	Male	1	0
A002	Female	0	1
A003	Female	0	1
A004	Male	1	0
A005	Female	0	1
A006	Male	1	0
A007	Male	1	0

Scatter Point

Money Invested	Profit
60	3.1
61	3.6
62	3.8
63	4
65	4.1

Data Modelling

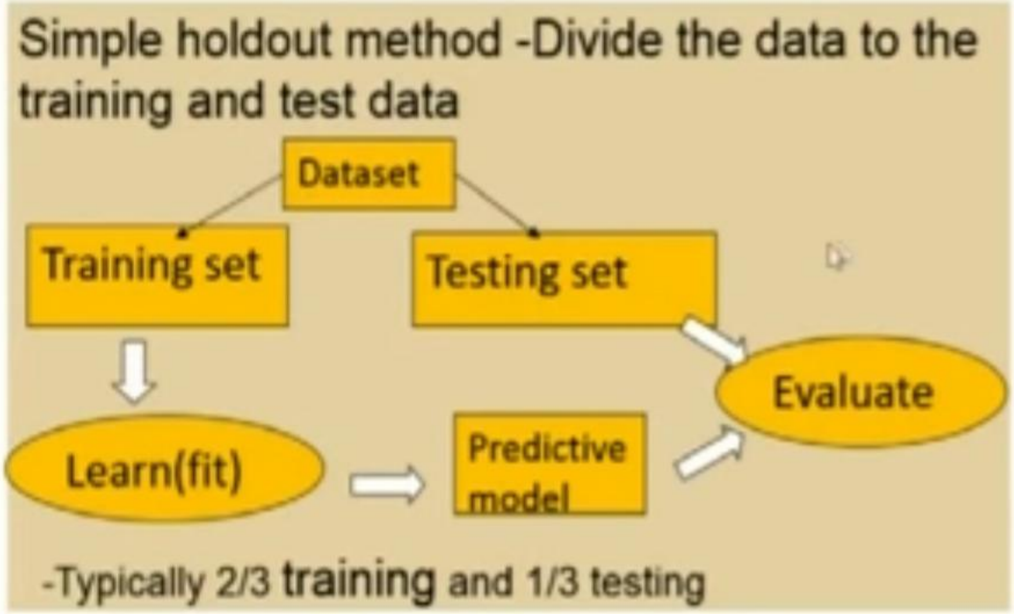
- **Data modeling** (**data modelling**) is the analysis of **data** objects and their relationships to other **data** objects.
- **Data modeling** is the process of producing a descriptive diagram of relationships between various types of information

Data Modeling

- Linear Regression
- Decision Tree
- Neural Network
- Bayesian Learning

Testing or Evaluation

- Apply learned model to new data.
- Predict output for new inputs using learned function.
- Evaluate on test data.



DATA NORMALIZATION

Transform the data into smaller or common range such as [-1,1] or [0,1]

Data Normalization methods

1. Min-Max normalization
2. Z-score normalization
3. Decimal scaling

Min – Max Normalization

Performs linier transformation on the original data.

$$v' = \frac{v - \min_A}{\max_A - \min_A}(new\_\max_A - new\_\min_A) + new\_\min_A$$

Min – Max Normalization new range - [0,1]

Marks	Marks after min-max normalization
8	0
10	0.16
15	0.58
20	1

Z-Score Normalization

$$z_i = \frac{x_i - \bar{x}}{s}$$

$\bar{x}$: average
S: standard deviation

Formula of standard deviation

$$\mathrm{SD} = \sqrt{\frac{\sum |x - \bar{x}|^2}{n}}$$

Decimal scaling

Salary bonus	Normalized after decimal scaling
40000	0.31
35000	0.35
31000	0.31

Issues in Machine Learning

Which algorithms perform best for which types of problems and representations?

Issues in Machine Learning

How much training data is sufficient?

Issues in Machine Learning

How does number of training examples influence accuracy?

Issues in Machine Learning

What is the best strategy for choosing a useful next training experience ?

Issues in Machine Learning

How does noisy data influence accuracy?

Issues in Machine Learning

How to develop a theoretical understanding of algorithms?

Issues in Machine Learning

Scale for distributed big data.

Issues in Machine Learning

Extract information with unlabeled data. Still is it useful?

CONCEPT LEARNING

Acquiring the definition of a general category from given sample positive and negative training examples of the category.

Ms. Krishna Modi, DCS

Concept learning - Example

- learning of bird-concept from the given examples of birds (positive examples) and non-birds (negative examples).
- Enjoy sports or not from given situations

A Concept Learning Task – Enjoy Sport

Training Examples

Example	**Sky**	**AirTemp**	**Humidity**	**Wind**	**Water**	**Forecast**	**EnjoySport**
1	Sunny	Warm	Normal	Strong	Warm	Same	**YES**
2	Sunny	Warm	High	Strong	Warm	Same	**YES**
3	Rainy	Cold	High	Strong	Warm	Change	**NO**
4	Sunny	Warm	High	Strong	Warm	Change	**YES**

ATTRIBUTES **CONCEPT**

A Concept Learning Task – Enjoy Sport

- A set of example days, and each is described by six attributes.
- The task is to learn to predict the value of EnjoySport for arbitrary day, based on the values of its attribute values.
- Each hypothesis consists of a conjunction of constraints on the instance attributes.
- Each hypothesis will be a vector of six constraints, specifying the values of the six attributes – (Sky, AirTemp, Humidity, Wind, Water, and Forecast).

A Concept Learning Task – Enjoy Sport

Hypothesis Representation

Each attribute will be:

- **?** - indicating any value is acceptable for the attribute **(don’t care)**
- **single value** – specifying a single required value (ex. Warm) **(specific)**
- Ø - indicating no value is acceptable for the attribute **(no value)**

A Concept Learning Task – Enjoy Sport

Hypothesis Representation

- Most general hypothesis: ?, ?, ?, ?, ?, ?ﬂ
- Most specific hypothesis: Ø, Ø, Ø, Ø, Ø, Øﬂ

Sport example

- Concept to be learned:
  Days in which Aldo can enjoy water sport
  Attributes:
  Sky: *Sunny, Cloudy, Rainy* ***Wind***: *Strong, Weak*
  AirTemp: *Warm, Cold* ***Water***: *Warm, Cool*
  Humidity: *Normal, High* ***Forecast***: *Same, Change*

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

Hypotheses representation

- *h* is a set of constraints on attributes:
  - a specific value: e.g. *Water = Warm*
  - any value allowed: e.g. *Water* = ?
  - no value allowed: e.g. *Water* = Ø
- Example hypothesis:

  Sky AirTemp Humidity Wind Water Forecast

  Sunny, ?, ?, *Strong,* ?, *Same*ﬂ

  Corresponding to boolean function:

  Sunny(Sky) ∧ Strong(Wind) ∧ Same(Forecast)

Hypothesis satisfaction

- An instance x *satisfies* an hypothesis h iff all the constraints expressed by h are satisfied by the attribute values in x.

- Example 1:

  x_1: *Sunny, Warm, Normal, Strong, Warm, Same*ﬁ

  h_1: *Sunny*, ?, ?, *Strong*, ?, *Same*ﬁ

  Satisfies? Yes

- Example 2:

  x_2: *Sunny, Warm, Normal, Strong, Warm, Same*ﬁ

  h_2: *Sunny*, ?, ?, Ø, ?, *Same*ﬁ Satisfies? No

FIND-S

FINDING A MAXIMALLY SPECIFIC HYPOTHESIS

Ms. Krishna Modi, DCS

Find-S

- Begin with the most specific possible hypothesis in H, then generalize this hypothesis each time it fails to cover an observed positive training example.
- Most specific Hypothesis.
- Consider only positive examples.

Find-S Algorithm

1. Initialize h to the most specific hypothesis

    H={Ø, Ø, Ø, …, Ø}

2. For each positive training instance ***x***

    For each attribute constraint *a,* in h

        If the attribute value=hypothesis value

            Then do nothing

        Else replace *hypothesis value* with the more general constraint ‘?’

3. Output hypothesis h

Find-S example

Example	Sky	AirTemp	Humidity	Wind	Water	Forecast	EnjoySport
1	Sunny	Warm	Normal	Strong	Warm	Same	YES
2	Sunny	Warm	High	Strong	Warm	Same	YES
3	Rainy	Cold	High	Strong	Warm	Change	NO
4	Sunny	Warm	High	Strong	Warm	Change	YES

ATTRIBUTES CONCEPT

Find-S example

Ex.	Sky	Temp	Humid	Wind	Water	Forecast	Enjoy Sport?	Hypothesis
0								$\langle \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset \rangle$

Find-S example

Ex.	Sky	Temp	Humid	Wind	Water	Forecast	Enjoy Sport?	Hypothesis
0								⟨ ∅, ∅, ∅,∅,∅,∅ ⟩
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes	⟨ Sunny,Warm,Normal,Strong,Warm,Same⟩

Find-S example

Ex.	Sky	Temp	Humid	Wind	Water	Forecast	EnjoySport	Hypothesis
0								⟨ ∅, ∅, ∅,∅,∅,∅ ⟩
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes	⟨ Sunny,Warm,Normal,Strong,Warm,Same⟩
2	Sunny	Warm	High	String	Warm	Same	Yes	⟨ Sunny,Warm,?,Strong,Warm,Same⟩

Find-S example

Ex.	Sky	Temp	Humid	Wind	Water	Forecast	Enjoy Sport?	Hypothesis
0								⟨ ∅, ∅, ∅,∅,∅,∅ ⟩
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes	⟨ Sunny,Warm,Normal, Strong,Warm,Same⟩
2	Sunny	Warm	High	String	Warm	Same	Yes	⟨ Sunny,Warm,?,Strong,Warm,Same⟩
3	Rainy	Cold	High	Strong	Warm	Change	No	⟨ Sunny,Warm,?,Strong,Warm,Same⟩
4	Sunny	Warm	High	Strong	Cool	Change	Yes	⟨ Sunny,Warm,?,Strong,?,?⟩

Example : 2 Concept – Finding Malignant tumor from MRI Scan

Concept – Malignant Tumor

Shape	Size	Color	Surface	Thickness	Tumor_prediction
Circular	Large	Light	Smooth	Thick	malignant
Circular	Large	Light	Irregular	Thick	malignant
Oval	Large	Dark	Smooth	Thin	Not malignant
Oval	Large	Light	Irregular	Thick	malignant
Circular	Small	Light	Smooth	thick	Not malignant

Example 3 : Concept - Japanese Economy Car

Origin	Manufacturer	Color	Decade	Type	Example Type
Japan	Honda	Blue	1980	Economy	Positive
Japan	Toyota	Green	1970	Sports	Negative
Japan	Toyota	Blue	1990	Economy	Positive
USA	Chrysler	Red	1980	Economy	Negative
Japan	Honda	White	1980	Economy	Positive
Japan	Toyota	Green	1980	Economy	Positive

CANDIDATE ELIMINATION ALGORITHM

The Candidate elimination algorithm finds all describable hypotheses that are consistent with the observed training examples.

Candidate Elimination

- Finds version space
- Consider both positive and negative results.
- This algorithm represents the set of *all* hypotheses consistent with the observed training examples.

Version Space

- Set of all hypothesis is called as “Version Space”.
- The **version space,** with respect to hypothesis space ***H*** and training examples D, is the subset of hypotheses from ***H*** consistent with the training examples in D.

$$\boldsymbol{VS} = \{\, h \in \boldsymbol{H} \,/\, \boldsymbol{Conistent(h, D)} \}$$

Candidate Elimination Algorithm

1. Generalize G and S as most general and most specific hypothesis.
2. For each example e,

    if e is +ve,

        make specific Hypothesis more general. [Find-S]

    else

        make general hypothesis more specific.

3. End

Candidate elimination algorithm

Input: training set

Output:

- G = maximally general hypotheses in H
- S = maximally specific hypotheses in H

Algorithm:

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
      - (a) h is consistent with d, and
      - (b) some member of G is more general than h
    - ∗ Remove from S any hypothesis that is more general than another hypothesis in S

Candidate elimination algorithm

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

(a) h is consistent with d, and

(b) some member of S is more specific than h

∗ Remove from G any hypothesis that is less general than another hypothesis in G

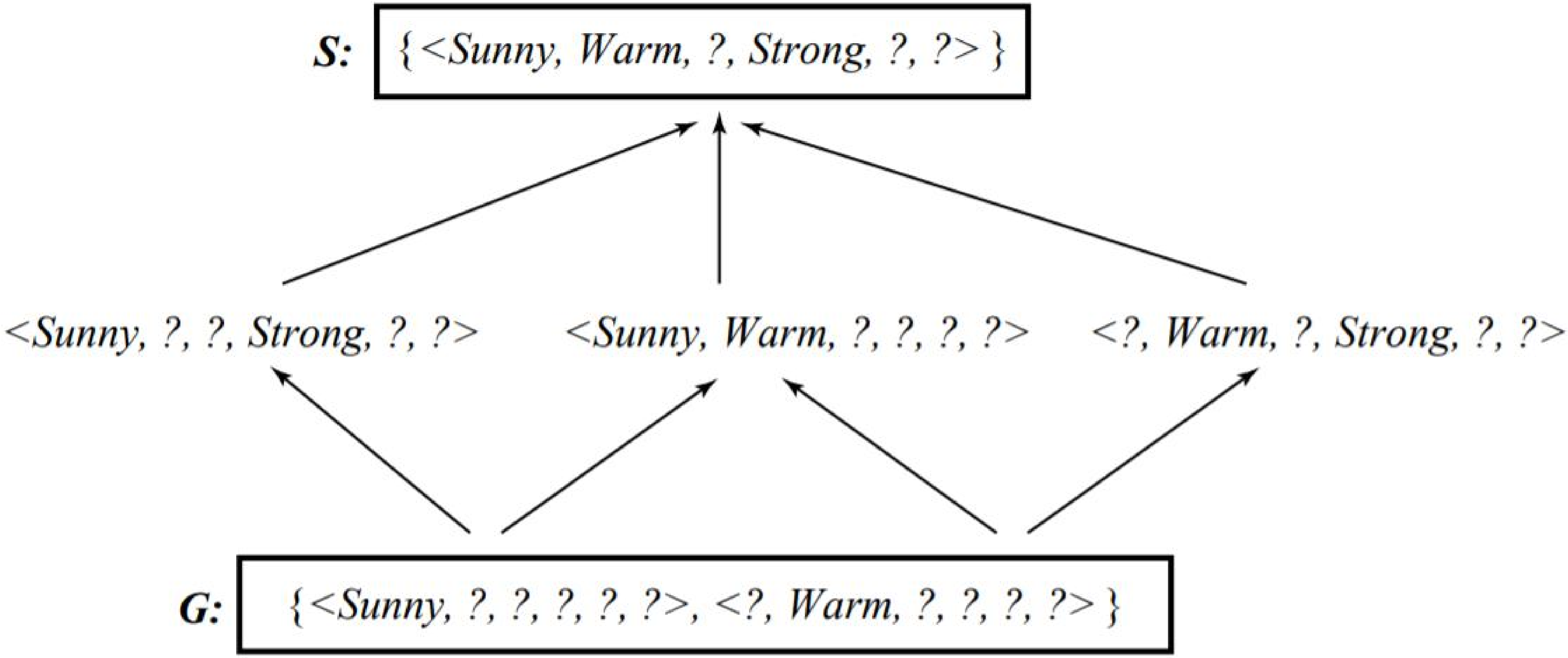
Example

Example	Sky	AirTemp	Humidity	Wind	Water	Forecast	EnjoySport
1	Sunny	Warm	Normal	Strong	Warm	Same	YES
2	Sunny	Warm	High	Strong	Warm	Same	YES
3	Rainy	Cold	High	Strong	Warm	Change	NO
4	Sunny	Warm	High	Strong	Warm	Change	YES

ATTRIBUTES

CONCEPT

Version space



Example

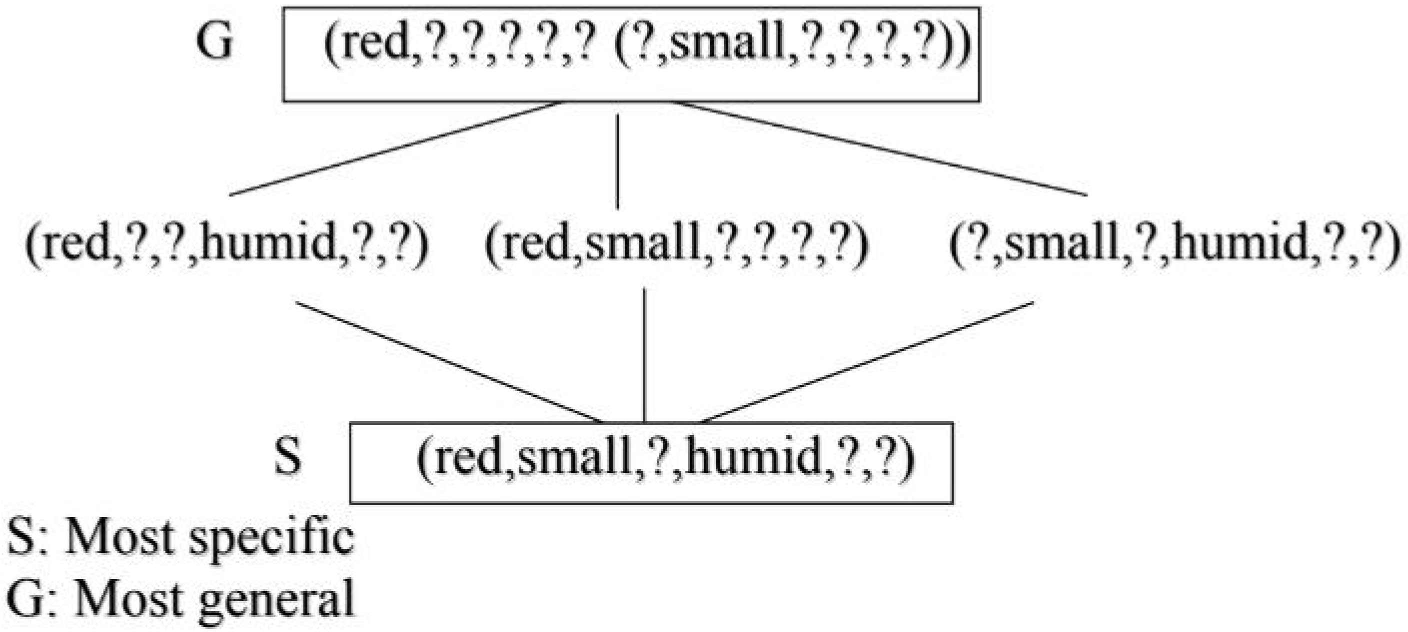
((red,small,round,humid,low,smooth), poisonous)

((red,small,elongated,humid,low,smooth), poisonous)

((gray,large,elongated,humid,low,rough), not-poisonous)

((red,small,elongated,humid,high,rough), poisonous)

Version Space – previous example



The inductive learning assumption

- We can at best guarantee that the output hypothesis fits the target concept over the training data
- *Assumption*: an hypothesis that approximates well the training data will also approximate the target function over unobserved examples
- i.e. given a significant training set, the output hypothesis is able to make predictions

Thank you..!!

References

1. Tom M Mitchell, “Machine Learning”, McGraw Hill