LCC-SJCE

Machine Learning
Session-1

Machine Learning – Definition

- Tom Mitchell (1998) Well-posed Learning Problem:

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.

The concept of learning in a ML system

- **Learning** = Improving with experience at some task
  - Improve over **task *T***,
  - With respect to **performance measure, *P***
  - Based on **experience, *E***.

- Motivating Example
  Learning to Filter Spam
  (Spam v/s Non-Spam)

Motivating Example
Learning to Filter Spam

Example: Spam Filtering

Spam - is all email the user does not want to receive and has not asked to receive

T: Identify Spam Emails

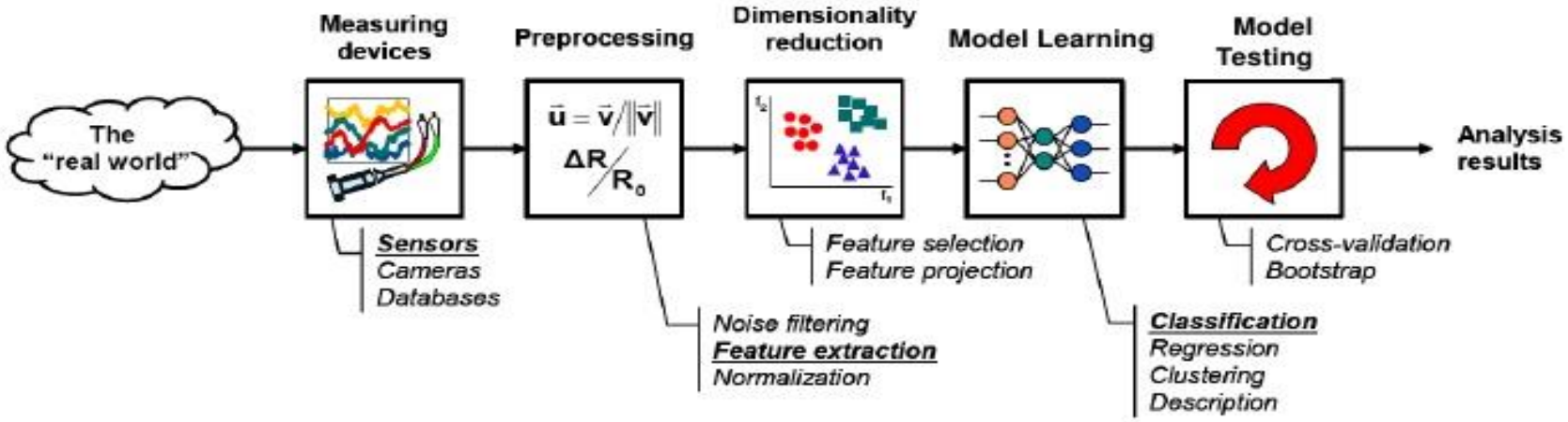
P:

% of spam emails that were filtered

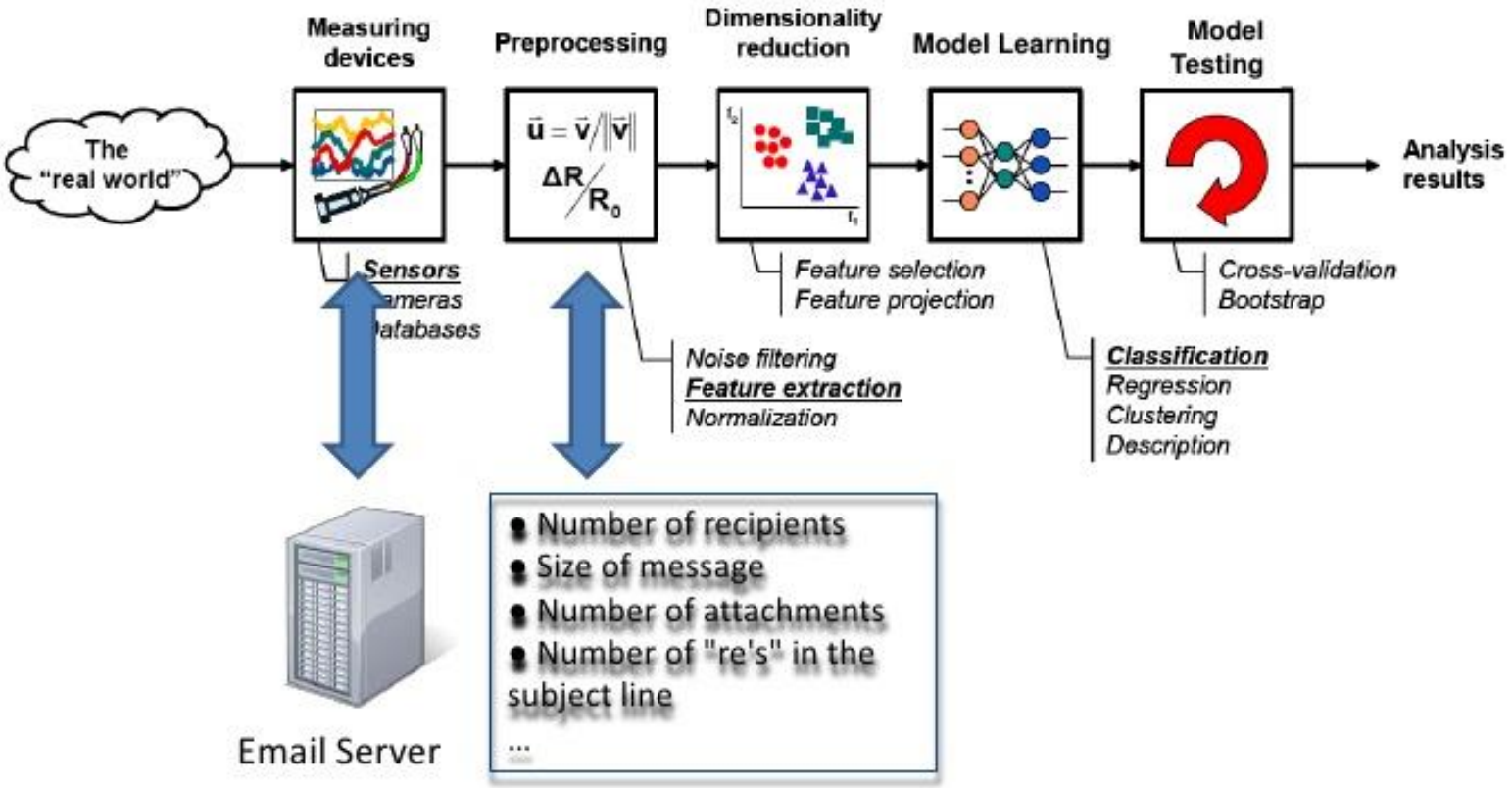
% of ham/ (non-spam) emails that were incorrectly filtered-out

E: a database of emails that were labelled by users

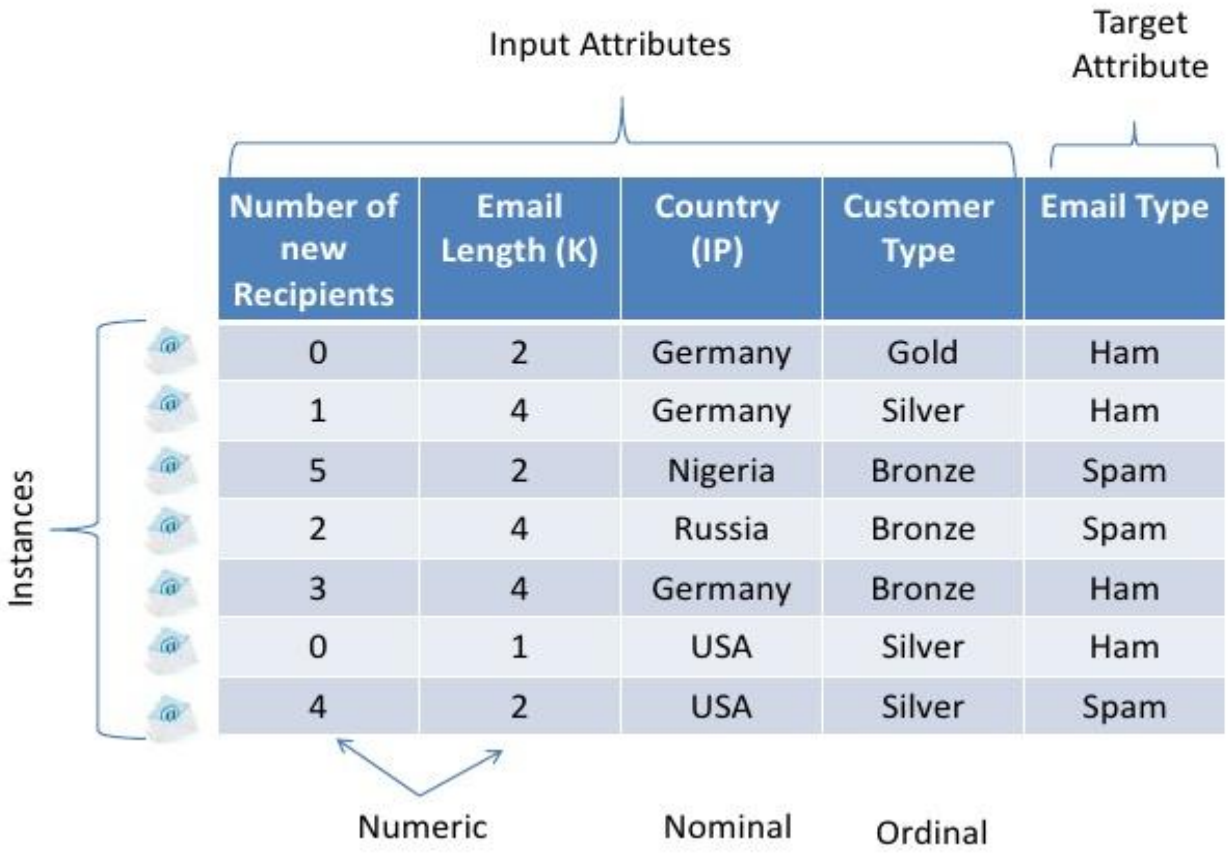
The Learning Process



The Learning Process in our Example



Data Set



Number of new Recipients	Email Length (K)	Country (IP)	Customer Type	Email Type
0	2	Germany	Gold	Ham
1	4	Germany	Silver	Ham
5	2	Nigeria	Bronze	Spam
2	4	Russia	Bronze	Spam
3	4	Germany	Bronze	Ham
0	1	USA	Silver	Ham
4	2	USA	Silver	Spam

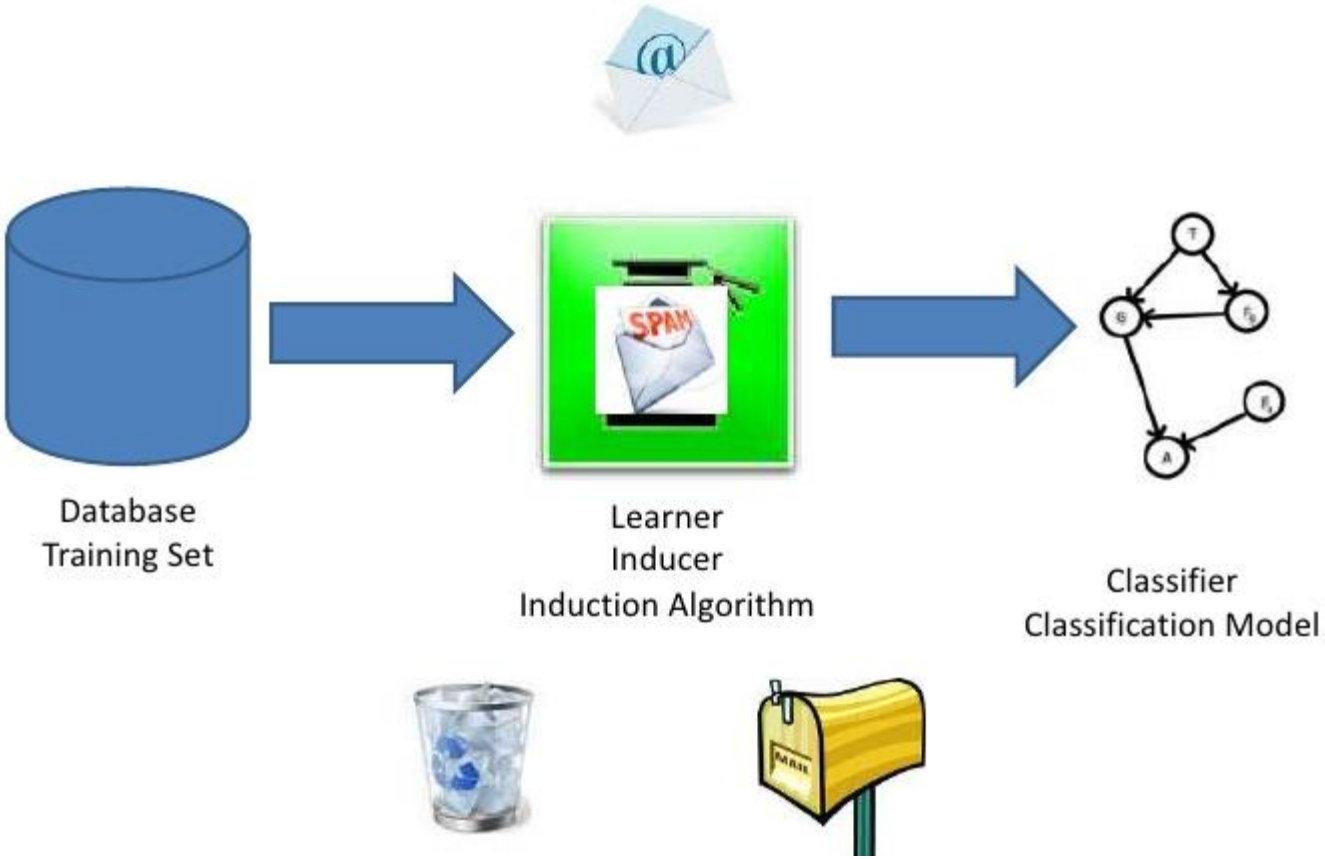
Step 4: Model Learning

Database
Training Set

Learner
Inducer
Induction Algorithm

Classifier
Classification Model

Step 5: Model Testing



Why Machine Learning?

- Develop systems that can automatically adapt and customize themselves to individual users.
  - Personalized news or mail filter
- Discover new knowledge from large databases (***data mining***).
  - Market basket analysis (e.g. diapers and beer)
- Ability to mimic human and replace certain monotonous tasks - which require some intelligence.
    - like recognizing handwritten characters
- Develop systems that are too difficult/expensive to construct manually because they require specific detailed skills or knowledge tuned to a specific task (knowledge engineering bottleneck).

Why now?

- Flood of available data (especially with the advent of the Internet)
- Increasing computational power
- Growing progress in available algorithms and theory developed by researchers
- Increasing support from industries

Machine learning algorithms

Categories: (Based on the way they **“learn”** about data to make predictions)

- **Supervised:** All data is labelled and the algorithms learn to predict the output from the input data.
- **Unsupervised:** All data is unlabelled and the algorithms learn to inherent structure from the input data.
- **Others:**
  - Reinforcement learning
  - Recommender systems.

Supervised Learning

- In supervised learning, we are given a data set and already know what our correct output should look like, having the idea that there is a relationship between the input and the output.

- Categories:

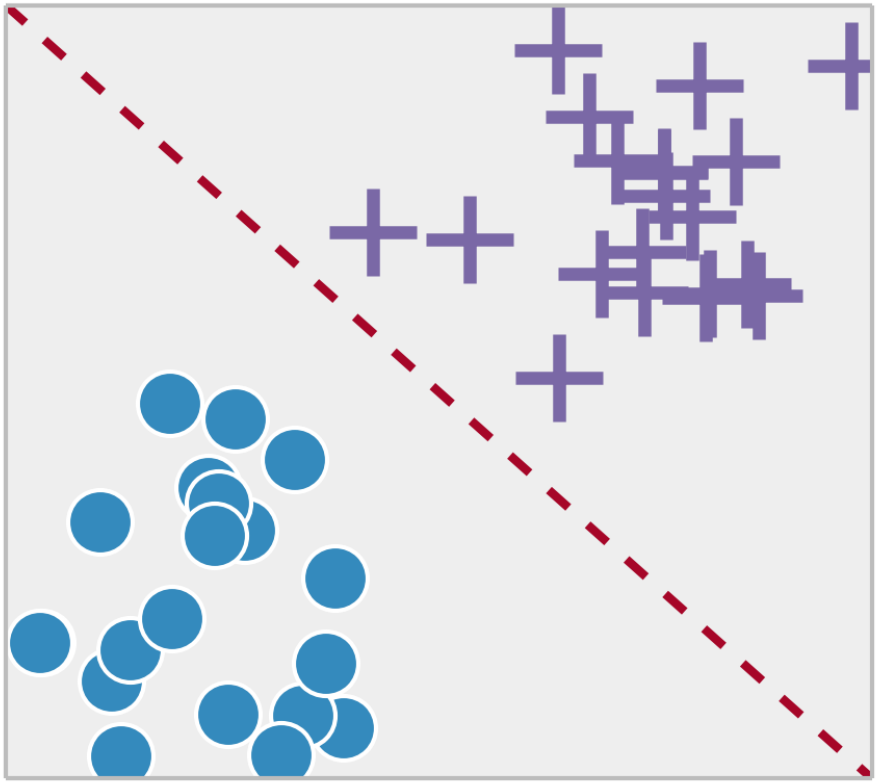
  - **Regression problem**: we try to predict results within a continuous output, meaning that we are trying to map input variables to some continuous function.

  - **Classification problem**: we are instead trying to predict results in a discrete output. In other words, we are trying to map input variables into discrete categories.

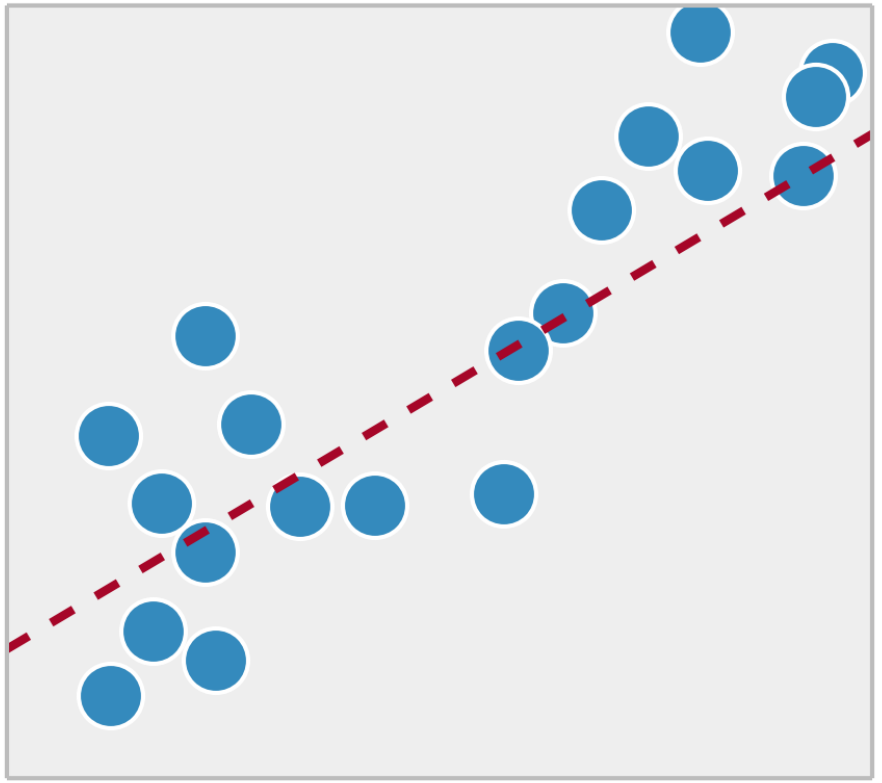
Example

- Given data about the size of houses on the real estate market, try to predict their price.
- Price as a function of size is a continuous output, so this is a **regression problem.**
- We could turn this example into a **classification problem** by instead making our output about whether the house "sells for more or less than the asking price." Here we are classifying the houses based on price into two discrete categories.

Classification



Regression

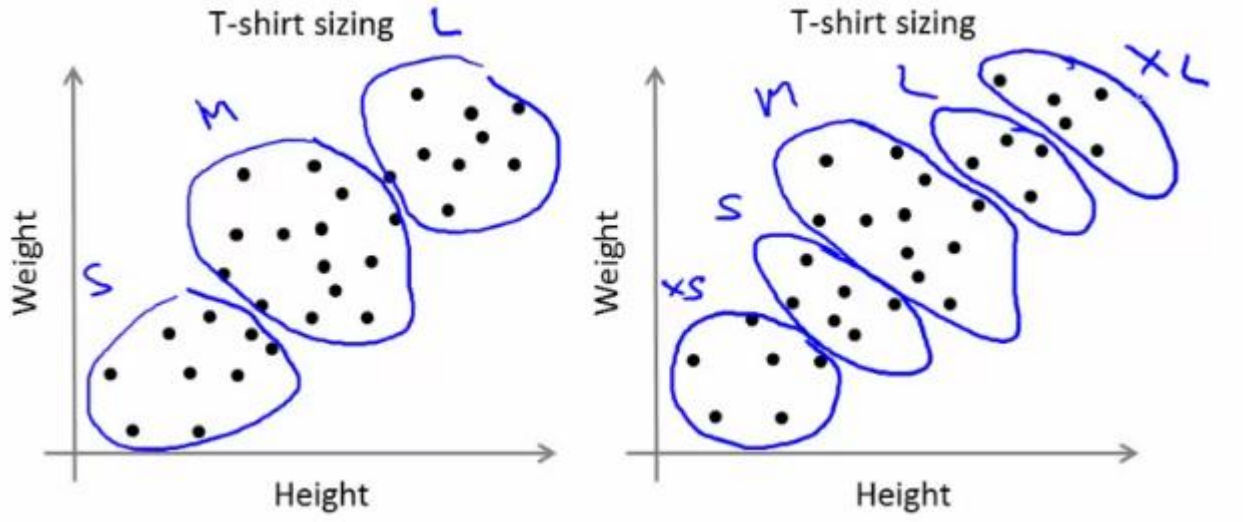


Unsupervised Learning

- Unsupervised learning is where you only have input data (X) and no corresponding output variables.
- The goal for unsupervised learning is to model the underlying structure or distribution in the data in order to learn more about the data.
- These are called unsupervised learning because unlike supervised learning above there is no correct answers and there is no teacher. Algorithms are left to their own devises to discover and present the interesting structure in the data.

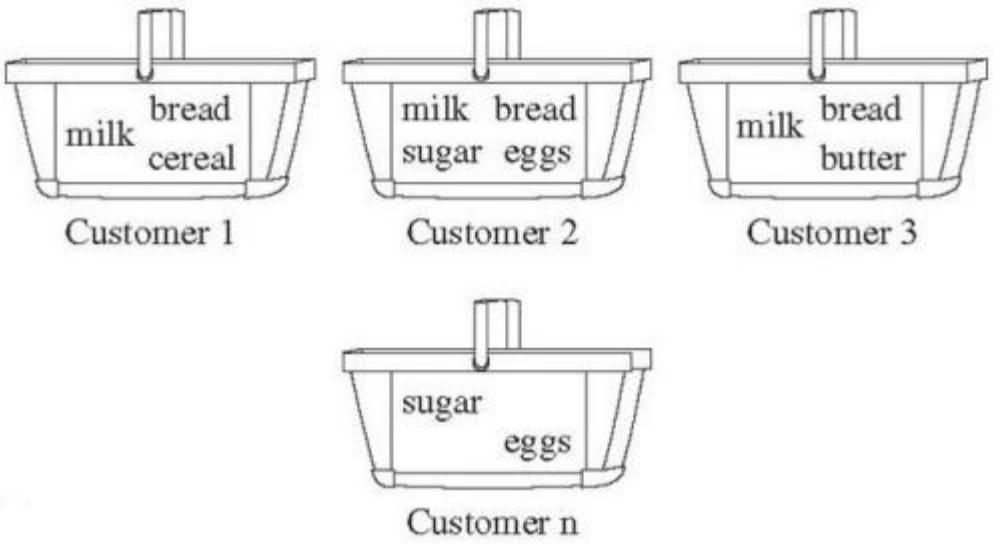
Example

- **Clustering**: A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behaviour.

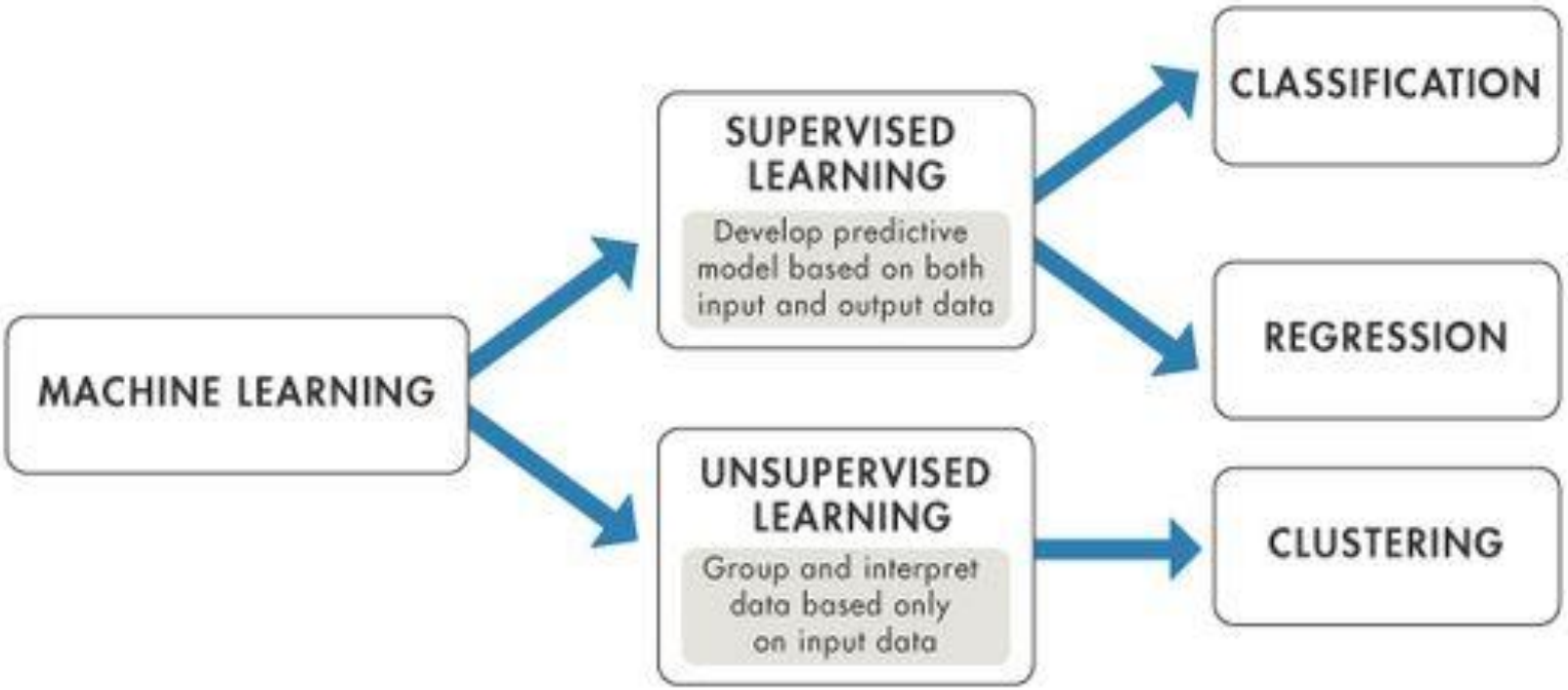


Example

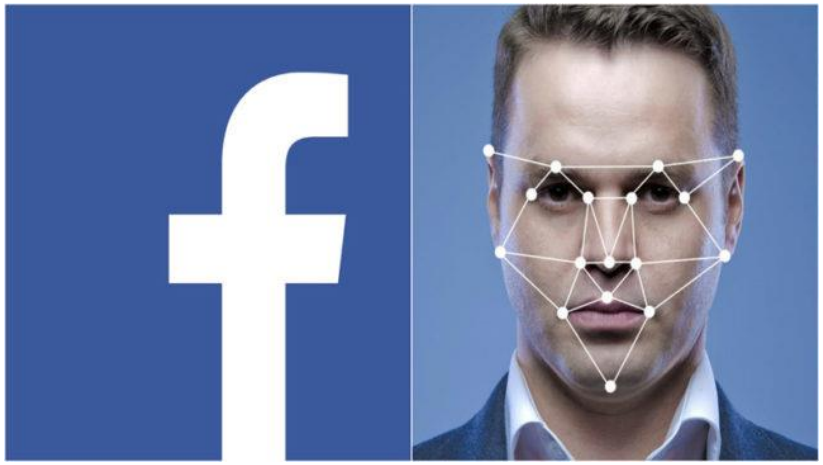
- **Association**: An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

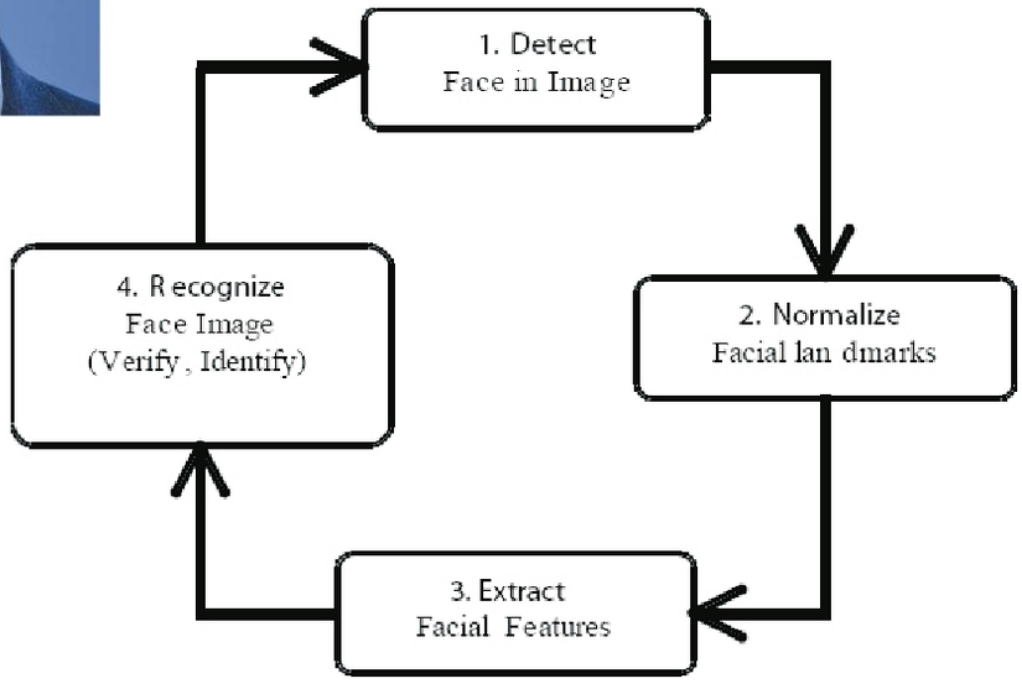


Important areas

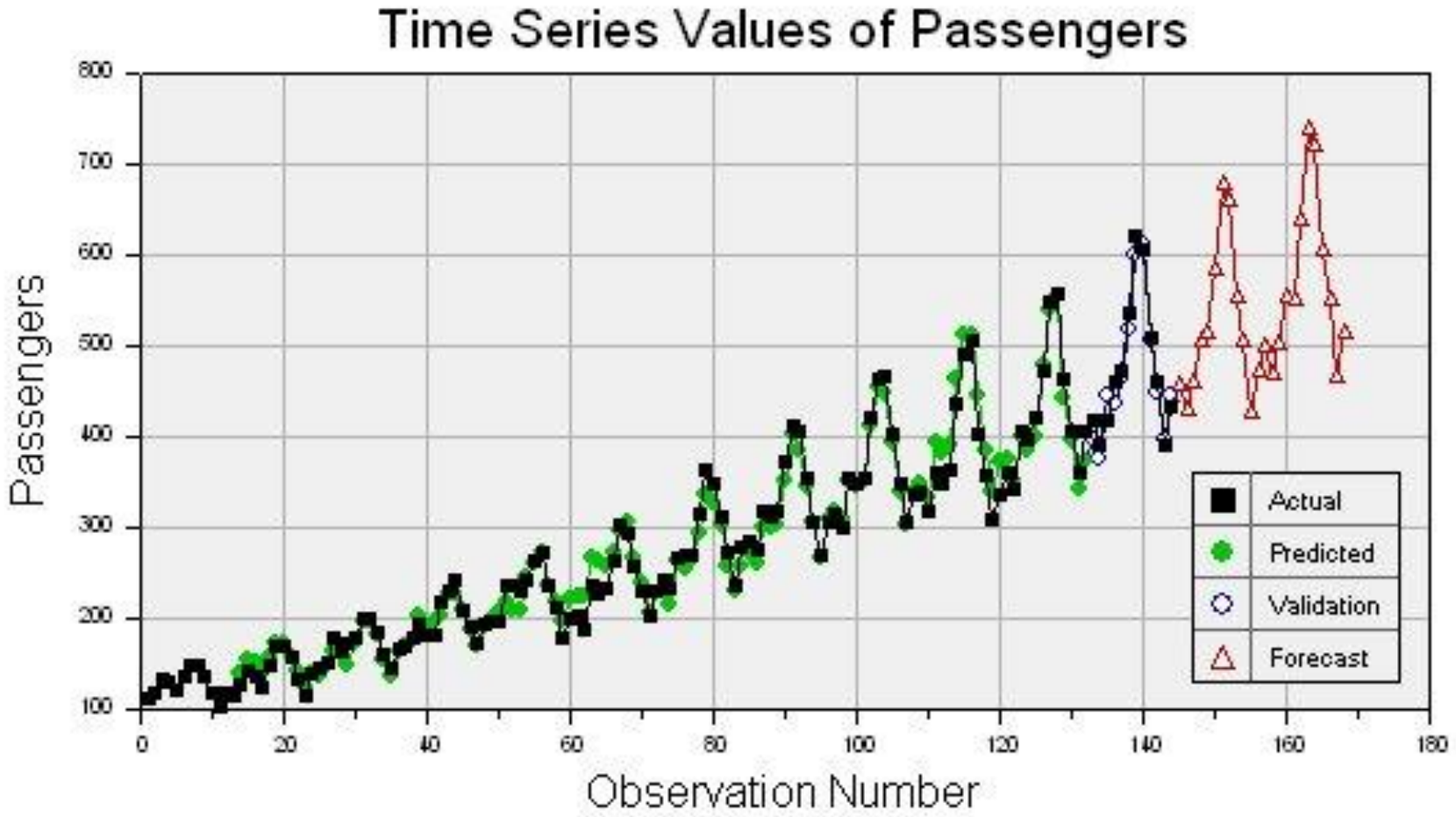


Facebook- Facial Recognition





Autonomous Driving


Time Series Values of Passengers
Passengers
Observation Number
100
200
300
400
500
600
700
800
0
20
40
60
80
100
120
140
160
180
Actual
Predicted
Validation
Forecast

Other applications –

- Virtual Personal Assistants
- Online Customer Support
- Product Recommendations
- Assistive and Medical Tech
- Drug Discovery/Manufacturing
- Personalized Treatment/Medication
- Machine Learning Examples in Finance for Fraud Detection
- Machine Learning Examples in Travel for Dynamic Pricing – Uber

Concepts covered so far..

- What is Machine Learning?
- Steps involved in learning!
- Supervised & Unsupervised Learning.
- Regression, Classification & Clustering
- Few applications of Machine learning

Thank you!