Kmit

SONET

Data Science

(Level-1)

Ajeet K. Jain, M. Narsimlu

(ML TEAM)- SONET, KMIT, Hyderabad

Session - 1

This session deals with

Why Data Science

About Data Science

Definitions of Data Science

What is machine Learning

Where Machine Learning is used

Types of Machine Learning

Course Structure

Why Data Science...?

Huge amount of data is generated from last 3 decades

Examples of data generators:Websites, Smart phones, Stock Markets

Lot of hidden information is in that data which no one can observe.

Analysis is required on data to find the hidden facts

Data might be in structured/semi-structured/unstructured form.

It is difficult to analyze the structured data with unstructured data

Analysis of Real-Time data is challenging task

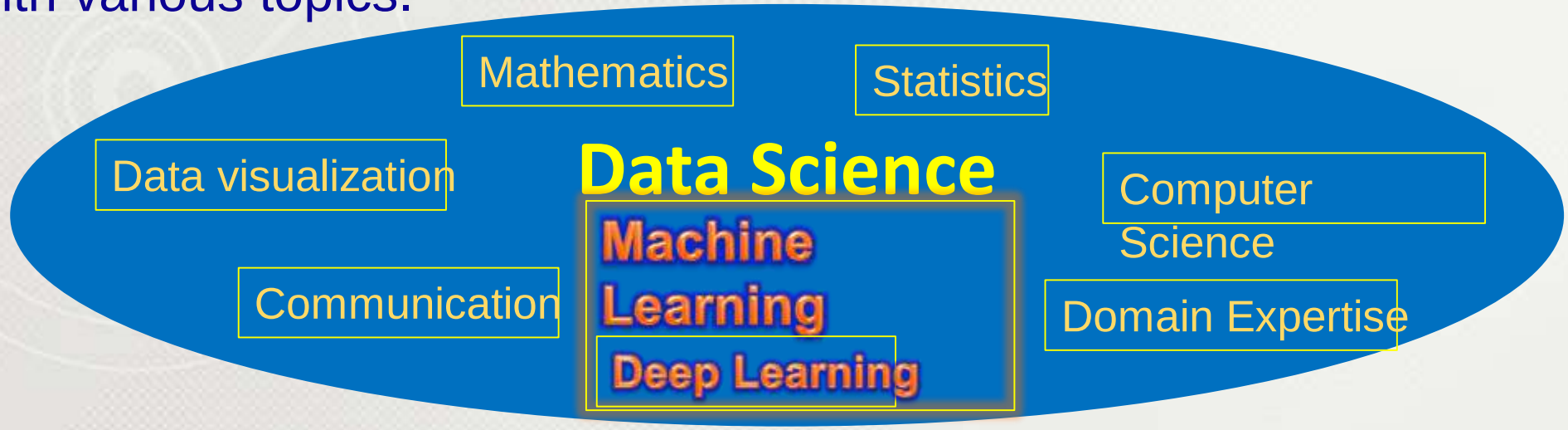
Analysis of Big Data is challenging task

Companies like Facebook, Yahoo, Google, Linkedin started to analyze data to improve their product.

About Data Science...!

Data Science is an emerging field with rapid changes and opportunities

It is linked with various topics.



The scope of data science is measureless

Manufacturing	Education	
Healthcare	Sports & Games	
Finance	Telecom	Many more….
Biotech	Transport & travel	

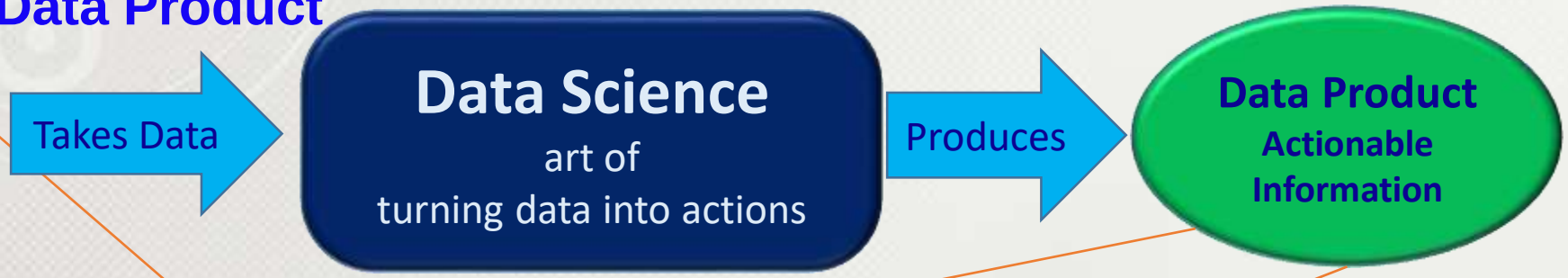
Definition of Data Science

Sources of Data

Data is available from various sources.
Data can be in various formats : structured/unstructured/semi-structured

It might be useful or irrelevant for a particular purpose.

Data science is an art of turning data into actions by producing Data Product



Examples:

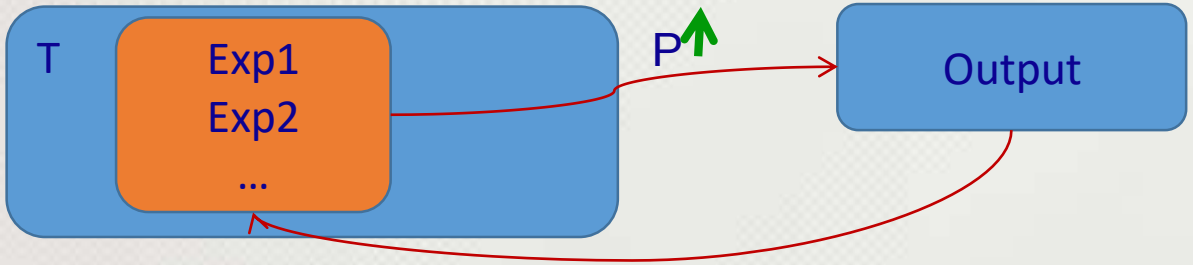
Recommendations
Weather forecasts
Stock market predictions
Etc..

What Is Machine Learning...?

Machine Learning is the science (and art) of programming computers so they can learn from data.

Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed.

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.

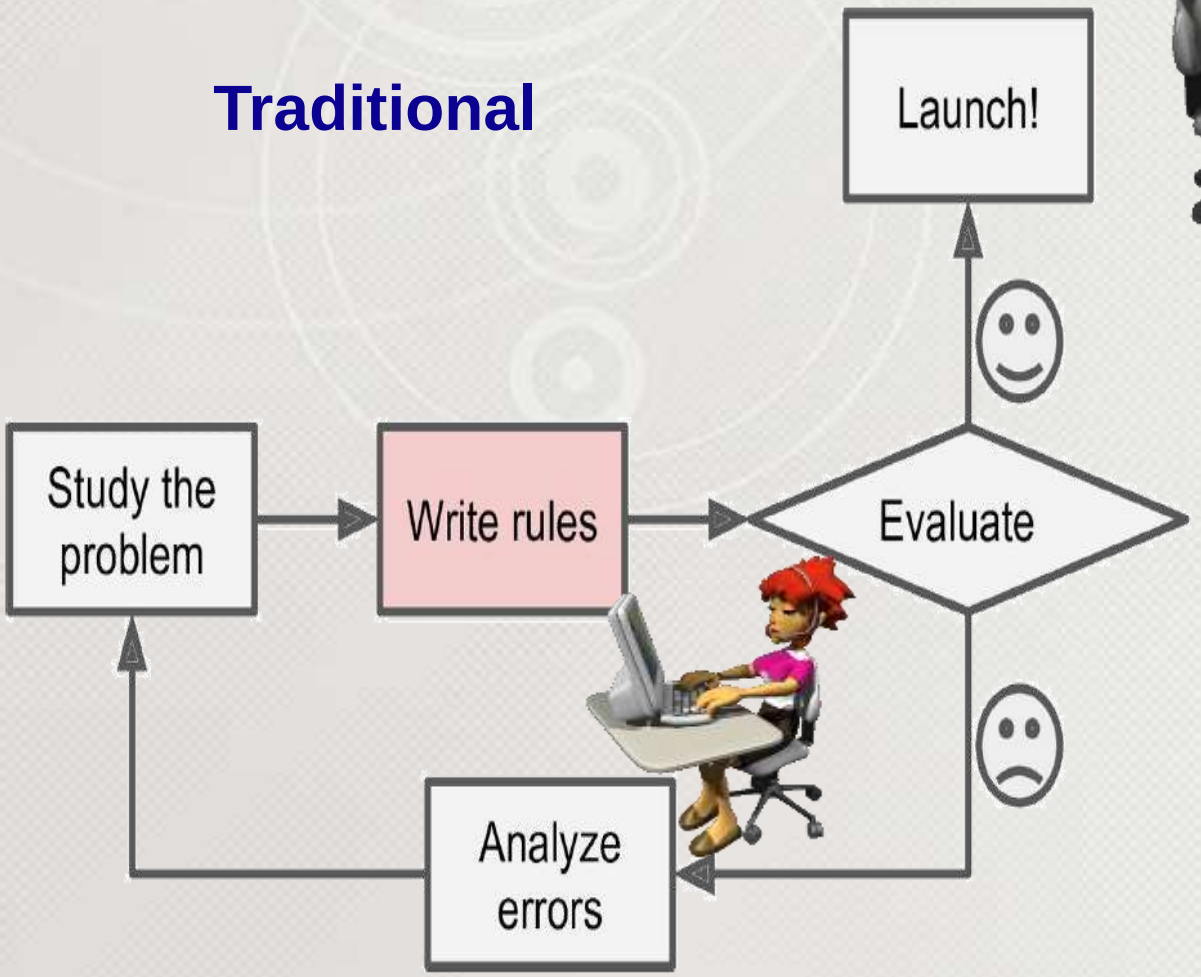


SONET

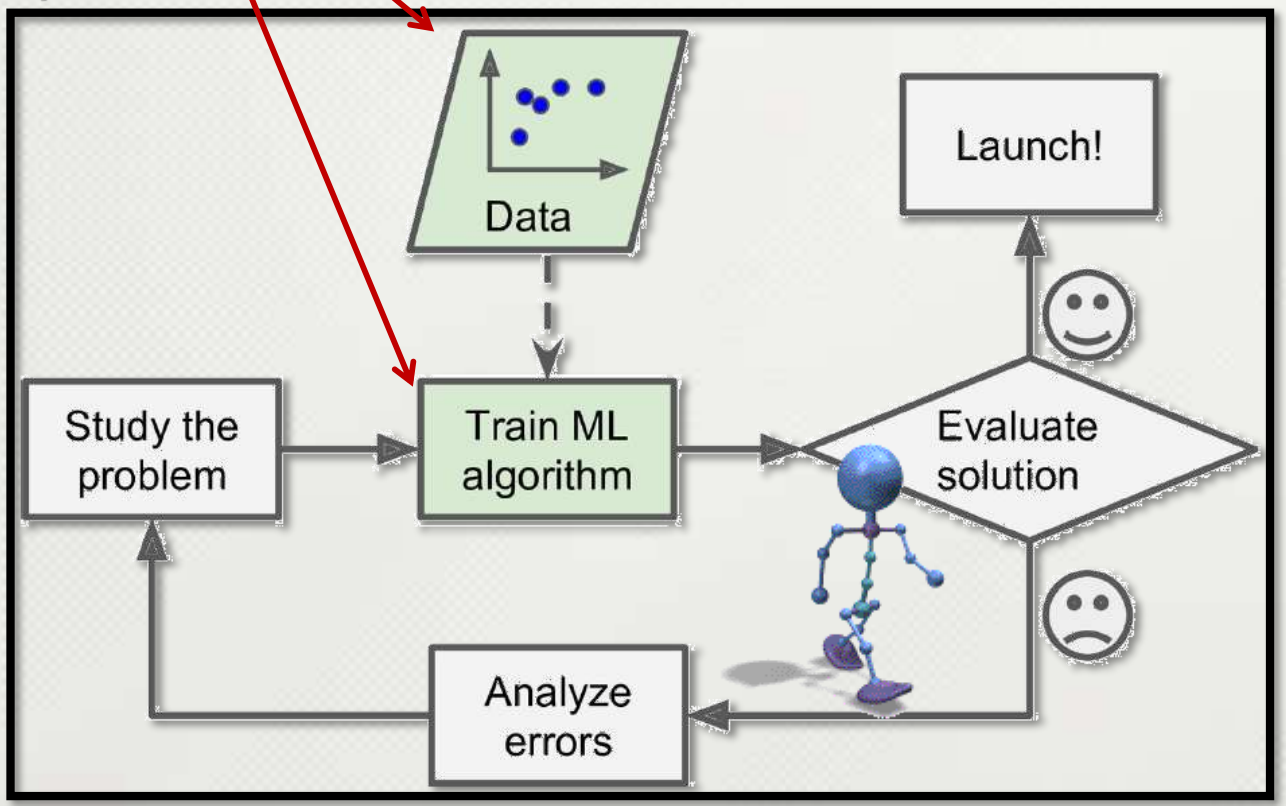
Traditional / ML

DATA SCIENCE

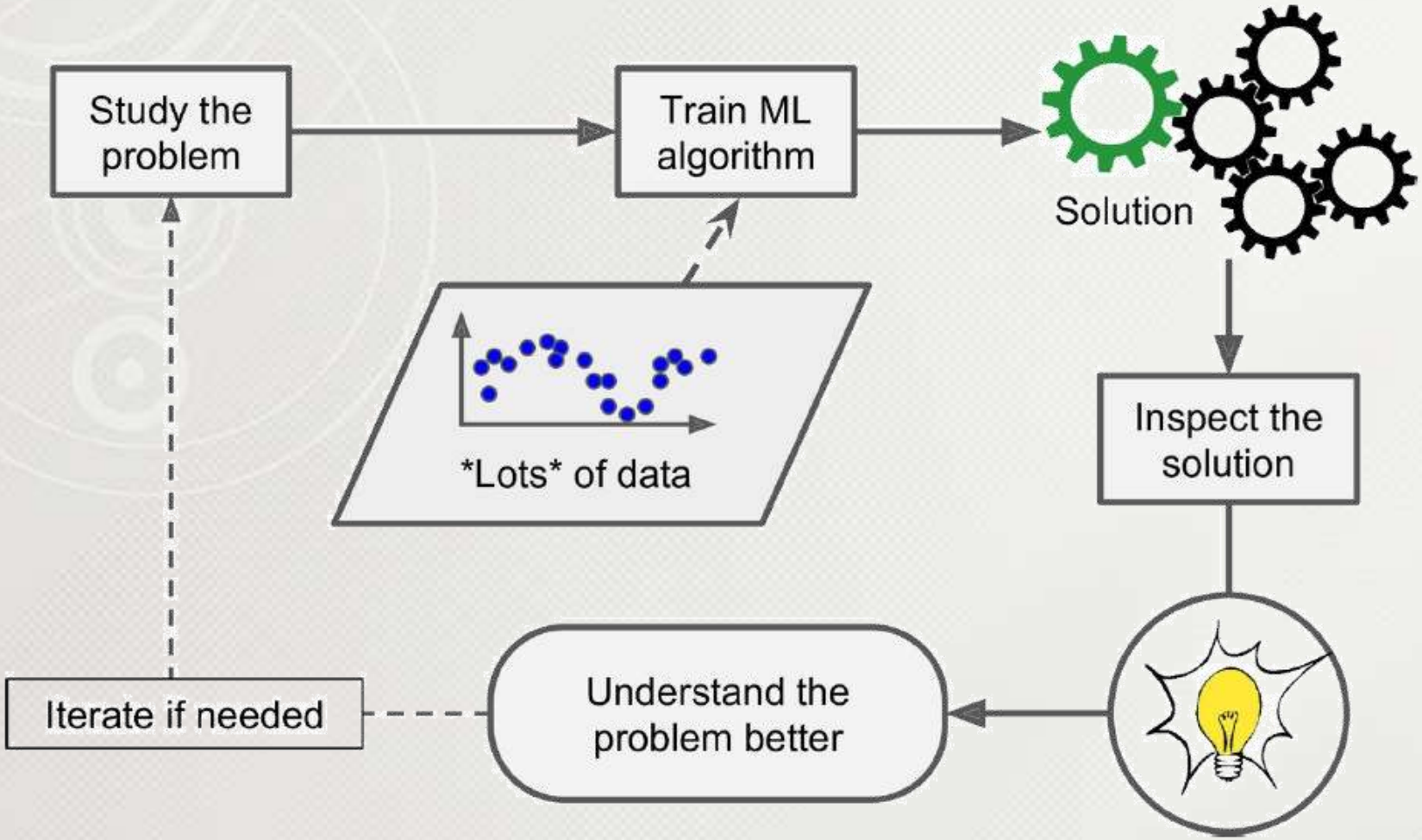
Traditional



Machine Learning



Machine Learning Flow



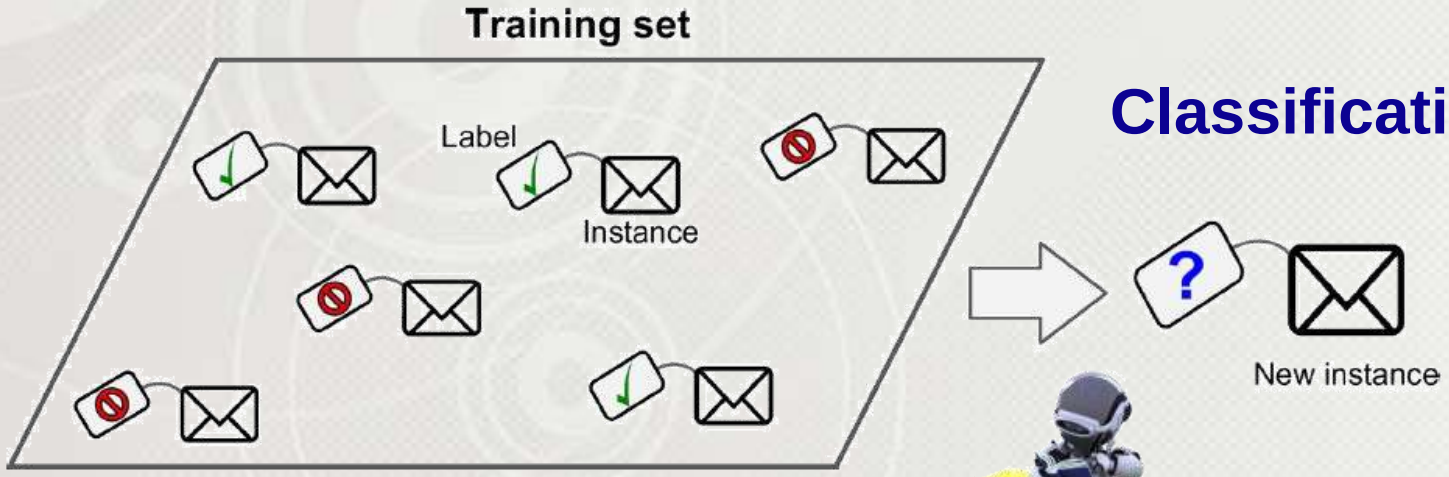
Where ML fit for...?

- Problems for which existing solutions require a lot of hand-tuning or long lists of rules: one Machine Learning algorithm can often simplify code and perform better.
- Complex problems for which there is no good solution at all using a traditional approach: the best Machine Learning techniques can find a solution.
- Fluctuating environments: a Machine Learning system can adapt to new data.
- Getting insights about complex problems and large amounts of data.

SONET

Supervised Learning

DATA SCIENCE



Classification

- k-Nearest Neighbors
- Linear Regression
- Logistic Regression
- Support Vector Machines (SVMs)
- Decision Trees and Random Forests
- Neural networks[2]

Class	Percentage
10th	62
Intermediate	78
Graduation	X
PG	85

Regression

Class	Percentage
Graduation	82

Unsupervised Learning

DATA SCIENCE

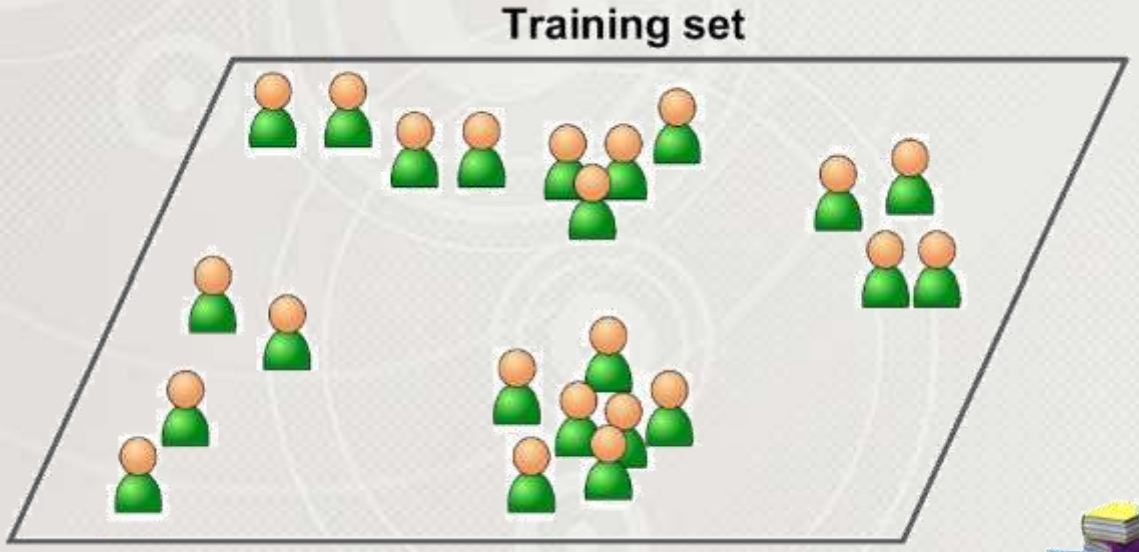


Figure 1-7. An unlabeled training set for unsupervised learning

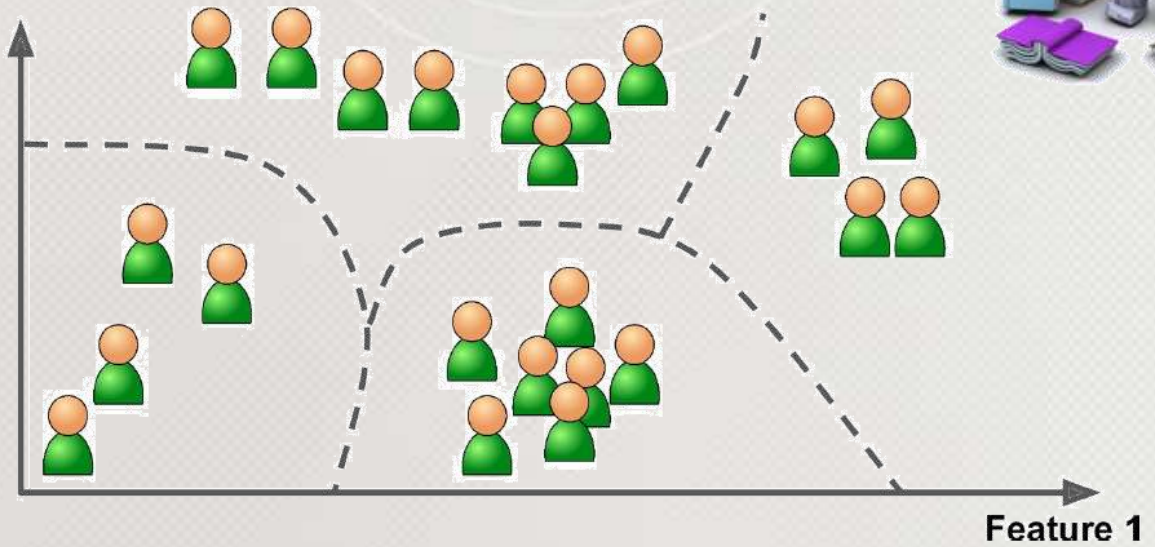


Figure 1-8. Clustering

- Clustering
  - k-Means
  - Hierarchical Cluster Analysis (HCA)
  - Expectation Maximization
- Visualization and dimensionality reduction
  - Principal Component Analysis (PCA)
  - Kernel PCA
  - Locally-Linear Embedding (LLE)
  - t-distributed Stochastic Neighbor Embedding (t-SNE)
- Association rule learning
  - Apriori
  - Eclat

Kmit SONET

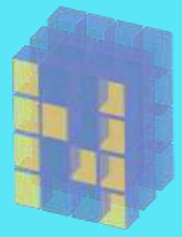
Course Structure

DATA SCIENCE

IDE

BASICS OF PYTHON

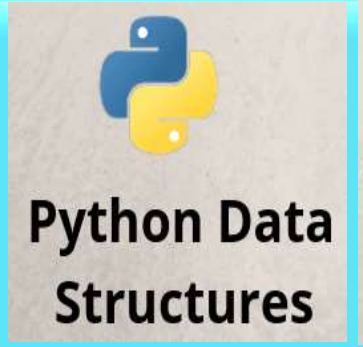
python



NumPy

Data Analysis

Pandas



Python Data Structures

DATA VISUALIZATION

Data Science Case Studies

scikit learn

MATPLOTLIB

Data Visualization in Python

Seaborn

ANACONDA

spyder

Session - 1

Conclusion

You are aware of

- What is Data Science
- What is Machine Learning
- Types of Machine Learning

We will proceed with

- Python

Kmit
SONET

DATA SCIENCE

THANK YOU