Introduction to Machine Learning

Dr. Arijit Ghosal
Associate Professor, Department of Information Technology

St. Thomas' College of Engineering & Technology, Kolkata

What is CAT?

What is Machine Learning?

- Herbert Alexander Simon:

  "Learning is any process by which a system improves performance from experience."

- "Machine Learning is concerned with computer programs that automatically improve their performance through experience."

When Do We Use Machine Learning?

- Human expertise does not exist (navigating on Mars)
- Speech recognition
- Develop systems that can automatically adapt and customize themselves to individual users (Personalized news or mail filter)

When Do We Use Machine Learning?

- Develop systems that are too difficult/expensive to construct manually because they require specific detailed skills or knowledge tuned to a specific task

- You cannot code the rules

- You cannot scale

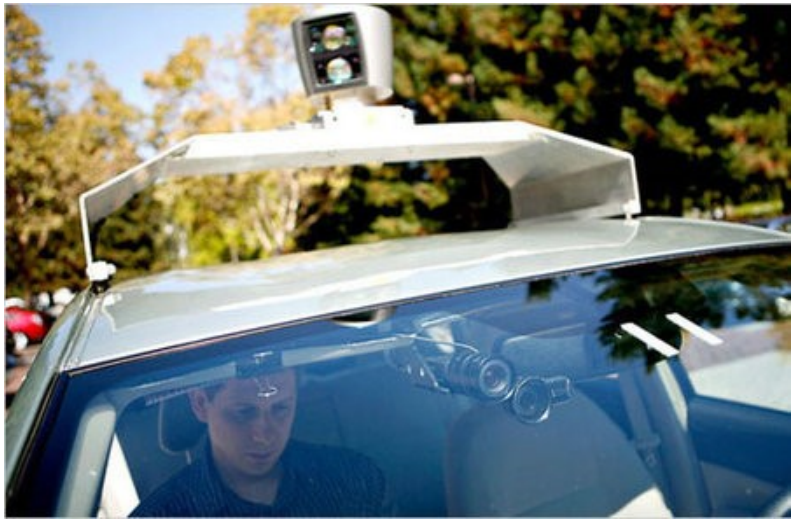
Examples of Tasks That Are Best Solved By Using A Learning Algorithm

- Recognizing patterns: facial expressions, Handwritten or spoken words, Medical images
- Recognizing anomalies: Unusual transactions
- Prediction: Future stock prices, Weather forecasting

State of the Art Applications of Machine Learning

Autonomous Cars

- Nevada made it legal for autonomous cars to drive on roads in June 2011



Types of Machine Learning Problems

Supervised

- Learn through examples of which we know the desired output (what we want to predict)

  - Is this a cat or a dog?

  - Are these emails spam or not?

- Two types – Classification, Regression

Types of Machine Learning Problems

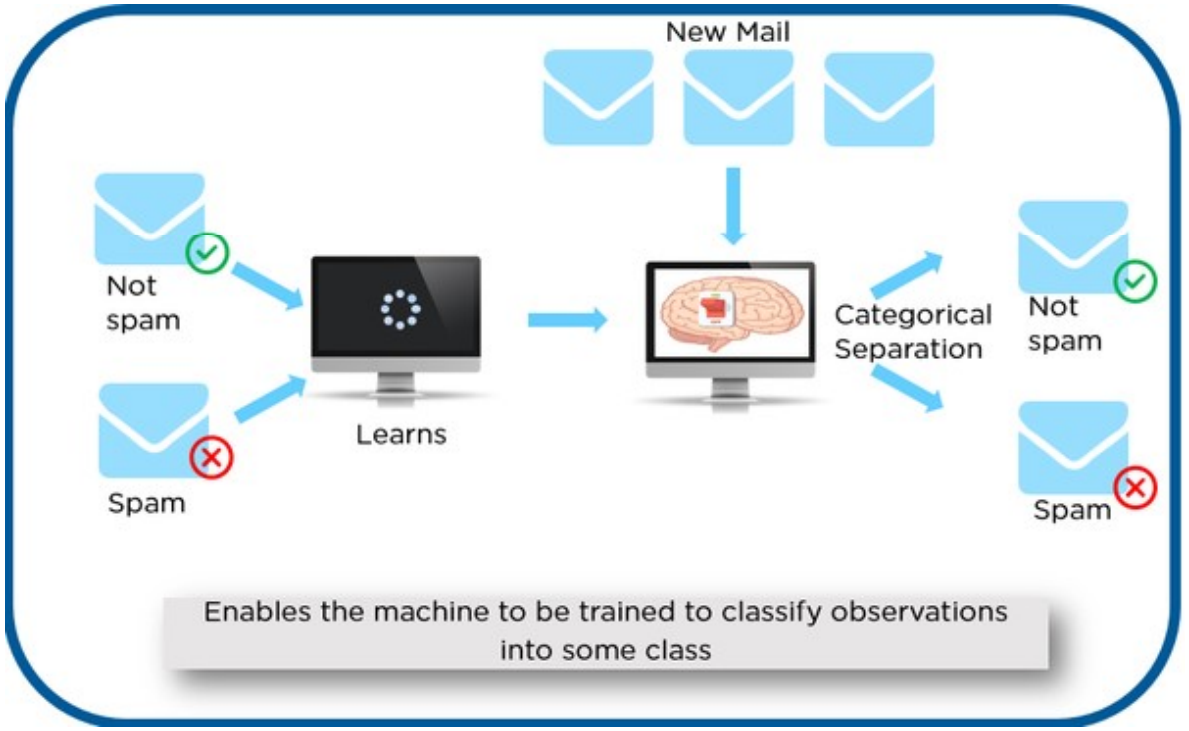
Supervised

- Classification - Output is a discrete variable (e.g., cat/dog), basically a category

- Regression - Output is continuous (e.g., price, temperature)

Types of Machine Learning Problems

Supervised



Types of Machine Learning Problems

Unsupervised

- There is no desired output. Learn something about the data

  - I have photos and want to put them in 20 groups

Types of Machine Learning Problems

Unsupervised

- Two types:

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

Types of Machine Learning Problems

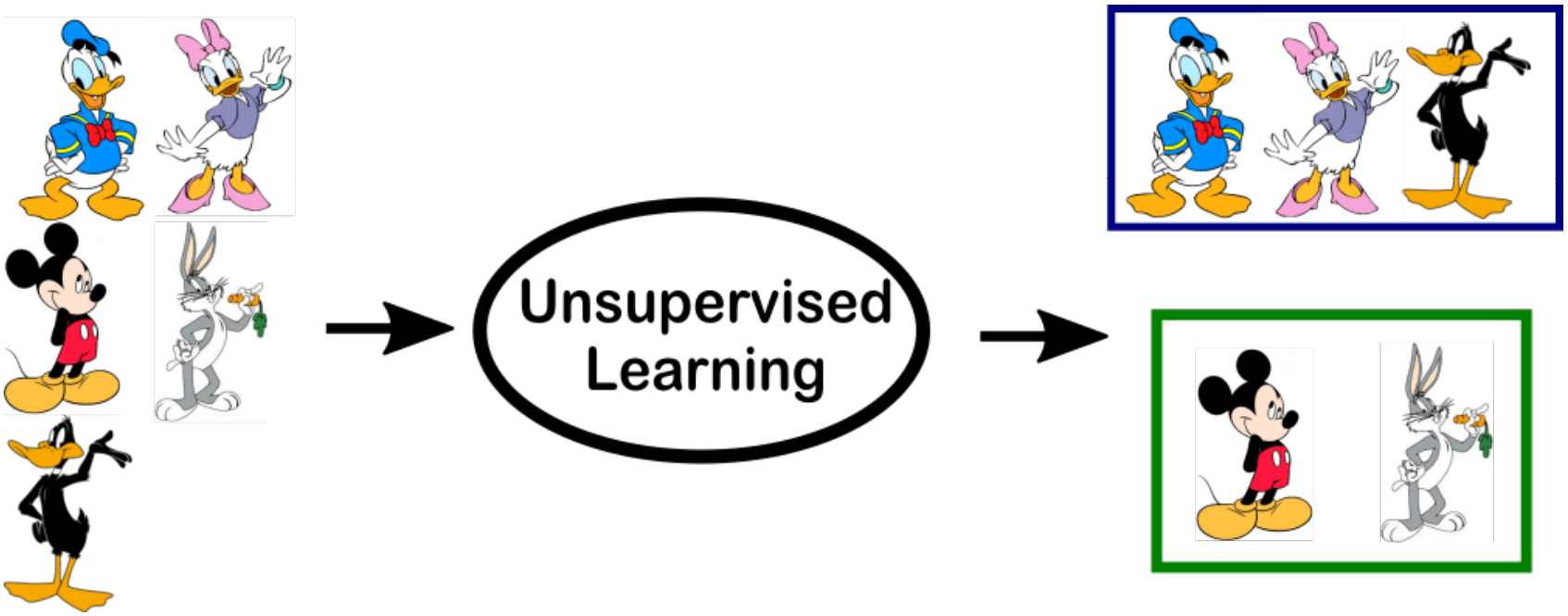
Unsupervised

- Two types:

  - 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

Types of Machine Learning Problems

Unsupervised



Types of Machine Learning Problems

Semi-supervised

- It is used for the same applications as supervised learning. But it uses both labeled and unlabeled data for training – typically a small amount of labeled data with a large amount of unlabeled data (because unlabeled data is less expensive and takes less effort to acquire)

Types of Machine Learning Problems

Semi-supervised

- Semi-supervised learning is useful when the cost associated with labeling is too high to allow for a fully labeled training process

Types of Machine Learning Problems

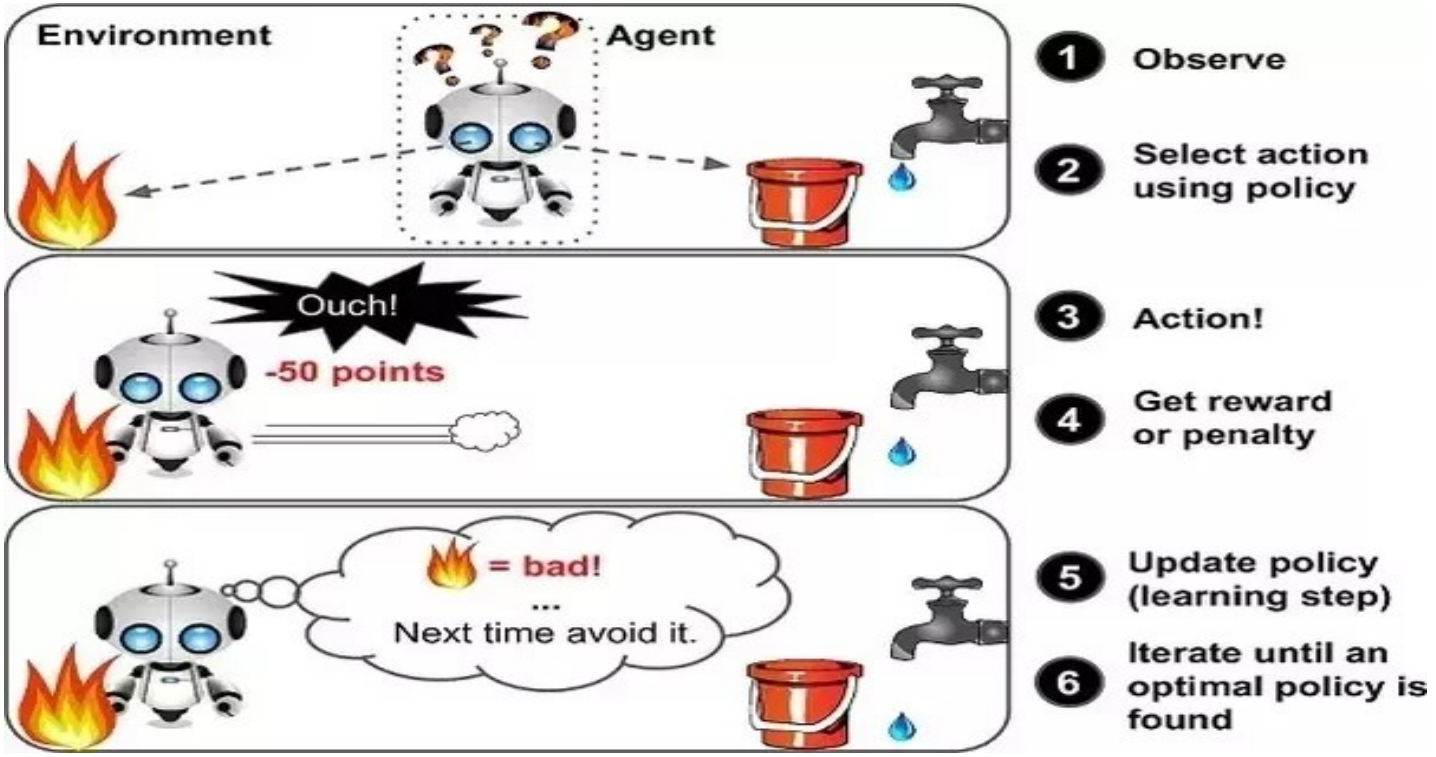
Reinforcement

- An agent interacts with an environment and watches the result of the interaction

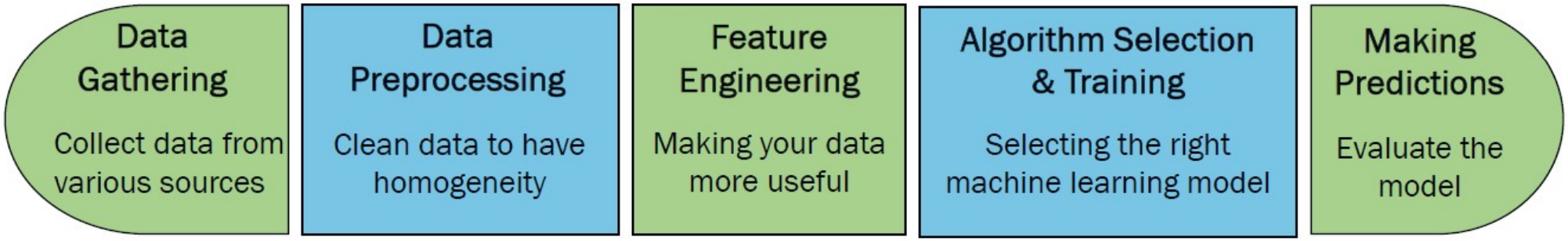
- Environment gives feedback via a positive or negative reward signal

Types of Machine Learning Problems

Reinforcement



Steps to Solve a Machine Learning Problem



Steps to Solve a Machine Learning Problem

Data Gathering

- Might depend on human work
- Some algorithms need large amounts of data to be useful (e.g., neural networks)
- The quantity and quality of data dictate the model accuracy

Steps to Solve a Machine Learning Problem

Data Pre-processing

- It refers to filtering the raw data for noise suppression and other operations performed on the raw data to improve its quality
- It also refers to bad encoding (for text)
- Biased data: Do I have many more samples of one class than the rest?

Steps to Solve a Machine Learning Problem

Feature Engineering

- What is a feature? - *A feature is an individual measurable property of a phenomenon being observed*
- Inputs are represented by a set of features
- Requires thought and knowledge of the data

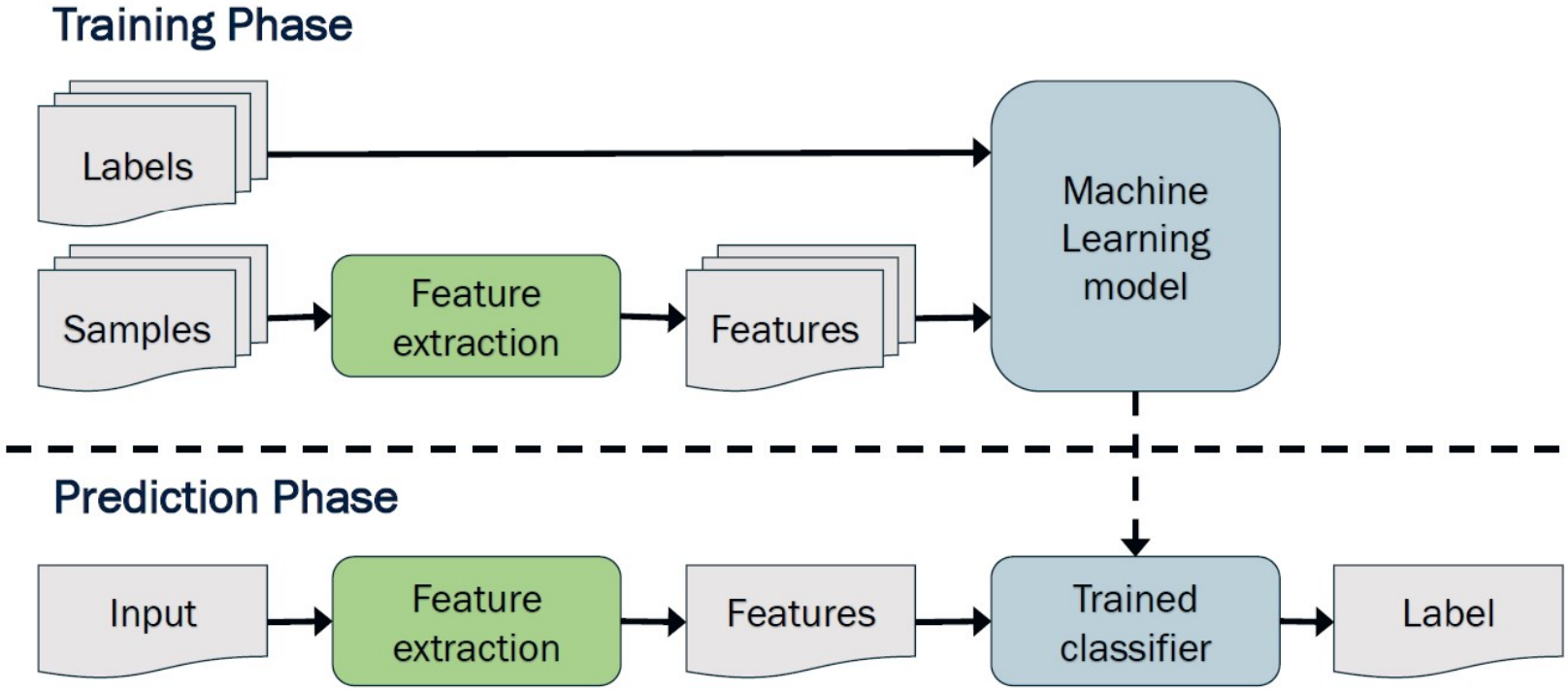
Steps to Solve a Machine Learning Problem

Algorithm Selection & Training

- Supervised
- Unsupervised
- Reinforcement

Steps to Solve a Machine Learning Problem

Making Predictions



Development Platform

Language Choice

- Python
- R
- Matlab
- Octave
- Julia
- C++
- C

Development Platform

IDEs

- R Studio
- Pycharm
- iPython/Jupyter Notebook
- Julia
- Spyder
- Anaconda
- Rodeo
- Google –Colab

Thank You