BCSE209L- Machine Learning
Module 1

Dr. G. Praveen Kumar
SMEC (Spl in AI&ML)

Course Objectives

- To teach the theoretical foundations of various learning algorithms.
- To train the students better understand the context of supervised and unsupervised learning through real-life examples.
- To understand the need for Reinforcement learning in real – time problems.
- Apply all learning algorithms over appropriate real-time dataset.
- Evaluate the algorithms based on corresponding metrics identified.

Expected Course Outcome:

At the end of this course, student will be able to:

1. Understand, visualize, analyze and preprocess the data from a real-time source.
2. Apply appropriate algorithm to the data.
3. Analyze the results of algorithm and convert to appropriate information required for the real – time application.
4. Evaluate the performance of various algorithms that could be applied to the data and to suggest most relevant algorithm according to the environment

Hours/ Week : 3

L	T	P	C
2	1	0	3

Internal Mark Configuration Rubrics :

DA 1 : 10 marks

Quiz 1 : 10 marks

Quiz 2 : 10 marks

CAT 1 : 50 marks (5*10=50 marks, no choice)

CAT 2 : 50 marks (5*10=50 marks, no choice)

Fat : 100 marks (10*10=100 marks, 10 out of 12)

Modules

- Module 1- Introduction to Machine Learning and Pre-requisites (CO1)
- Module 2- Supervised Learning -I(CO2)
- Module 3- Supervised Learning –II (CO3)
- Module 4- Unsupervised Learning (CO4)
- Module 5- Ensemble Learning (CO5)
- Module 6- Machine Learning in Practice (CO6)
- Module 7- Reinforcement Learning (RL) (CO7)
- Module 8- Contemporary Issues

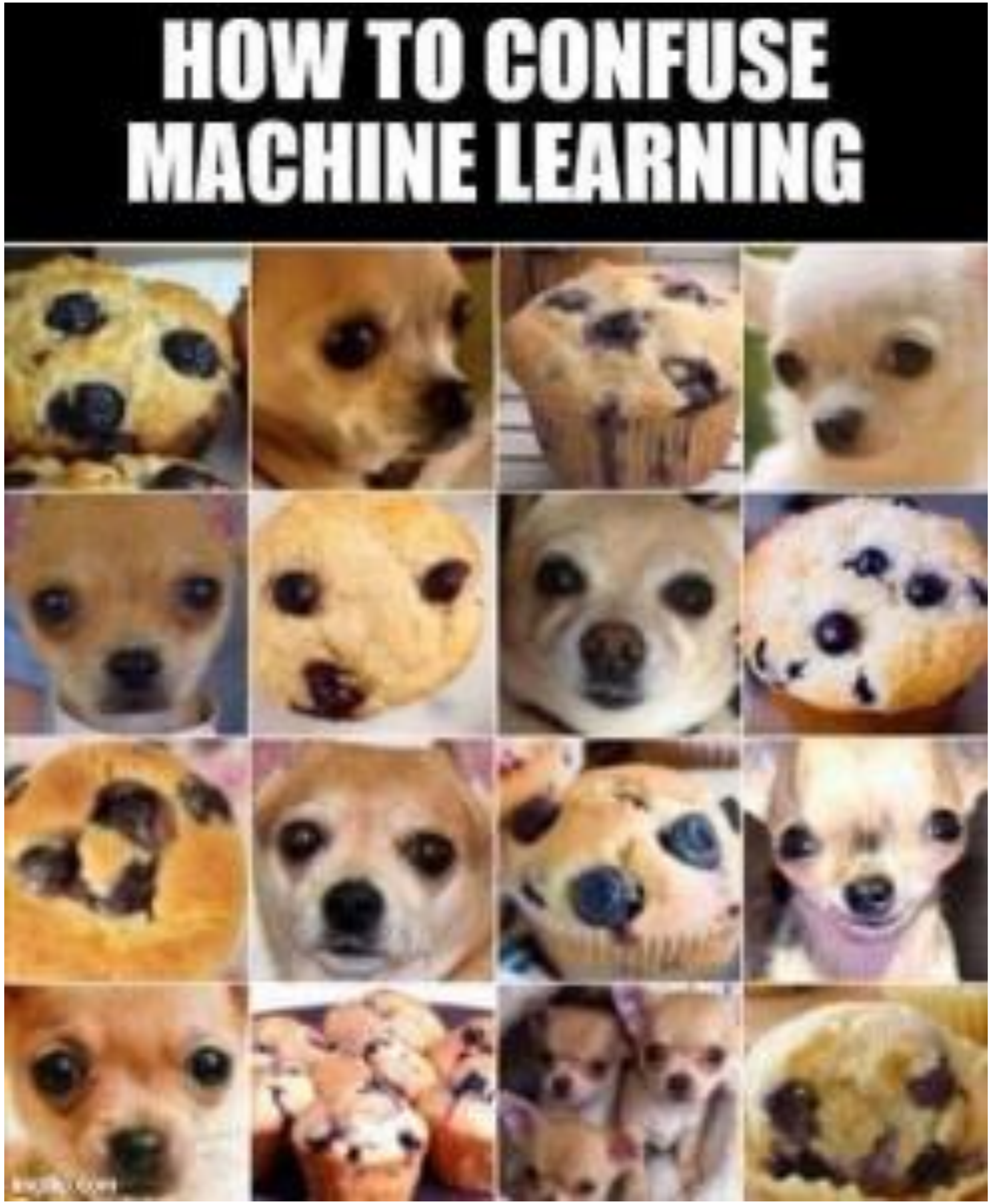
Books

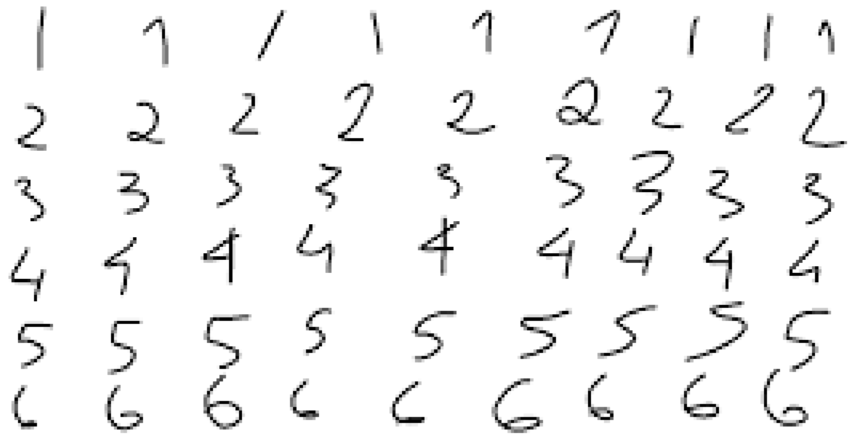
Textbook

1. **Ethem Alpaydin,"Introduction to Machine Learning", MIT Press, Prentice Hall of India**.
2. Reinforcement Learning: An Introduction (Adaptive Computation and Machine Learning series) 2nd edition, Richard **S. Sutton and Andrew G. Barto, A** Bradford Book; 2018,

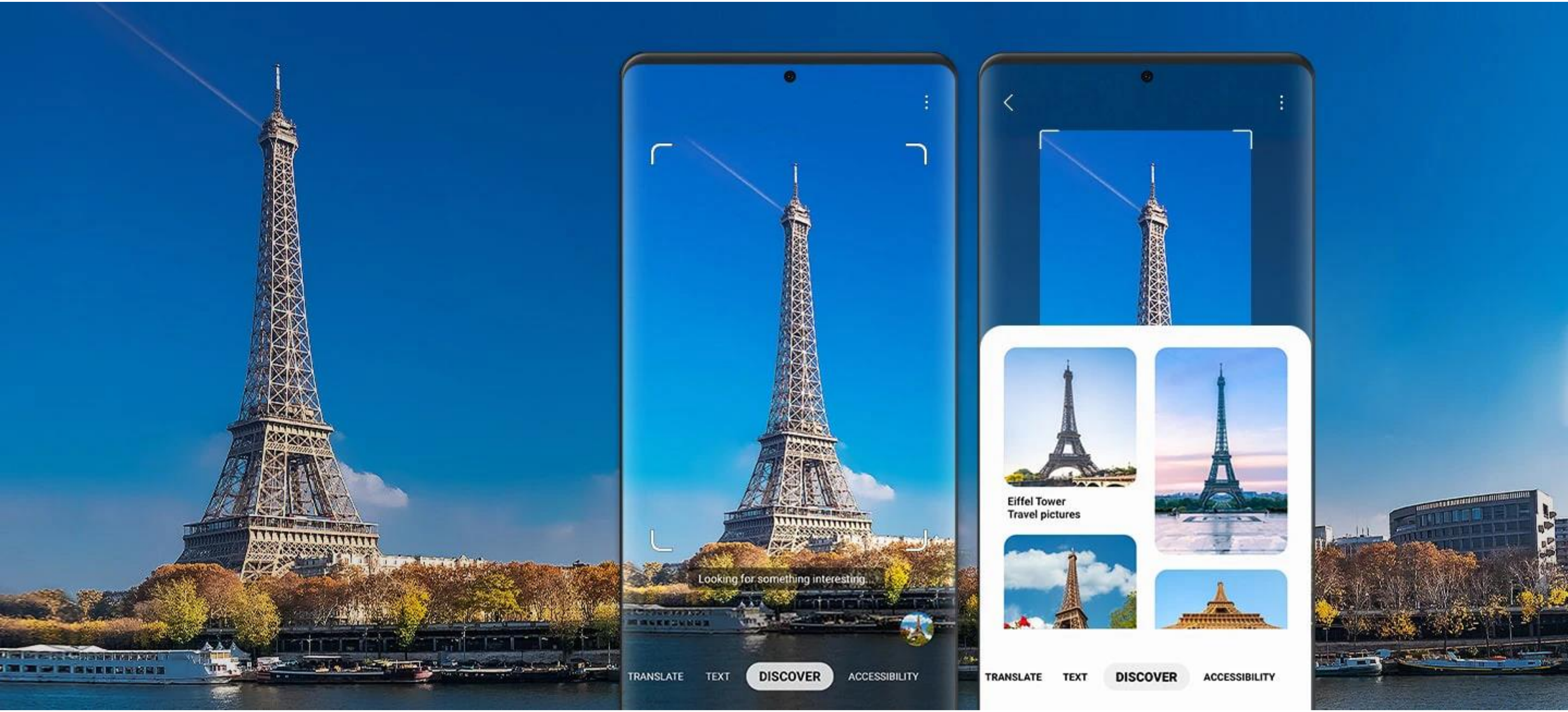
Reference Books

1. **Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar "Foundations of Machine Learning", MIT Press, 2012.**
2. **Tom Mitchell, "Machine Learning", McGraw Hill, 3rd Edition,1997.**
3. **Charu C. Aggarwal, "Data Classification Algorithms and Applications" , CRC Press, 2014**


HOW TO CONFUSE
MACHINE LEARNING



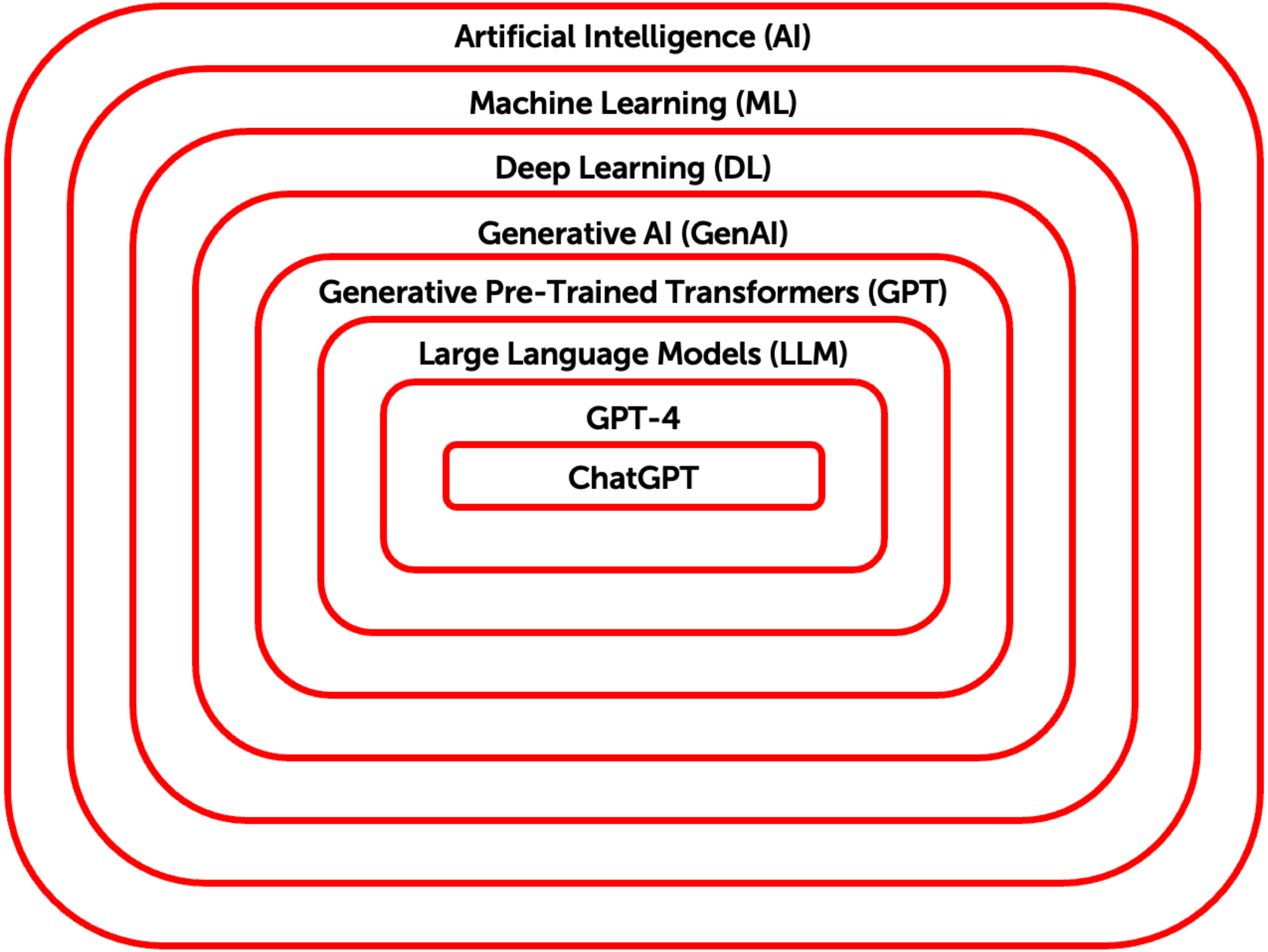
Bixby Vision



Bixby Vision

AI Terminology

@AndreRetterath
datadrivenvc.io



Evolution of Machine Learning

1805

Linear Regression

(The method of least squares): Developed by Adrien-Marie Legendre (1805) and Carl Friedrich Gauss (1809).

1957

Perceptron

(Precursor to neural networks): Invented by Frank Rosenblatt.Simple model simulating a neuron for binary classification tasks.

1959

Reinforcement Learning

Inventor: Richard Bellman. Algorithmic method teaching agents to make decisions through rewards and penalties in environments.

1964

Support Vector Machines

(SVMs): Introduced by Vladimir Vapnik and Alexey Chervonenkis. excel in handwriting recognition, face detection, and various classification tasks.

1986

Backpropagation for Neural Networks

Popularized by David Rumelhart, Geoffrey Hinton, and Ronald Williams.Essential algorithm for training neural networks through gradient descent optimization.

1990

Boosting Algorithms

(AdaBoost) Introduced by Yoav Freund and Robert Schapire. Improves prediction by combining multiple weak learning models.

1995

Random Forests

Robust classification and regression through decision tree-like model ensembles. Introduced by Tin Kam Ho's

1997

RNN and LSTM

Inventors: Various, notable contributions by Sepp Hochreiter, Jürgen Schmidhuber (LSTM). Excelled in sequential data processing, enhancing NLP and speech recognition.

2012

Deep Convolutional Neural Networks

Inventors: Alex Krizhevsky, Ilya Sutskever, Geoffrey Hinton. Revolutionized image recognition with advanced, layered neural architectures.

2014

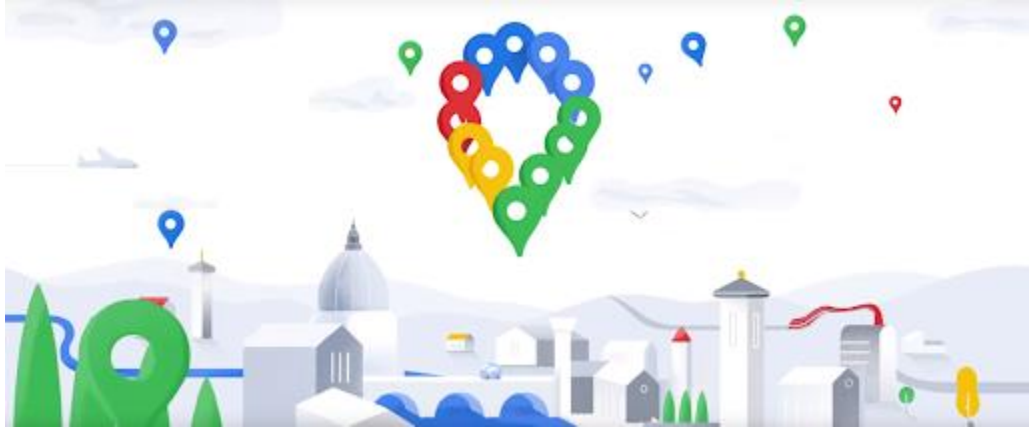
Generative Adversarial Networks

Inventor: Ian Goodfellow and colleagues. GANs have been groundbreaking in generating realistic images and videos, contributing significantly to both creative and analytical applications

2017

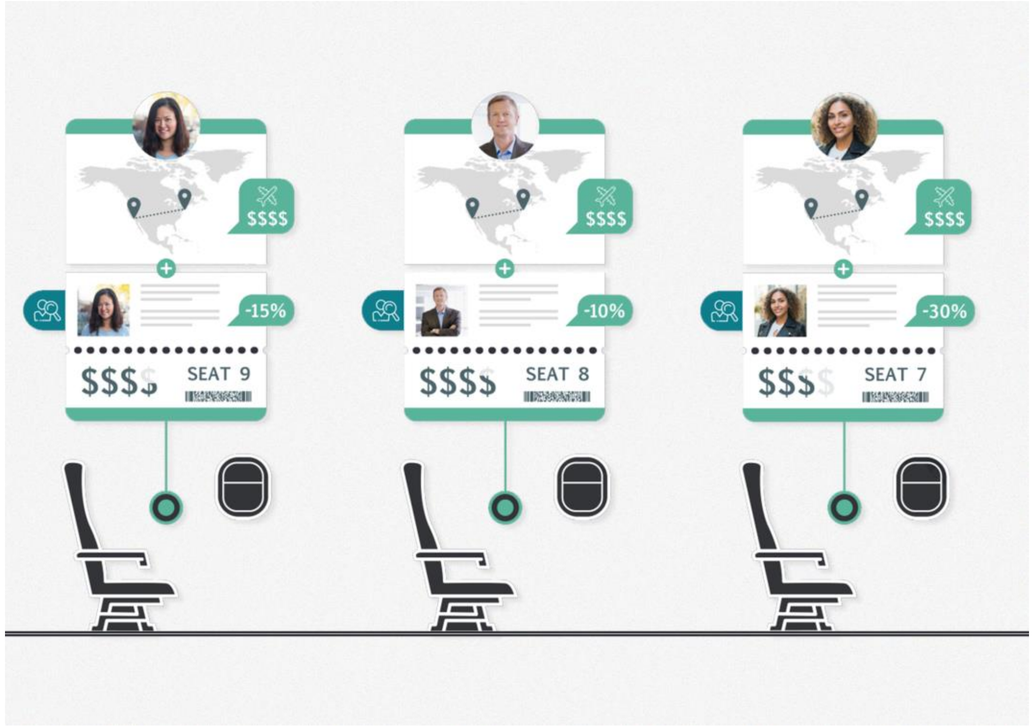
Transformer Networks

Inventors: Ashish Vaswani and colleagues. Pioneered NLP advancements with efficient context-aware processing.cornerstone for modern NLP tasks, leading to models like BERT, GPT series, and T5.



Identify which ML model they used?

Amazon Alexa/ Siri/ Google Assitant Voice Recognition
Spotify Song Recommendation
Tinder Recommendation based on your right and left swipe
Google Ads Recommendation based on your visit on webpages
Facebook/ Instagram post/ wall feed Recommendation based on your interest
Goibibo dynamic pricing of airline tickets based on demand
Face-unlock feature of your smartphone
E-mail segregation in Gmail folders (primary, spam, promotion, update, etc.)
Text prediction while composing mail in Gmail
Uber/ Ola predicting the accurate time of arrival based on real-time traffic
Uber predicting the fare estimate (surge price) during the peak hours to increase the profit



Praveen Kumar G

WhatsApp contact

Association for Computing Machinery (ACM is a non-profit professional membership group, reporting nearly 110,000 student and professional members as of 2022)

https://www.acm.org/

https://www.kaggle.com/

https://summerofcode.withgoogle.com/programs/2023/organizations/machine-learning-for-science-ml4sci

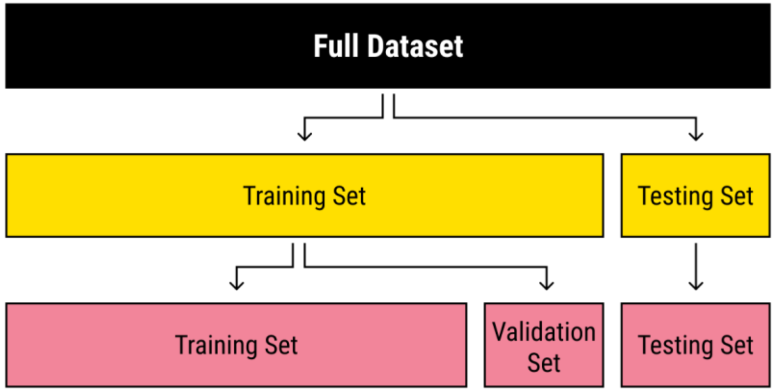
Dataset for machine learning

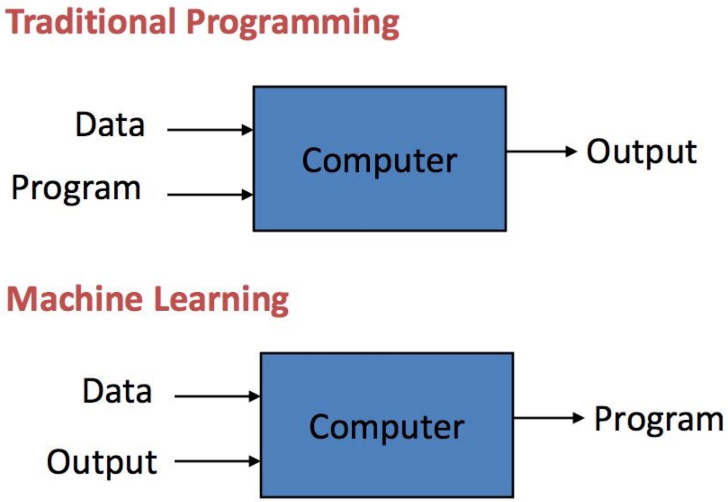
https://datasetsearch.research.google.com/search?ref=TDJjdk1URjNNakEzZDJob05BPT0sTDJjdk1URnpPVGxpTnpoa1pnPT0%3D&query=Largest%20student%20Association%20machine%20learning&docid=L2cvMTF3MjA3d2hoNA%3D%3D

Context :

Introduction to Machine Learning – Learning Paradigms – PAC learning – Version Spaces – Role of Machine Learning in Artificial Intelligence applications

Introduction to Machine Learning





- Machine learning is a subfield of artificial intelligence (AI) that allows computers to learn without being explicitly programmed.
- Machine learning is an application of artificial intelligence that involves algorithms and data automatically analyzing and making decisions without human intervention.
- It involves feeding data into algorithms that can then identify patterns and make predictions on new data.
- Machine learning is revolutionizing many industries, from healthcare and finance to manufacturing and retail.

The concept of learning in a ML system

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

Example: Spam Filtering

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

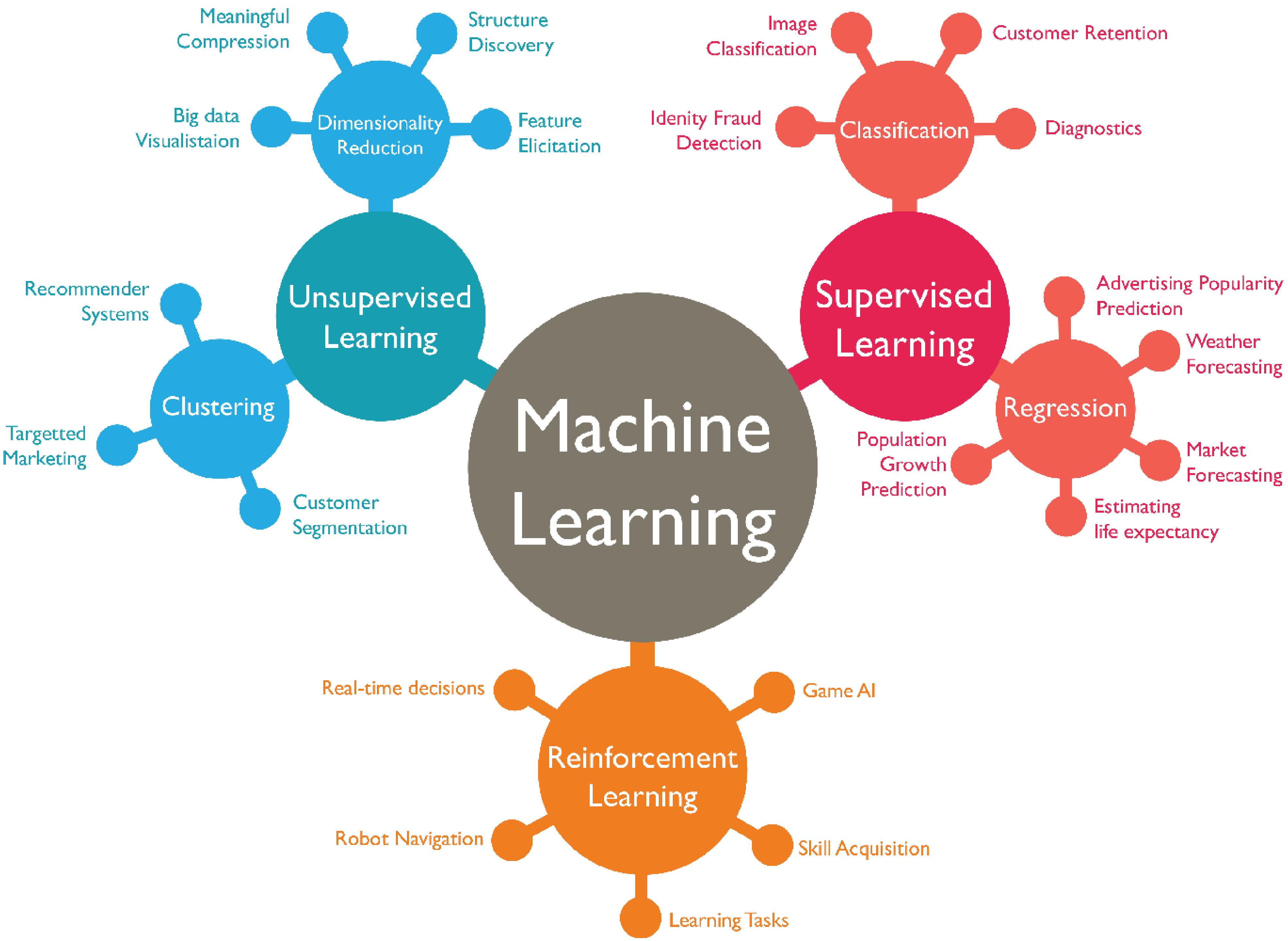
T: Identify Spam Emails

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


Machine Learning
Unsupervised Learning
Dimensionality Reduction
Meaningful Compression
Structure Discovery
Big data Visualistaion
Feature Elicitation
Clustering
Recommender Systems
Targetted Marketing
Customer Segmentation
Supervised Learning
Classification
Image Classification
Customer Retention
Idenity Fraud Detection
Diagnostics
Regression
Advertising Popularity Prediction
Weather Forecasting
Population Growth Prediction
Market Forecasting
Estimating life expectancy
Reinforcement Learning
Real-time decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks

Introduction to Machine Learning

Types of Data to ML

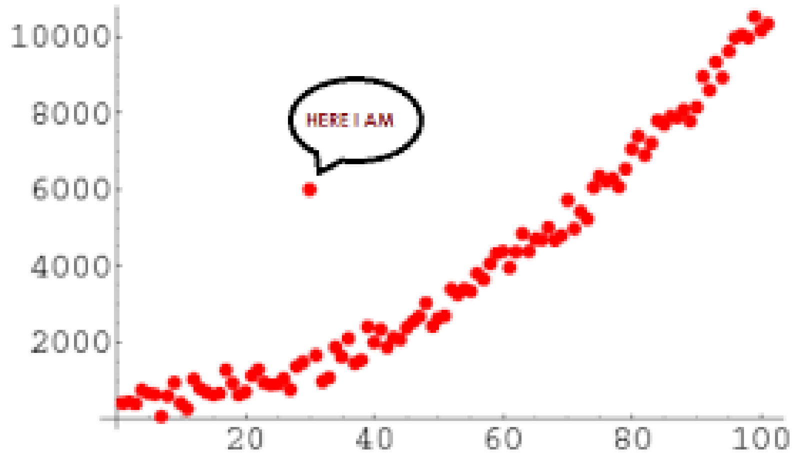
•**Numerical data:** Numbers like age, height, weight, temperature, etc.

•**Categorical data:** Labels or categories like gender, color, city, etc.

•**Text data:** Words, sentences, documents, etc.

•**Image data:** Pictures, photos, and other visual representations.

•**Audio data:** Speech, music, sound effects, etc.



Data Preparation

Before feeding data to a machine learning model, it typically undergoes several steps:

•**Data cleaning:** Handling missing values, outliers, and inconsistencies.

•**Data preprocessing:** Transforming data into a suitable format for the model (e.g., normalization, scaling, encoding).

•**Feature engineering:** Creating new features from existing ones to improve model performance.

Feeding Data to the Model

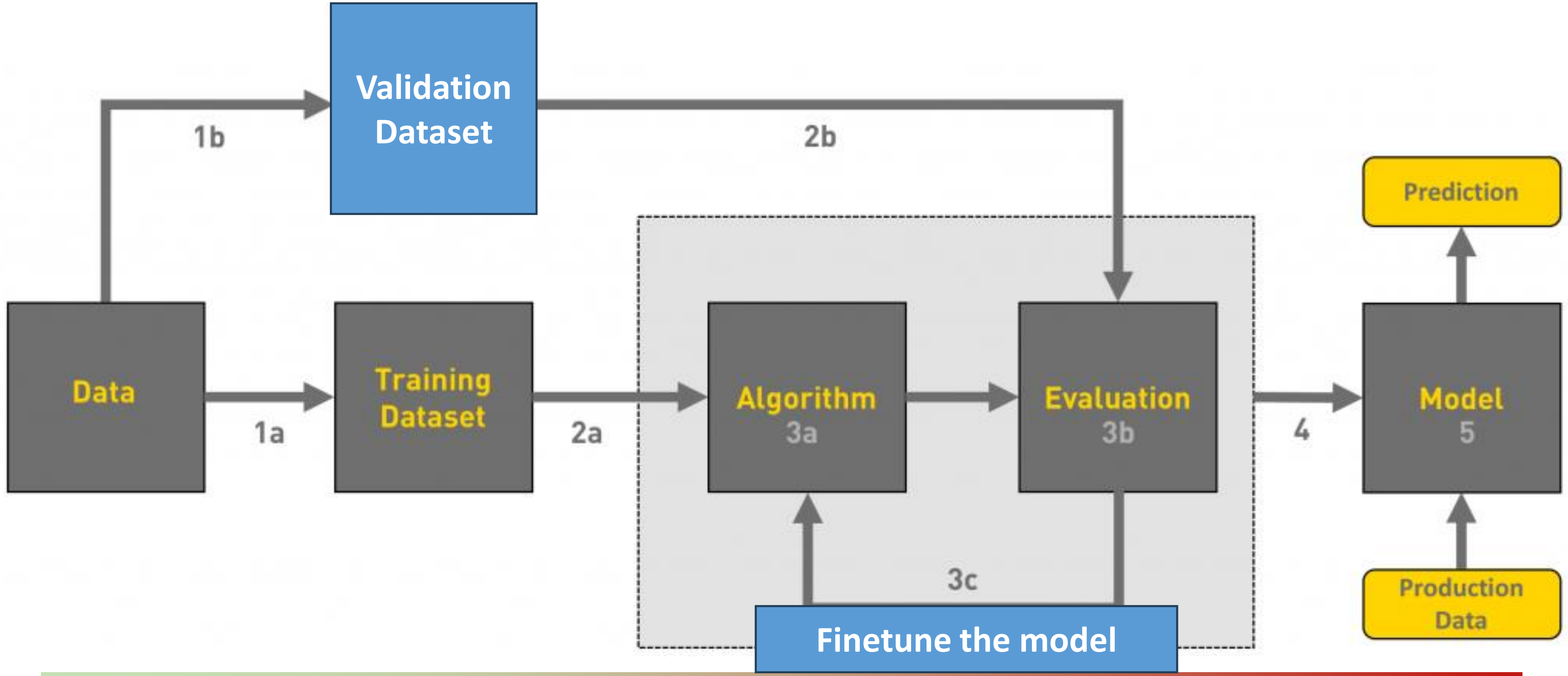
•**Training data:** Used to teach the model patterns and relationships in the data.

•**Testing data:** Used to evaluate the model's performance on unseen data.

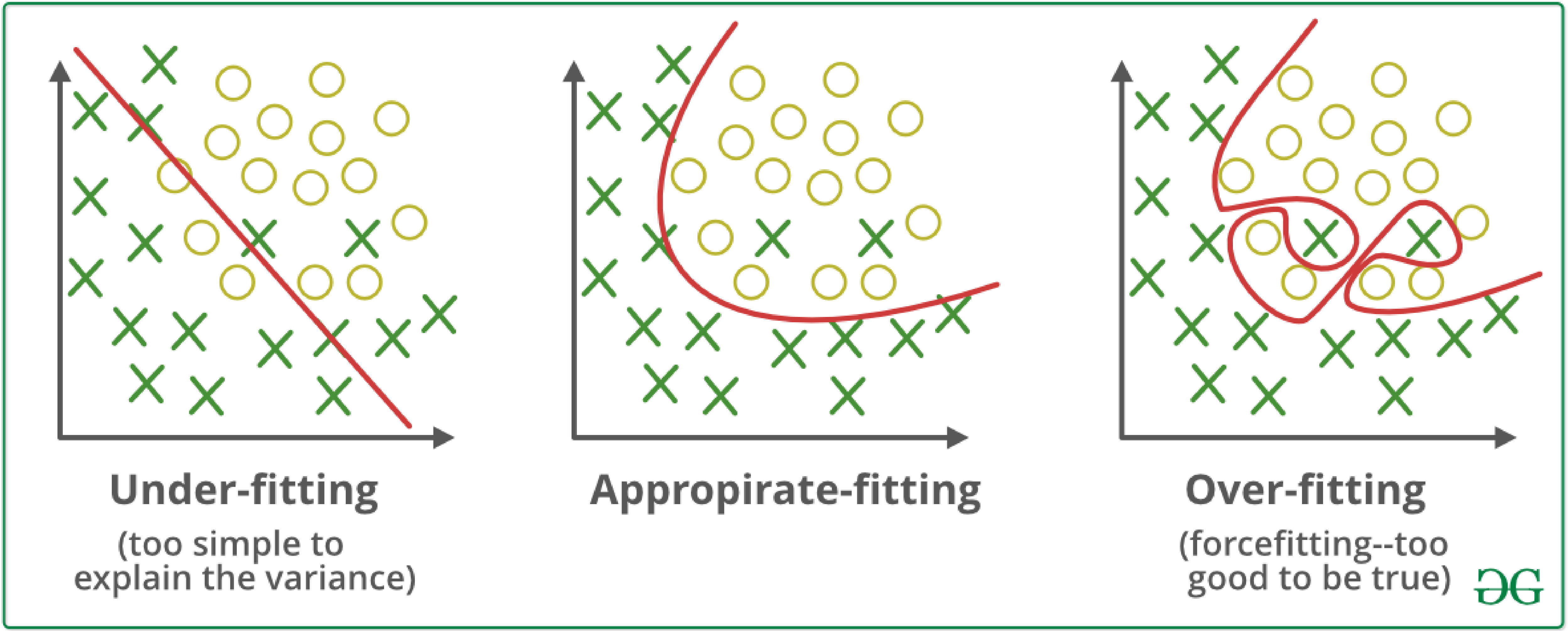
Introduction to Machine Learning

Workflow:

- Collecting and preprocessing data.
- Selecting appropriate algorithms.
- Training models.
- Evaluating performance.



Introduction to Machine Learning



Overfitting : The problem is that the model might simply memorize the training data and not be able to generalize to new data.

Bias : Bias refers to the error introduced by simplifying assumptions in the model. These assumptions make the model easier to understand but might miss the complexities of the data, leading to underfitting. High bias means the model performs poorly on both training and testing data.

Variance : Variance is the error due to the model's sensitivity to small fluctuations in the training data. High variance means the model captures noise and random fluctuations, leading to overfitting. As a result, the model performs well on training data but poorly on testing data.

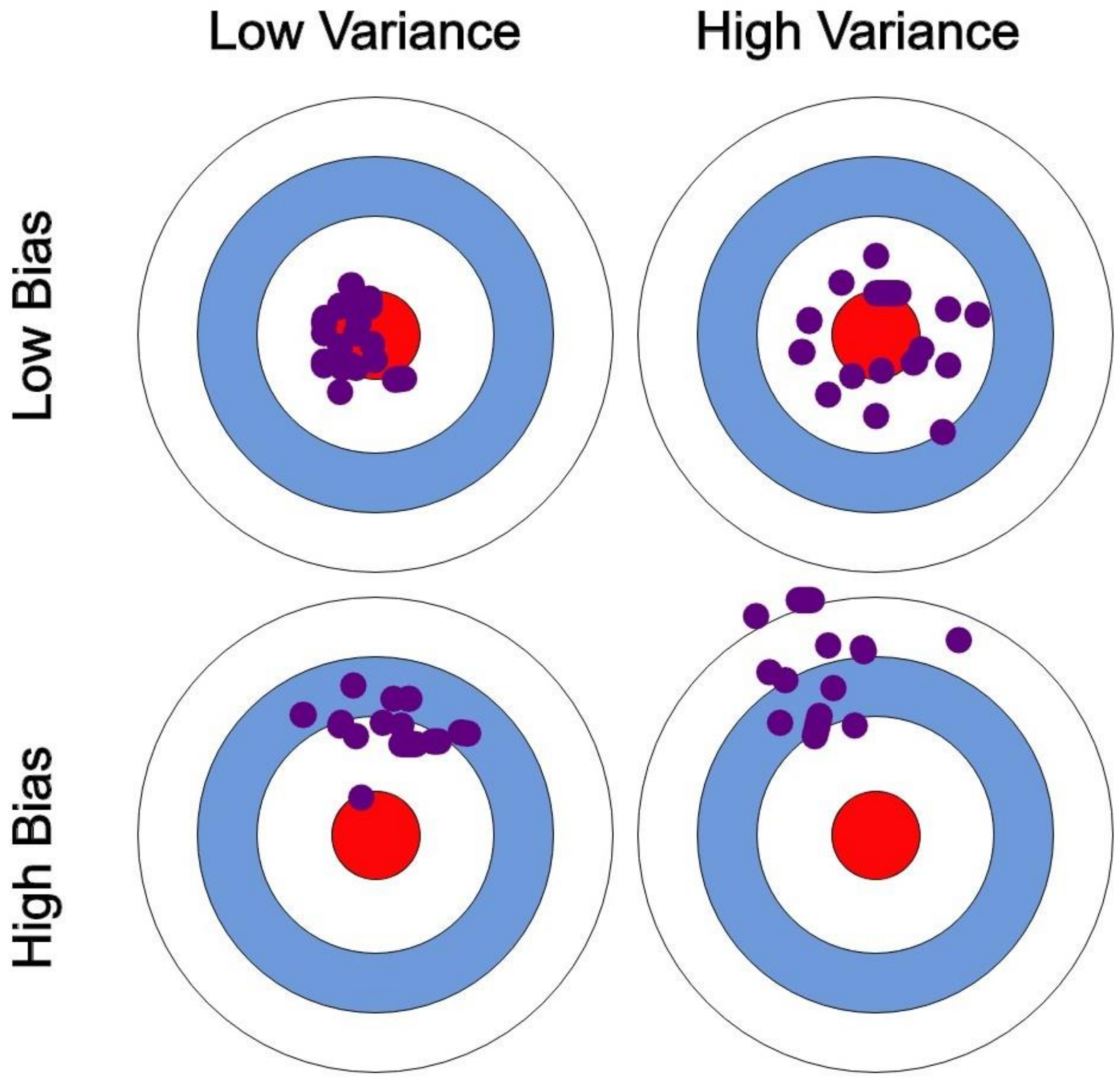
Simple linear regression : May be underfit, Polynomial regression with 10^{th} Degree : Overfit

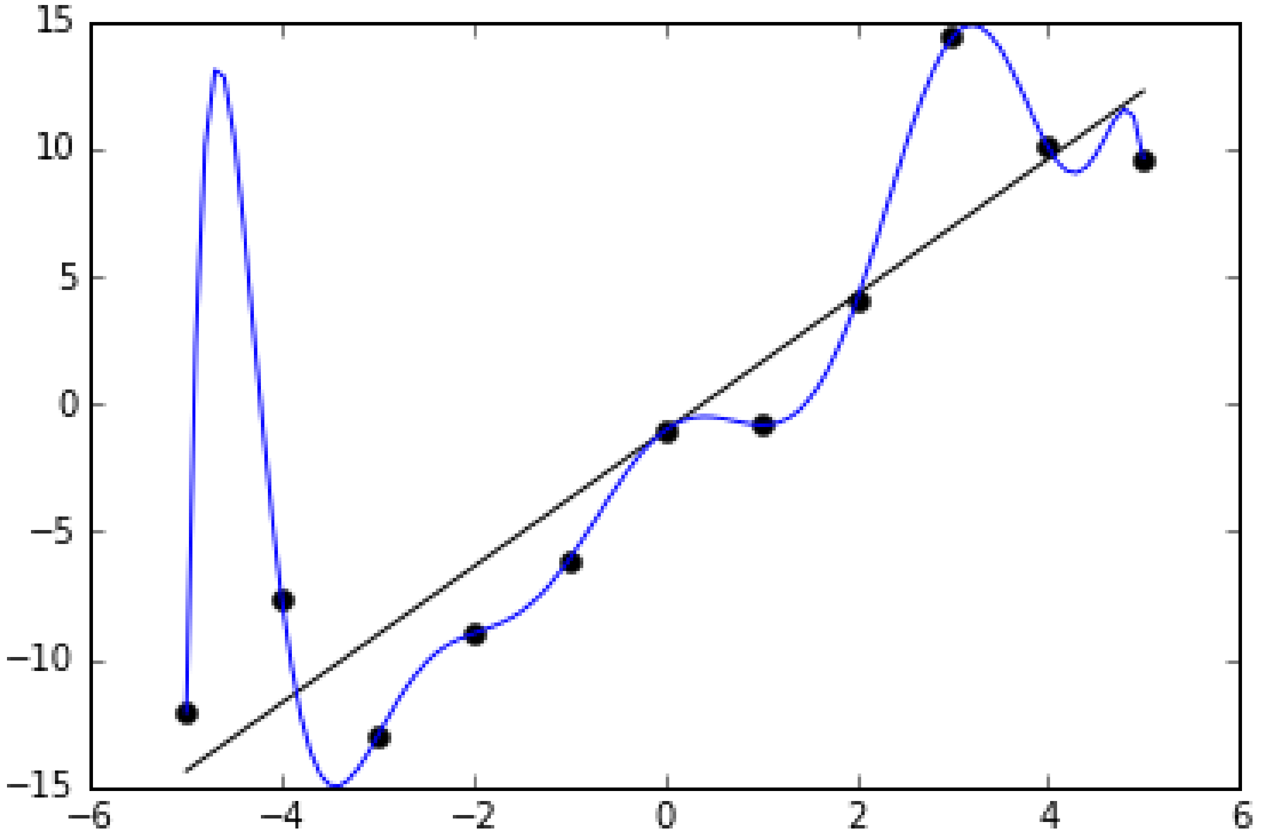
To avoid overfit : PAC model used VC (Vapnik-Chervonenkis) dimension, which is a measure of the complexity of a model. A model with a lower VC dimension is less complex.

The basic idea is that if the model is not too complex (has a low VC dimension) and is trained on a large enough dataset, then the probability of the model overfitting is low.

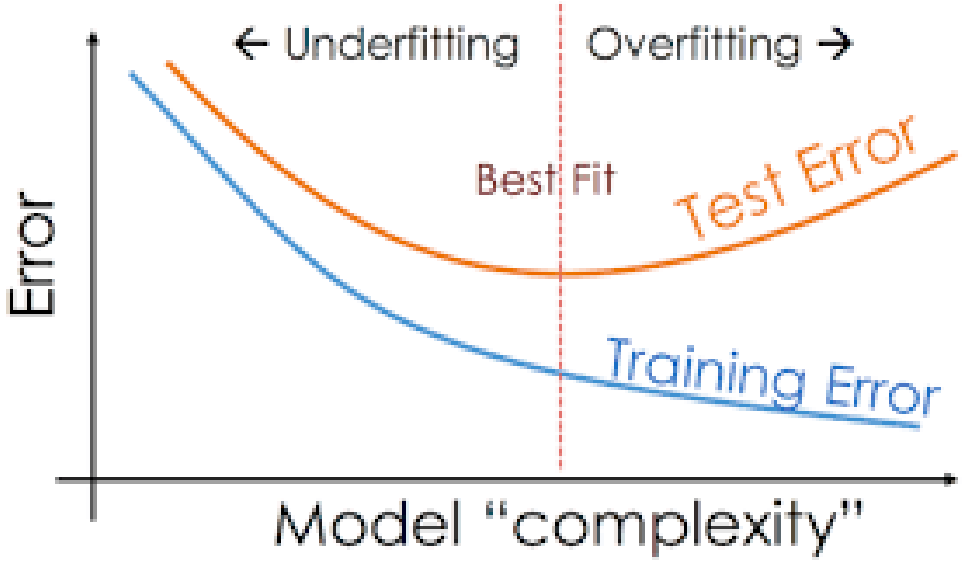
PAC learning is a different way of thinking about machine learning problems. It is a more theoretical approach, but it can be a powerful tool for understanding and improving machine learning models.

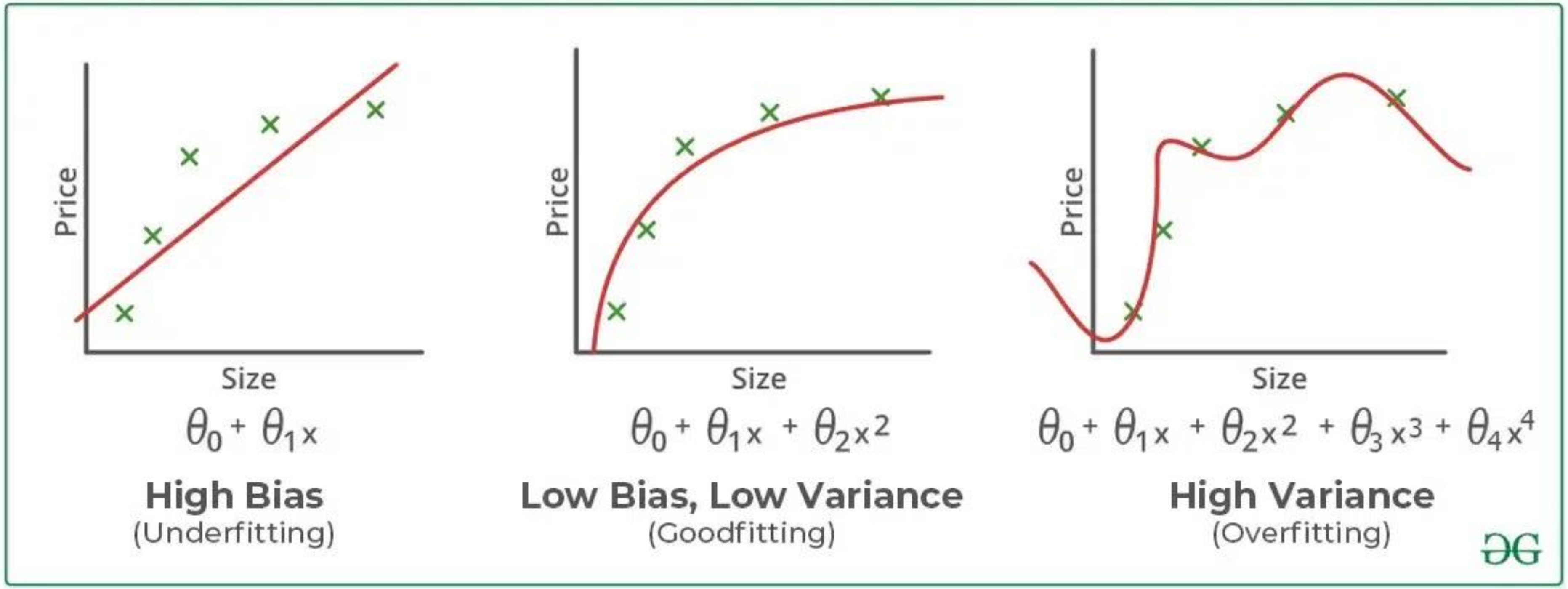
Introduction to Machine Learning





Introduction to Machine Learning





Probably Approximately Correct (PAC)

In machine learning, train algorithms to learn from data. The size of the dataset is important because more data usually helps the algorithm learn better. However, it's also crucial to understand **what** the algorithm can learn well and **how well** it can learn from the data. This is where the PAC learning framework comes in.

PAC learning helps us answer two main questions:

1. **What can the algorithm learn efficiently?**
2. **How many training examples are needed to achieve good results?**

Key Terms

- **Concept ©**: A feature or pattern we want to learn, like whether an email is spam or not.
- **Concept Class ©**: A set of all possible concepts we want to learn.
- **Hypothesis (H)**: A set of possible solutions or models that the algorithm can choose from.
- **Data Distribution (D)**: How the data is spread out or distributed.
- **Sample (S)**: A subset of data used for training.
- **Hypothesis for Sample (hS)**: The model or solution chosen based on the sample.
- **Accuracy Parameter (ε)**: How close the model's predictions are to the actual values.
- **Confidence Parameter (δ)**: The probability that the model's accuracy is within the desired range.

Probably Approximately Correct (PAC)

PAC Learning

A concept class (C) is PAC learnable if, after training on a number of samples (N), the hypothesis (H) returned by the algorithm has an error rate less than (epsilon) with a probability of at least (1 - delta). This means the model is “probably approximately correct.”

Example

Imagine we are training a model to classify emails as spam or not spam. Here’s how PAC learning applies:

Concept ©: Whether an email is spam (1) or not spam (0).

Concept Class ©: All possible ways to classify emails.

Hypothesis (H): Different models that can classify emails.

Data Distribution (D): The way emails are distributed in our dataset.(frequency of spam emails vs. non-spam emails, the length of email content, presence of specific keywords, etc)

Sample (S): A subset of emails used for training.

Hypothesis for Sample (hS): The model trained on the sample.

Accuracy Parameter (ε): We want the model’s error rate to be less than 5%. (During training)

Confidence Parameter (δ): We want to be 95% confident in our model’s accuracy. (after training)

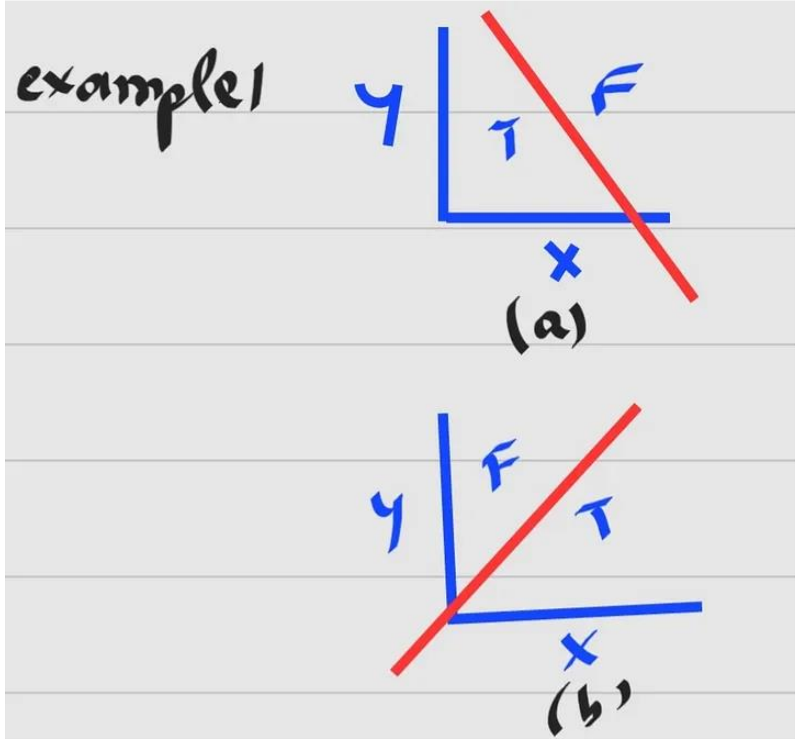
Probably Approximately Correct (PAC)

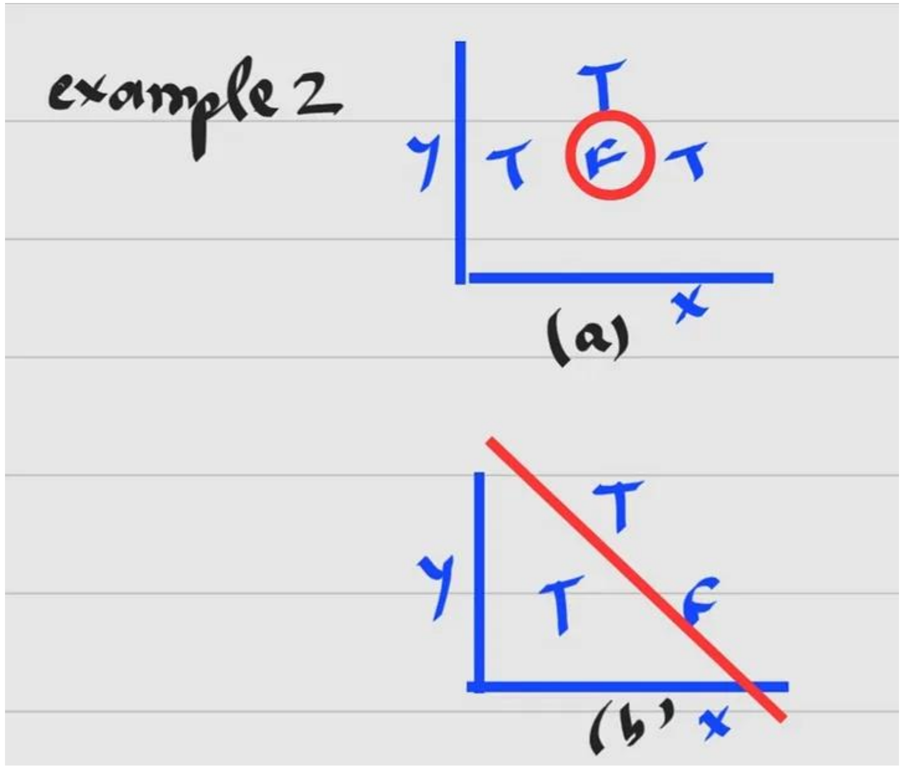
VC Dimension : Measure of the capacity of a hypothesis space (a set of functions or models). It tells us the largest number of points that can be shattered (perfectly split) by the hypothesis space.

Shattering means that for every possible way of labeling a set of points, there exists a hypothesis in the hypothesis space that can correctly classify those points.

Key Concepts

1. **Finite Hypothesis Space**: If the hypothesis space is finite, we can directly measure its capacity.
2. **Infinite Hypothesis Space**: For infinite hypothesis spaces, we use the concept of VC dimension to understand their capacity.





Types of Machine Learning:

Supervised Learning:

- Program given labeled input data and expected output data.
- Generates classification (class notification) or regression (numerical value prediction) results.

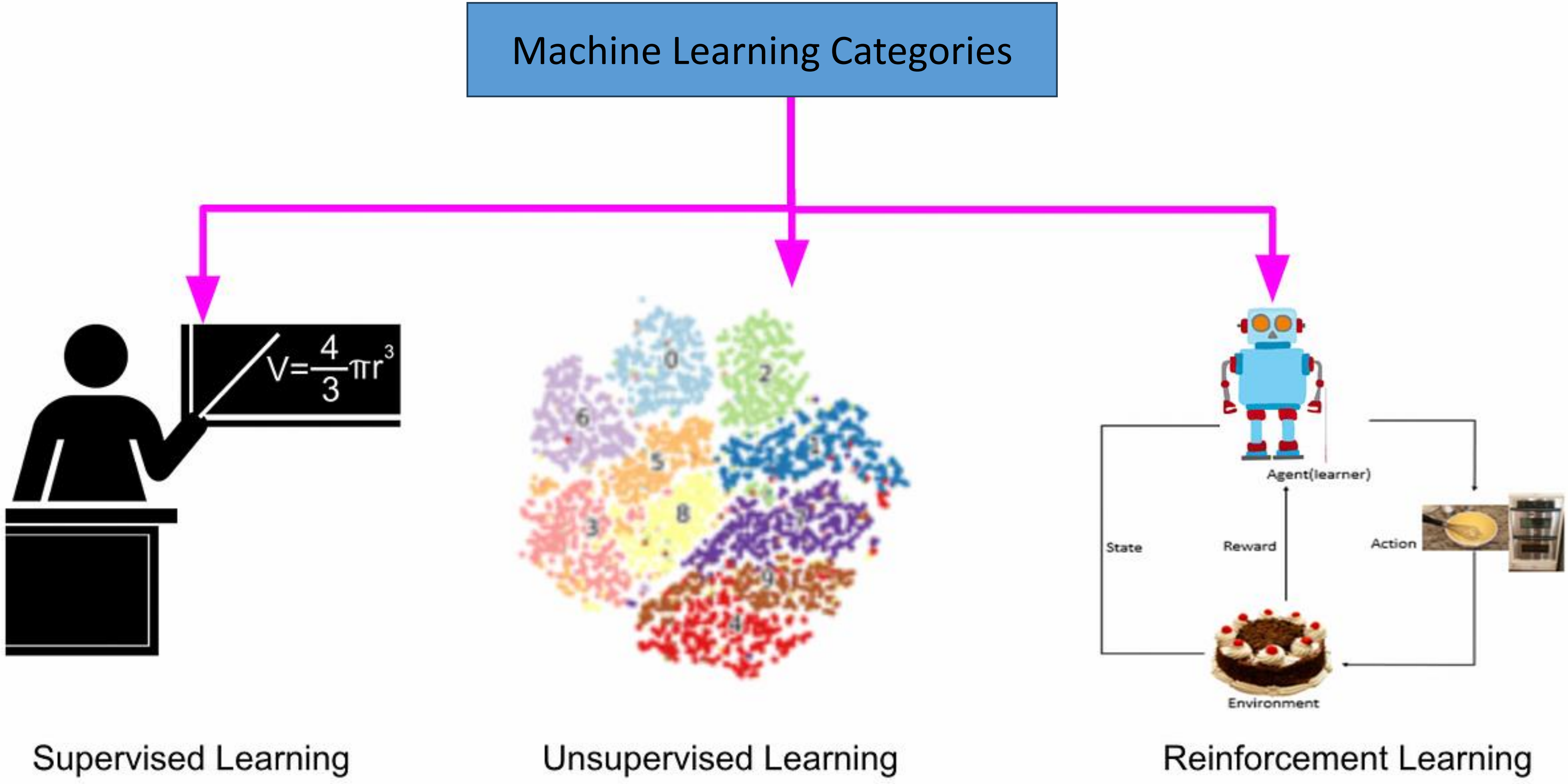
Unsupervised Learning:

- Works with input data without labeled responses.
- Automatically identifies structures in data.

Reinforcement Learning:

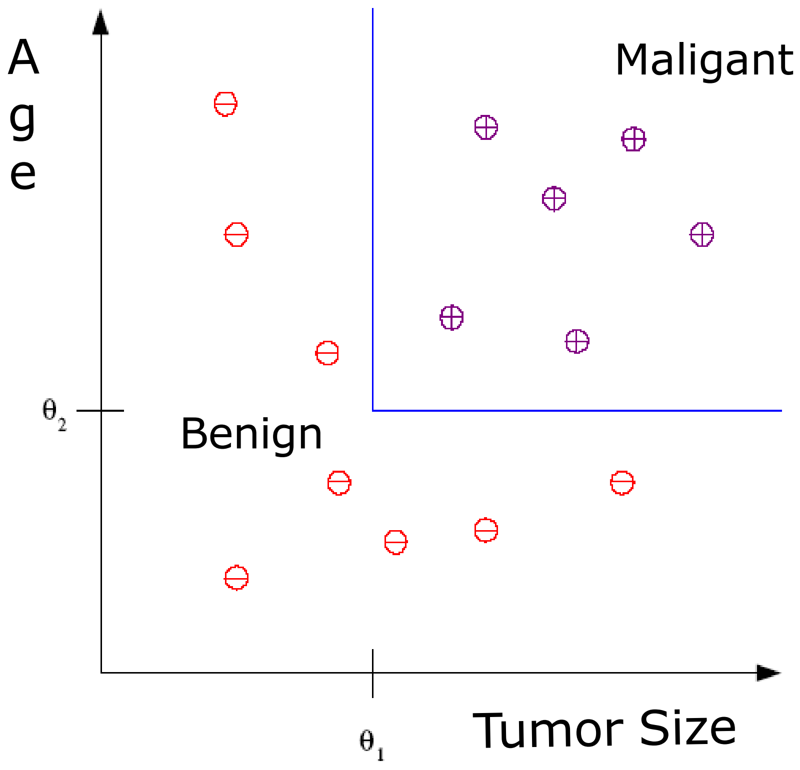
- Used for making a sequence of decisions.
- Learning by interacting with the environment.
- Based on rewarding and punishing actions.

Introduction to Machine Learning



Supervised Learning : Classification

- Example : Breast cancer
- Differentiating tumors as malignant or benign from patient's age and tumor size



Discriminant: IF *age* > θ_1 AND *tumor_size* > θ_2

THEN maligant ELSE benign

Supervised Learning : Regression

Example: Price of House

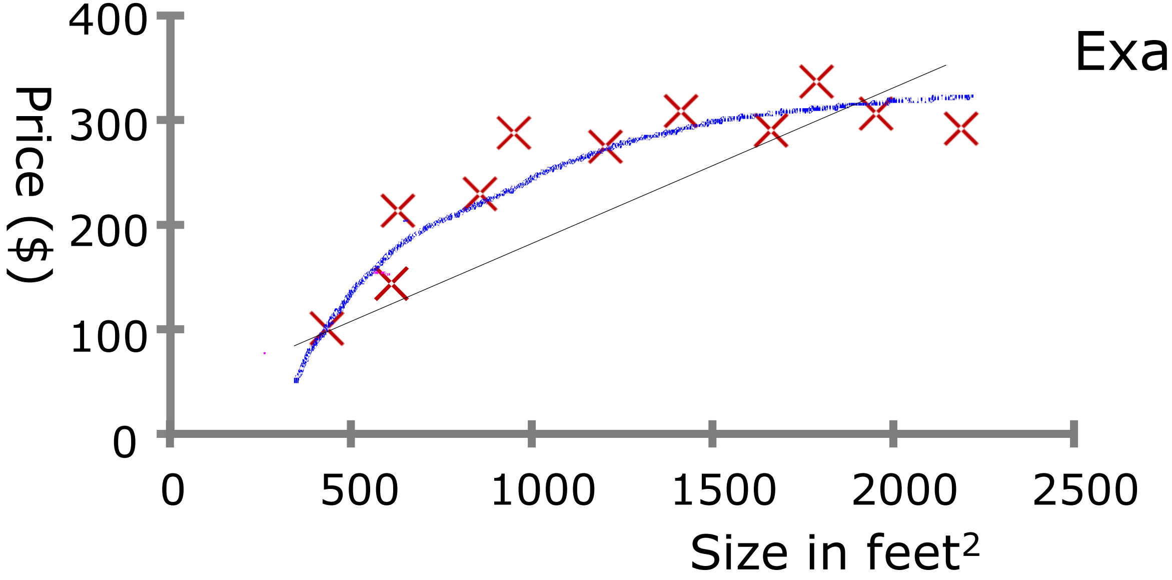
x : House attributes

y : price

$y = g\ (x \mid \theta\ )$

g ( ) model,

θ parameters



Example:

Learning "What normally happens"

No output

Clustering: Grouping similar instances

Other applications: Summarization, Association Analysis

Example applications

Customer segmentation in CRM

Image compression: Color quantization (K-means)

Bioinformatics: Learning motifs: DNA/ RNA protein sequence (expectation-maximization (EM)

Applications of Machine Learning:

- Traffic Prediction:

  GPS navigation services track locations to manage traffic.

- Virtual Personal Assistants:

  Smart speakers, smartphones, and apps like Google Allo.

- Online Transportation:

  Apps like Uber compute available vehicles, estimated cost, and travel distance.

- Social Media Services:

  Personalizing news feeds and suggesting connections.

- Email Spam Filtering:

  Identifying and filtering spam emails.

Introduction to Machine Learning

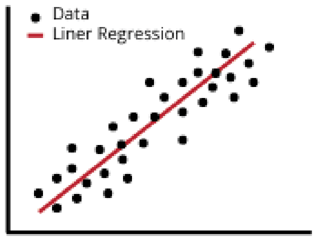
Role of Machine Learning in Artificial Intelligence:

- Machine learning is a crucial component of AI.
- AI systems leverage machine learning to act intelligently, adapt, and improve over time.
- Examples: Chatbots, recommendation engines, image recognition, and natural language processing.

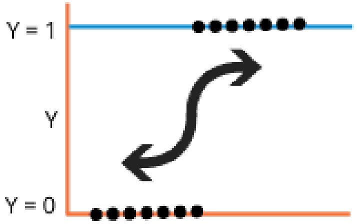
Applications include:

- Self-driving cars
- Facial recognition
- Natural language processing
- Fraud detection
- Medical diagnosis

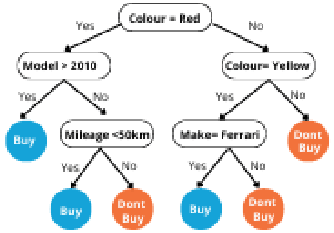
Linear Regression



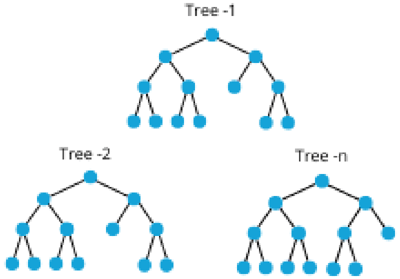
Logistic Regression



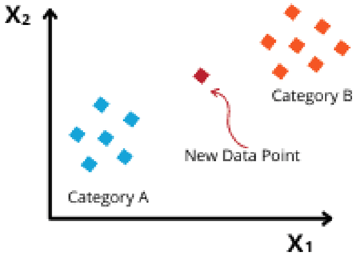
Decision Trees



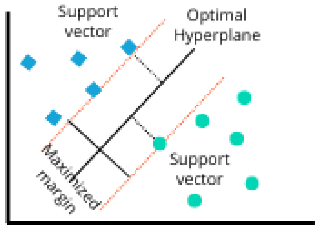
Random Forest



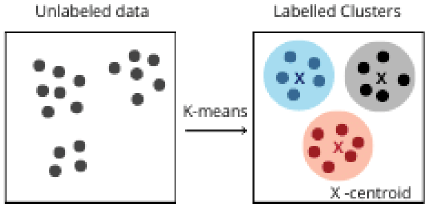
K-Nearest Neighbor



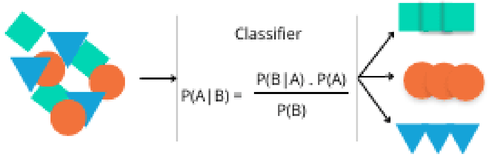
Support Vector Machine



K-Means Clustering

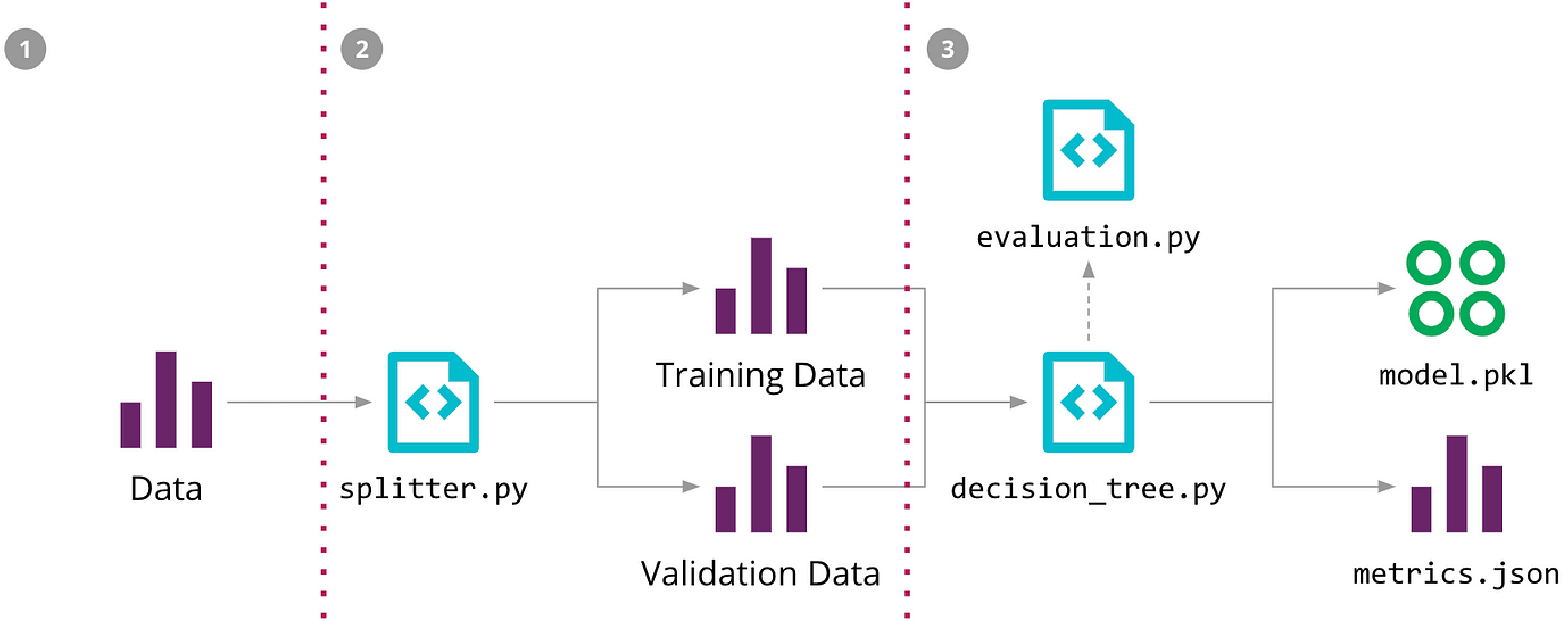


Naïve Bayes



Introduction to Machine Learning

1. Data Acquisition and Preprocessing
2. Model Development and Training
3. Model Evaluation and Deployment



Introduction to Machine Learning

Examples: Items or instances of data used for learning or evaluation.

Features: The set of attributes associated to an example.

Labels: Values or categories assigned to examples.

Training sample: Examples used to train a learning algorithm.

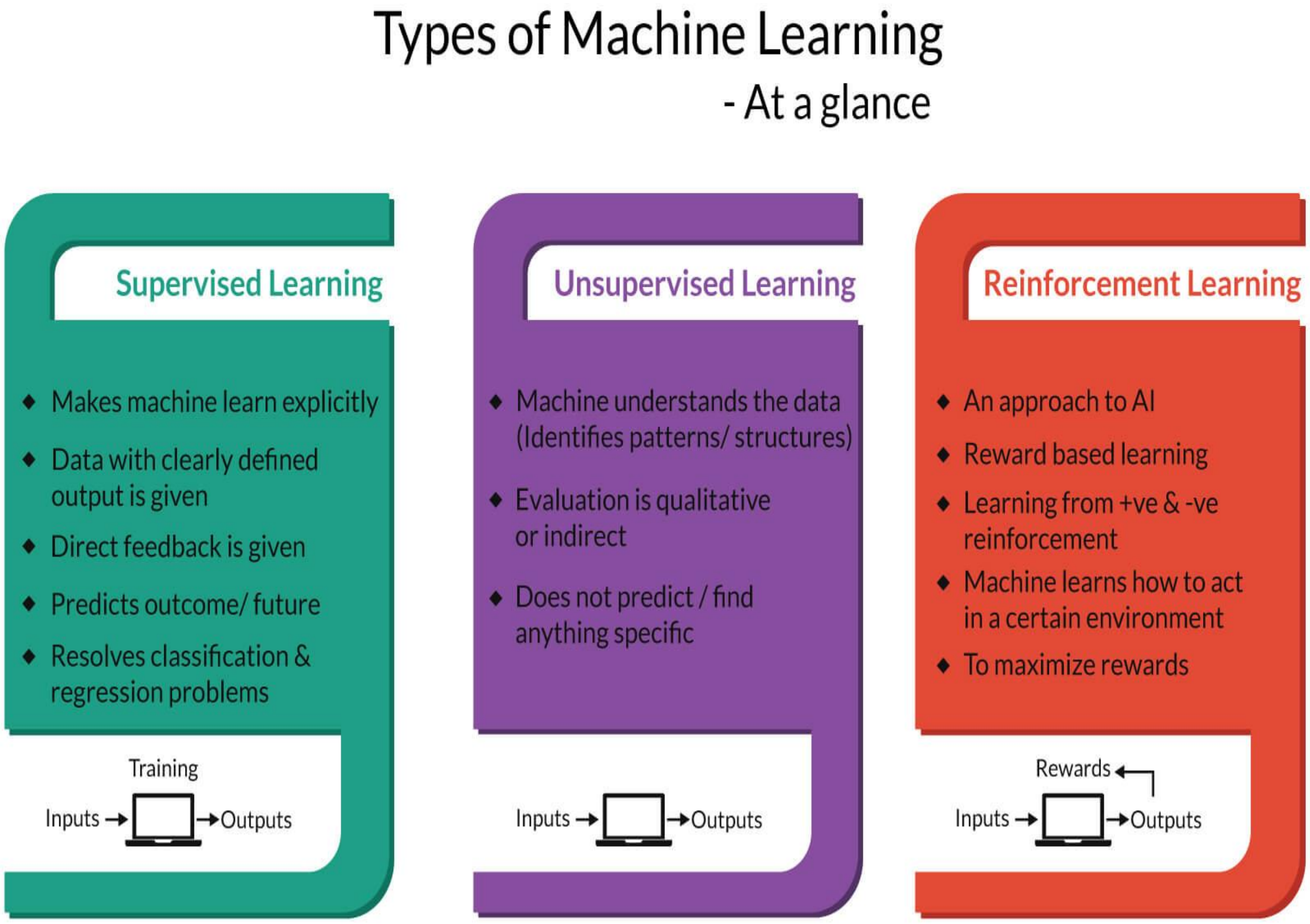
Validation sample: Examples used to tune the parameters of a learning algorithm

Test sample: Examples used to evaluate the performance of a learning algorithm.

Loss function: A function that measures the difference, or loss, between a predicted label and a true label.

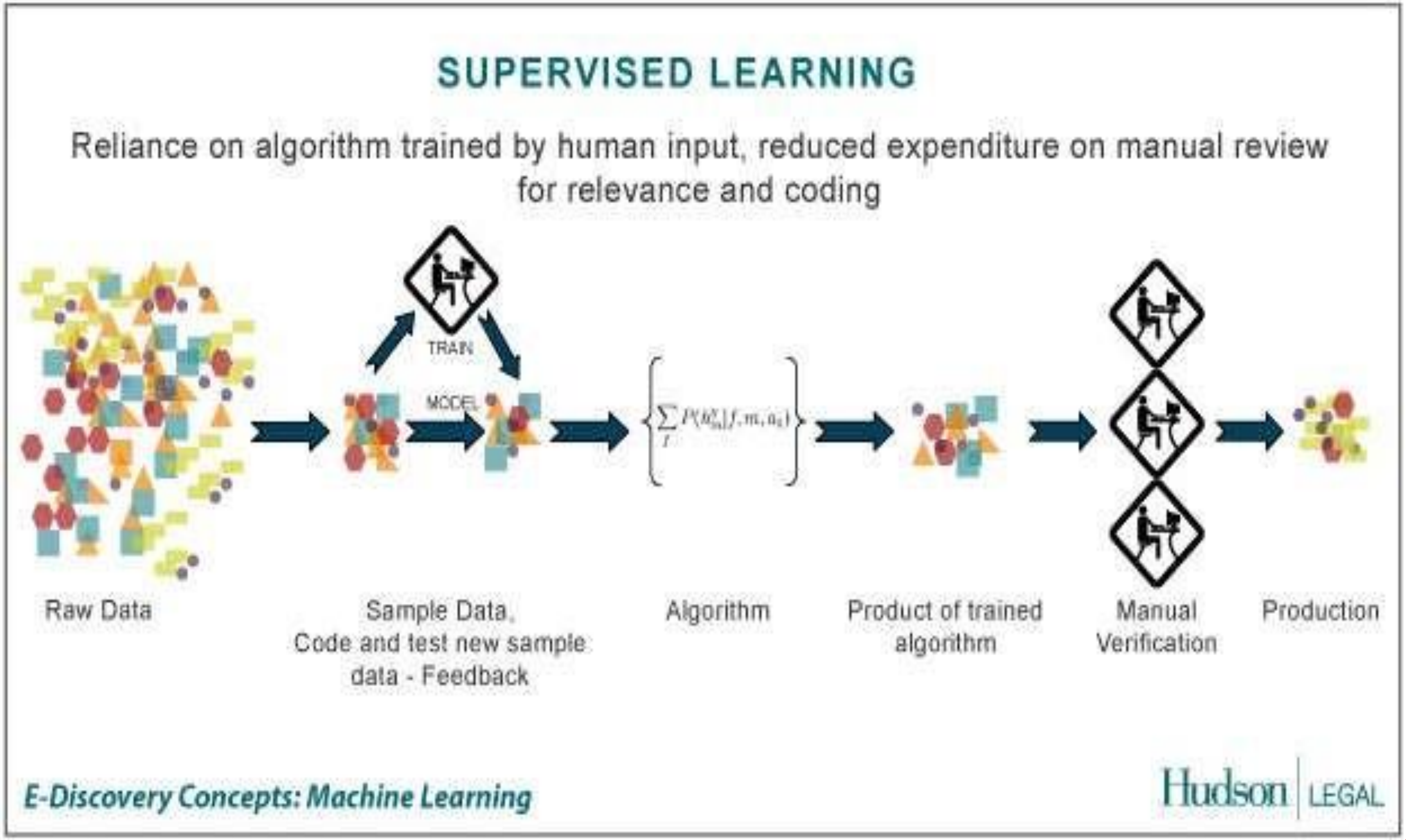
Hypothesis set: A set of functions mapping features (feature vectors) to the set of labels Y.

Introduction to Machine Learning



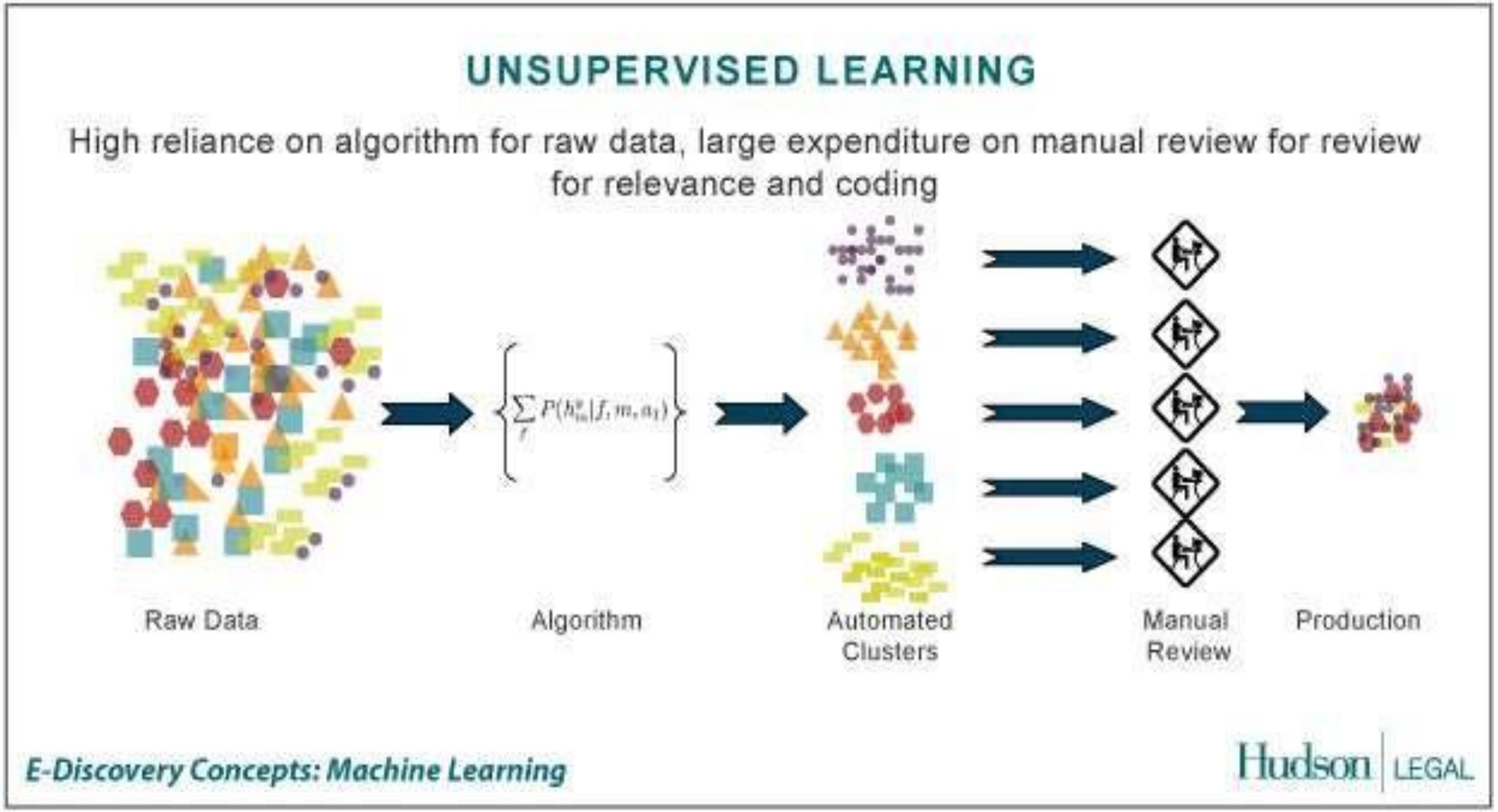
Supervised Learning

- the correct classes of the training data are known



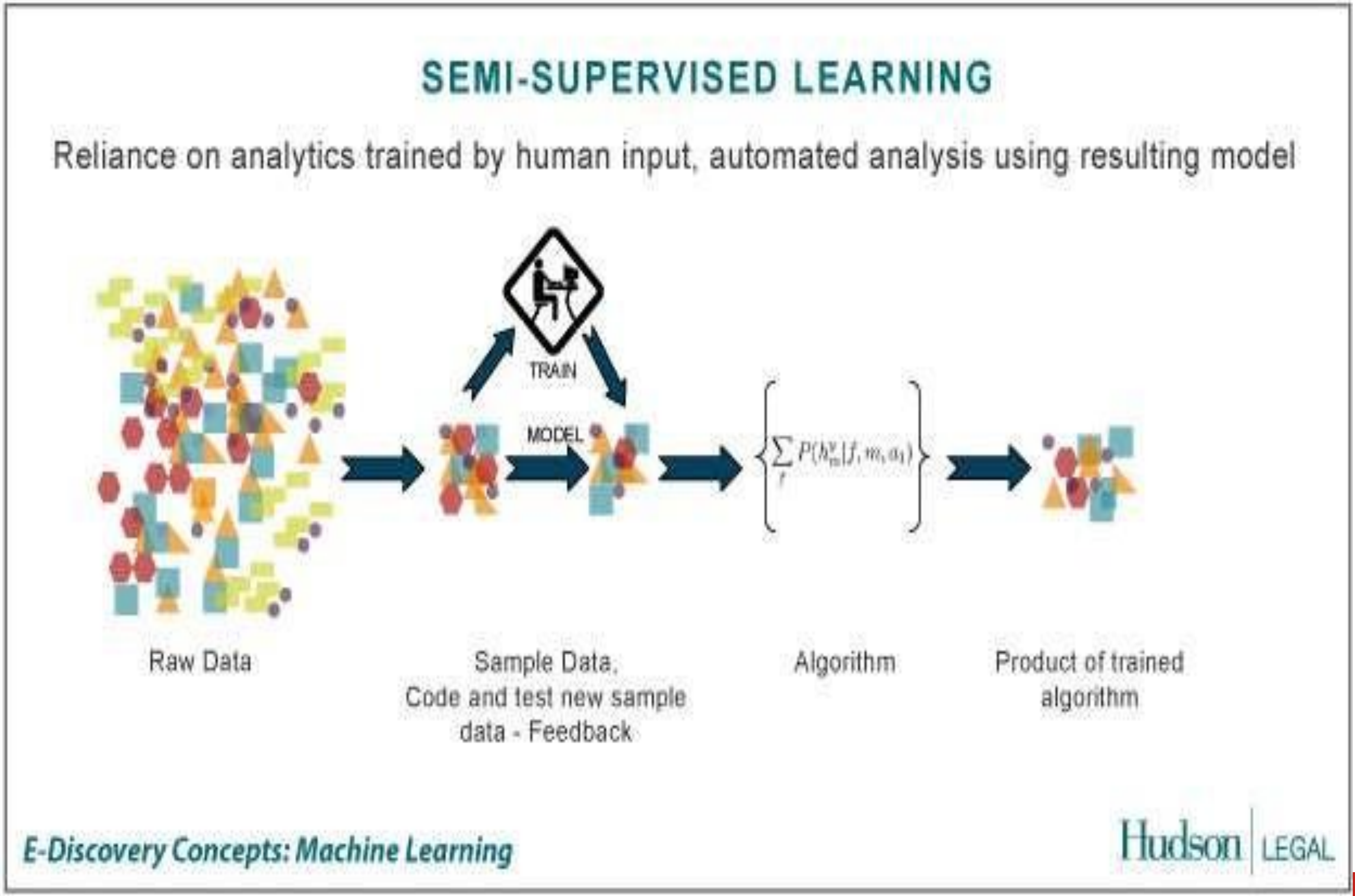
Unsupervised Learning

- the correct classes of the training data are not known



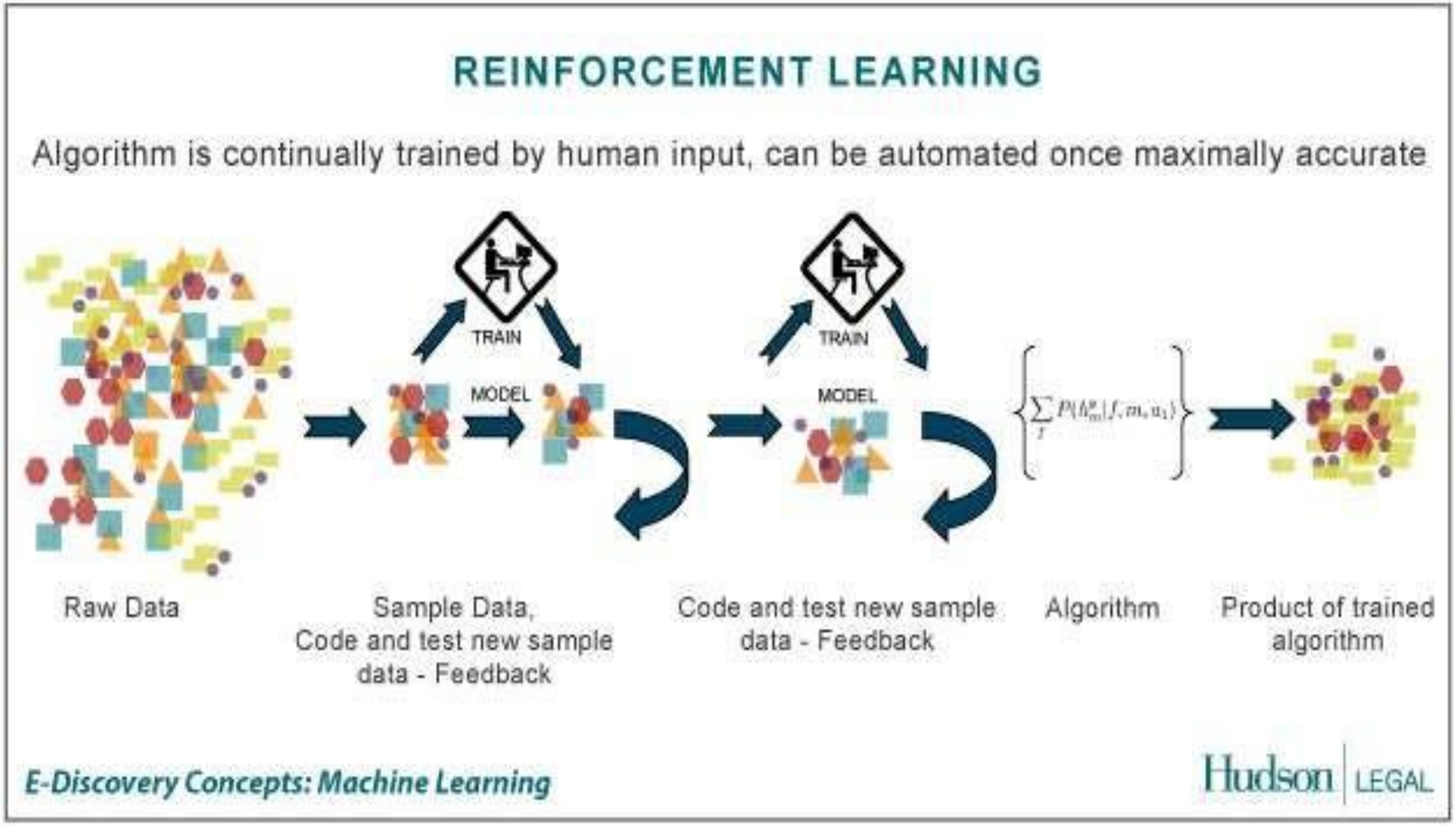
Semi-Supervised Learning

- A Mix of Supervised and Unsupervised learning



Reinforcement Learning

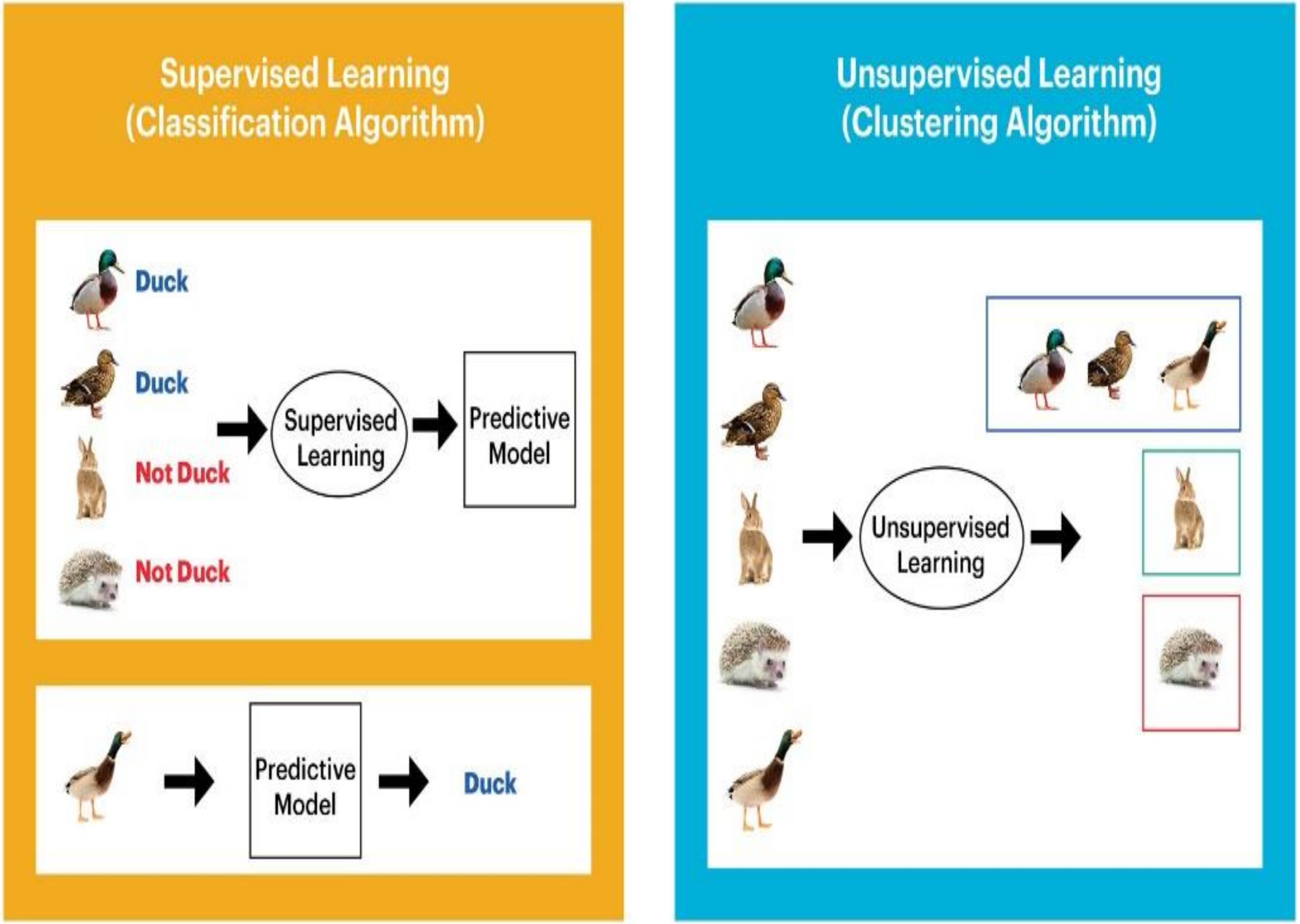
- allows the machine or software agent to learn its behavior based on feedback from the environment.
- This behavior can be learnt once and for all, or keep on adapting as time goes by.



Machine Learning Techniques

- classification: predict class from observations
- clustering: group observations into "meaningful" groups
- regression (prediction): predict value from observations

Introduction to Machine Learning

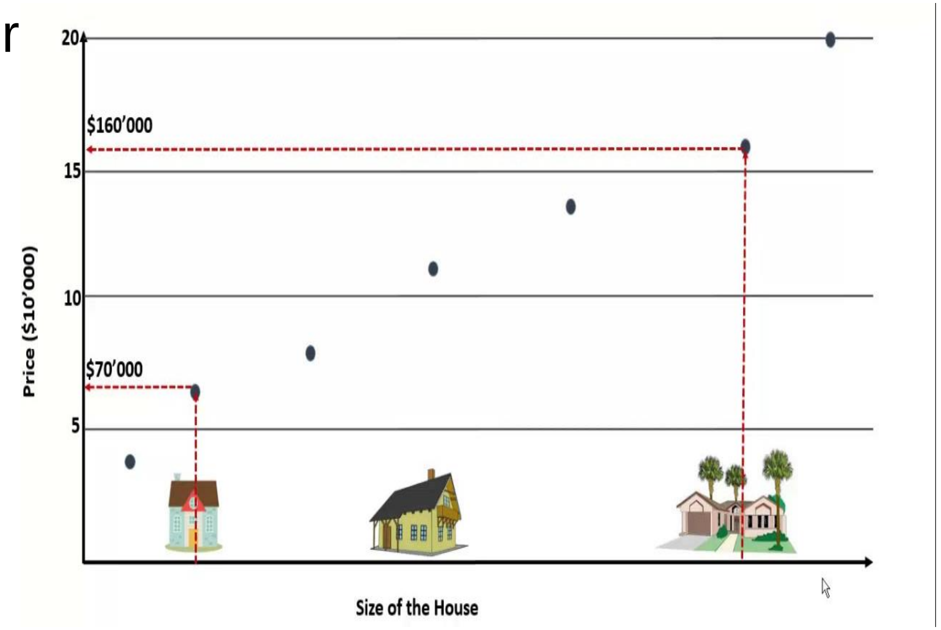


Classification

- classify a document into a predefined category.
- documents can be text, images
- Popular one is Naive Bayes Classifier.
- Steps:
  – Step1 : Train the program (Building a Model) using a training set with a category for e.g. sports, cricket, news,
  – Classifier will compute probability for each word, the probability that it makes a document belong to each of considered categories
  – Step2 : Test with a test data set against this Model

Regression

- is a measure of the relation between the mean value of one variable (e.g. output) and corresponding values of other variables (e.g. time andcost).
- **regression analysis** is a statistical process for estimating the relationships among variables.
- Regression means to **predict** the output value using training data.
- Popular one is Logistic regression (binary regression)



Classification vs Regression

- Classification means to group the output into a class.
- classification to **predict** the type of tumor i.e. harmful or not harmful using training data
- if it is discrete/categorical variable, then it is classification problem

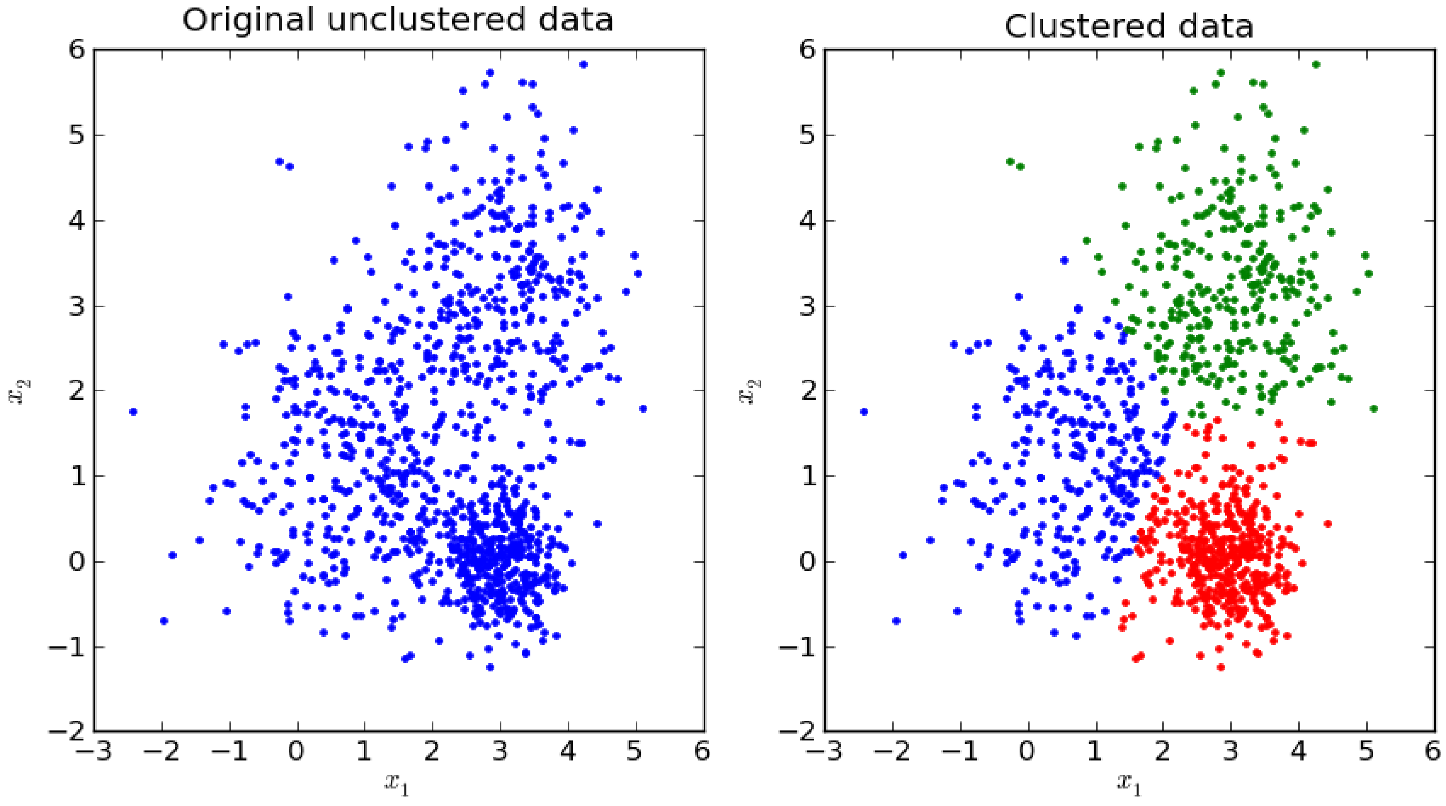
- Regression means to predict the output value using training data.
- regression to **predict** the house price from training data
- if it is a real number/continuous, then it is regression problem.

Clustering

- **clustering** is the task of grouping a set of objects in such a way that objects in the same group (called a **cluster**) are more similar to each other
- objects are not predefined
- For e.g. these keywords
  - "man's shoe"
  - "women's shoe"
  - "women's t-shirt"
  - "man's t-shirt"
  - can be cluster into 2 categories "shoe" and "t-shirt" or "man" and "women"
- Popular ones are **K-means clustering** and **Hierarchical clustering**

K-means Clustering

- partition ***n*** observations into ***k*** clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.
- http://en.wikipedia.org/wiki/K-means_clustering



Hierarchical clustering

- method of cluster analysis which seeks to build a hierarchy of clusters.
- There can be two strategies
  - **Agglomerative**:
    - This is a "bottom up" approach: each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.
    - Time complexity is O(n^3)
  - **Divisive**:
    - This is a "top down" approach: all observations start in one cluster, and splits are performed recursively as one moves down the hierarchy.
    - Time complexity is O(2^n)

- http://en.wikipedia.org/wiki/Hierarchical_clustering

Machine Learning Algorithms *(sample)*

Unsupervised

Continuous

- Clustering & Dimensionality Reduction
  - SVD
  - PCA
  - K-means

Categorical

- Association Analysis
  - Apriori
  - FP-Growth
- Hidden Markov Model

Supervised

Continuous

- Regression
  - Linear
  - Polynomial
- Decision Trees
- Random Forests

Categorical

- Classification
  - KNN
  - Trees
  - Logistic Regression
  - Naive-Bayes
  - SVM

Concept Learning as Search:

Find-S Algorithm

The Find-S algorithm is a simple and intuitive method used for concept learning. It focuses on finding the most specific hypothesis that fits all positive examples in the training data.

Steps:

1. Start with the most specific hypothesis (usually the null hypothesis).
2. For each positive example, generalize the hypothesis to include the example.
3. Ignore negative examples.
4. Repeat until all positive examples are covered.

Candidate Elimination Algorithm

The Candidate Elimination algorithm is more robust and flexible compared to Find-S. It maintains a version space, which is the set of all hypotheses consistent with the training data. The algorithm refines this space by considering both positive and negative examples.

Steps:

1. Initialize the version space with the most general (G) and most specific hypotheses (S).
2. For each example:
3. If positive, generalize the specific boundary.
4. If negative, specialize the general boundary.
5. Update the version space by removing inconsistent hypotheses.

Steps of Find-S Algorithm

Example: Diagnosing a Disease Based on Symptoms

Example	Fever	Cough	Fatigue	Disease
1	Yes	Yes	Yes	Yes
2	Yes	No	Yes	Yes
3	No	Yes	Yes	No
4	Yes	Yes	No	Yes
5	No	No	Yes	No

Initialize the most specific hypothesis: (h = (\text{Ø, Ø, Ø}))
Process each positive example and generalize the hypothesis:
Example 1: (h = (\text{Yes, Yes, Yes}))
Example 2: (h = (\text{Yes, Ø, Yes})) (since the second symptom is different)
Example 4: (h = (\text{Yes, Ø, Ø})) (since the third symptom is different)

Final Hypothesis
The final hypothesis (h = (\text{Yes, Ø, Ø})) suggests that the disease is likely present if the patient has a fever, regardless of the other symptoms.

Interpretation : Find-S algorithm helps identify that having a fever is a crucial symptom for diagnosing the disease.

Candidate Elimination Algorithm

Example	Color	Shape	Label
1	Red	Round	Apple
2	Green	Round	Apple
3	Red	Square	Not Apple
4	Yellow	Round	Apple

Initial Hypotheses
G: {?, ?} (most general)
S: {Ø, Ø} (most specific)

Processing Examples
Example 1 (Red, Round, Apple):
S: {Red, Round}
G: {?, ?}

Example 2 (Green, Round, Apple):
S: {?, Round}
G: {?, ?}

Example 3 (Red, Square, Not Apple):
S: {?, Round}
G: {?, Round}

Example 4 (Yellow, Round, Apple):
S: {?, Round}
G: {?, Round}

Final Hypotheses
G: {?, Round}
S: {?, Round}

The final hypothesis is that an apple is any fruit that is round, regardless of its color.

Candidate Elimination Algorithm

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	Cool	Change	Yes