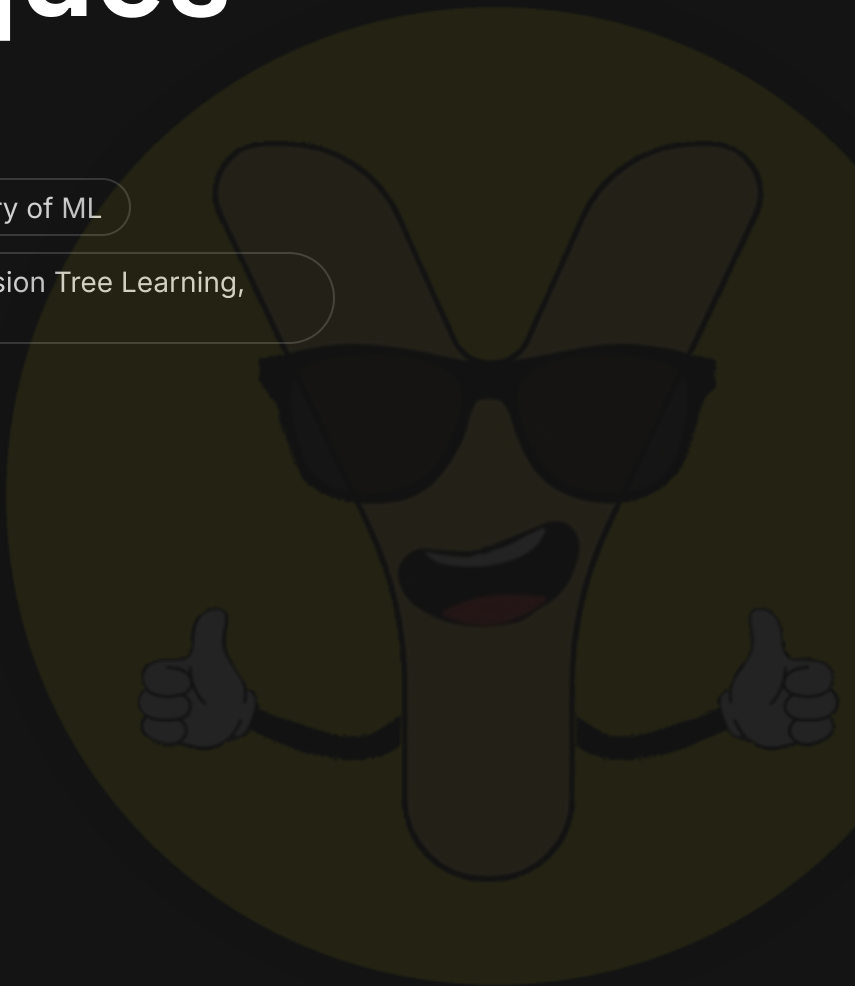
AKTU · 3RD YEAR · SEMESTER 6 · BCDS062

Machine Learning Techniques

Unit 1 — Introduction to Machine Learning

- Learning
- Types of Learning
- Well defined learning problems
- Designing a Learning System
- History of ML
- Introduction of Machine Learning Approaches (Artificial Neural Network, Clustering, Reinforcement Learning, Decision Tree Learning, Bayesian networks, Support Vector Machine, Genetic Algorithm)
- Issues in Machine Learning
- Data Science Vs Machine Learning



DECODED BY

Chirag Jain

Founder, Ycotes · Gold Medalist in Artificial Intelligence, IIT Madras · taught 10,000+ students

Introduction to Machine Learning

MACHINE LEARNING

A field of computer science that gives computers the ability to learn without being explicitly programmed, by extracting patterns from empirical data.

- Coined by Arthur Samuel in 1959.
- Shifts focus from manual rule creation to automated pattern recognition.
- Crucial for domains where algorithmic solutions are complex or unknown.

What is Learning?

DEFINITION OF LEARNING (TOM MITCHELL)

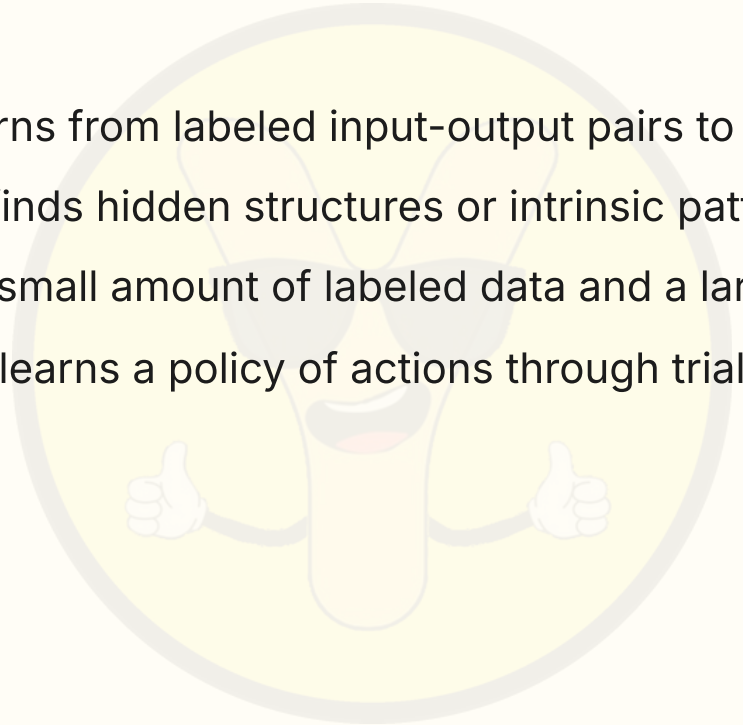
A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.

- Task (T): The application problem (e.g., classifying emails).
- Experience (E): The data or training examples provided (e.g., dataset of labeled emails).
- Performance (P): The quantitative metric to evaluate success (e.g., accuracy percentage).

Types of Machine Learning

Machine learning paradigms are categorized based on the nature of the learning signal and available feedback.

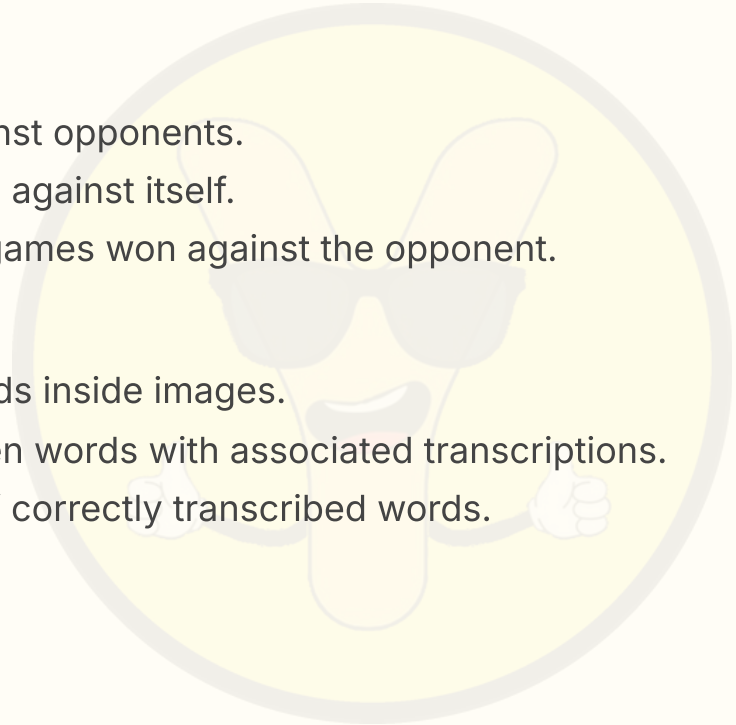
- Supervised Learning: Algorithm learns from labeled input-output pairs to map future inputs.
- Unsupervised Learning: Algorithm finds hidden structures or intrinsic patterns in unlabeled data.
- Semi-Supervised Learning: Uses a small amount of labeled data and a large amount of unlabeled data.
- Reinforcement Learning: Algorithm learns a policy of actions through trial and error to maximize a cumulative reward.



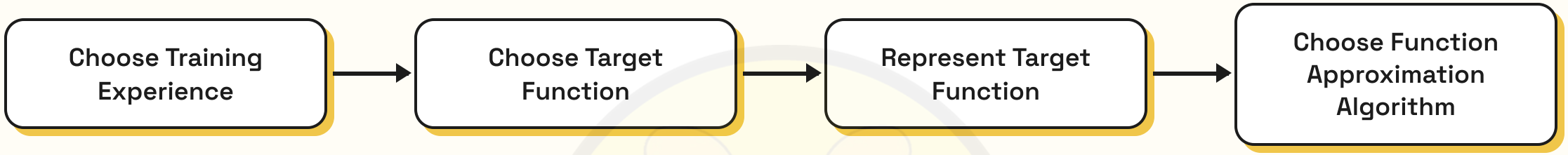
Well-Defined Learning Problems

To design a learning system properly, the task, experience, and performance must be precisely specified.

- Checkers Learning Problem
  - Task (T): Playing checkers games against opponents.
  - Experience (E): Playing practice games against itself.
  - Performance Measure (P): Percent of games won against the opponent.
- Handwriting Recognition
  - Task (T): Recognizing handwritten words inside images.
  - Experience (E): Database of handwritten words with associated transcriptions.
  - Performance Measure (P): Accuracy of correctly transcribed words.



Designing a Learning System



Step-by-step design process for constructing a machine learning system.

History of Machine Learning

The evolution of machine learning spans decades of theoretical breakthroughs and hardware advancements.

- 1950s: Alan Turing proposed the Turing Test; Arthur Samuel wrote the checkers program.
- 1960s-1970s: Development of Bayesian methods, nearest neighbor algorithms, and early neural networks (Perceptrons).
- 1980s: Revival with Decision Trees (ID3), Backpropagation algorithm, and symbolic learning.
- 1990s: Shift to probability-based methods, Support Vector Machines (SVM), and data-driven AI.
- 2000s-Present: Deep Learning boom, massive datasets, and ubiquitous commercial applications.

ML Approaches: Artificial Neural Networks & Clustering

Overview of foundational learning approaches used in predictive modeling and grouping.

- Artificial Neural Networks (ANN)
  - Inspired by biological nervous systems.
  - Composed of nodes (neurons) organized in input, hidden, and output layers.
  - Weights are adjusted iteratively using gradient descent and backpropagation.
- Clustering
  - An unsupervised learning approach to group similar data instances.
  - Minimizes intra-cluster distance and maximizes inter-cluster distance.
  - Examples: K-Means, Hierarchical Clustering.

ML Approaches: Reinforcement Learning & Decision Trees

Methods involving sequential decision making and hierarchical rule generation.

- Reinforcement Learning
  - An agent interacts with an environment by taking actions and receiving rewards or penalties.
  - Goal is to learn an optimal policy function $\pi : S \rightarrow A$.
  - Applications: Robotics, game playing (AlphaGo).
- Decision Tree Learning
  - Supervised approach representing a function as a flowchart-like tree structure.
  - Nodes represent attribute tests, branches represent outcomes, and leaves represent class labels.
  - Built using information gain or Gini impurity metrics.

ML Approaches: Bayesian Networks & SVM

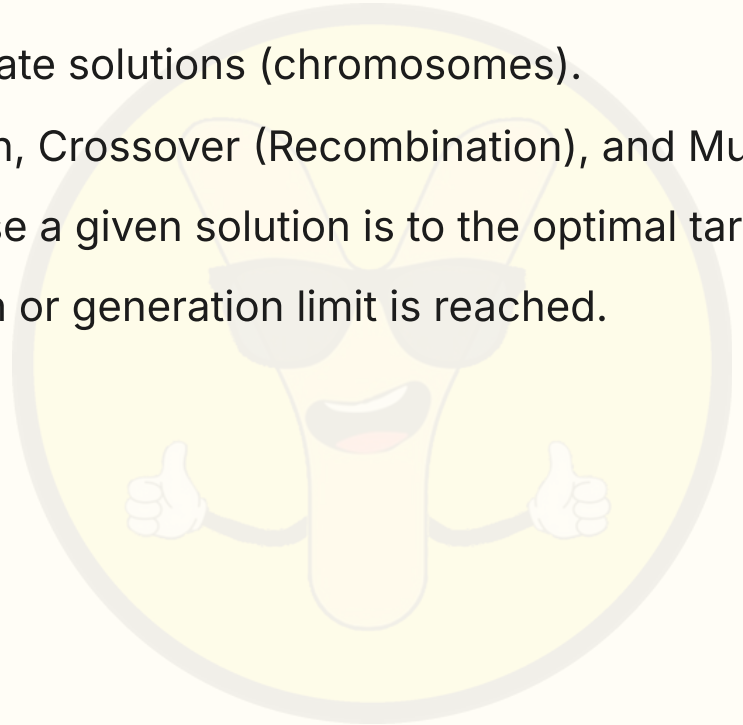
Probabilistic graphical models and geometric margin-maximizing classifiers.

- Bayesian Networks
  - Directed acyclic graph representing probabilistic relationships among variables.
  - Uses Bayes' theorem to compute conditional probabilities and handle uncertainty.
- Support Vector Machine (SVM)
  - Finds an optimal hyperplane in an N-dimensional space to separate classes with maximal margin.
  - Uses kernel tricks to handle non-linearly separable data.

ML Approaches: Genetic Algorithms

Search and optimization techniques inspired by natural selection and genetics.

- Operates on a population of candidate solutions (chromosomes).
- Applies genetic operators: Selection, Crossover (Recombination), and Mutation.
- Fitness function evaluates how close a given solution is to the optimal target.
- Iterates until a termination condition or generation limit is reached.



Issues in Machine Learning

Fundamental challenges faced by researchers and engineers when building practical ML models.

- Overfitting and Underfitting: High training accuracy vs generalizability on unseen test data.
- Curse of Dimensionality: Performance degradation as the number of input features grows excessively large.
- Data Quality: Presence of noise, missing values, outliers, and biased training samples.
- Computational Complexity: High training times and heavy hardware requirements for complex models.
- Interpretability: The 'black-box' nature of complex architectures like deep neural networks.

Data Science Vs Machine Learning

Feature	Data Science	Machine Learning
Definition	Broad interdisciplinary field focused on extracting insights from data.	Subset of AI focused on building algorithms that learn from data.
Scope	Includes data cleaning, visualization, statistical analysis, and ML.	Focused strictly on algorithmic prediction, pattern recognition, and modeling.
Primary Goal	To uncover hidden business trends, patterns, and actionable insights.	To enable systems to automatically learn and make accurate predictions.
Techniques	SQL, data wrangling, statistics, exploratory data analysis, visualization.	Supervised/unsupervised algorithms, neural networks, optimization.

Exam Corner

- **Q1** Define Machine Learning. Explain Tom Mitchell's definition of learning with a suitable example.
- **Q2** Differentiate between supervised, unsupervised, and reinforcement learning with real-world examples.
- **Q3** Explain the key design steps involved in building a machine learning system.
- **Q4** Discuss various approaches to machine learning, highlighting Artificial Neural Networks and Support Vector Machines.
- **Q5** What is a Genetic Algorithm? Explain its working steps and biological inspiration.
- **Q6** Discuss the major issues and challenges encountered in machine learning applications.
- **Q7** Differentiate clearly between Data Science and Machine Learning with appropriate parameters.

Unit 1 done. You are exam-ready.

Revise the Exam corner once more, then open Unit 2 on theycotes.com. Every AKTU subject, unit by unit, free. Notes by Chirag Jain, Ycotes.

