

Module 1

The logo of AIKTC (Asian Institute of Karsekar) is a circular emblem. It features a central illustration of a large, classical-style building with multiple domes and arches. The text around the building includes "ISLAM'S KALSEKAR ENGINEERING & TECHNOLOGY" on the left, "TECHNICAL CAMPUS, NEW PANVEL ARCHITECTURE PHARMACY" on the right, and "NAVI MUMBAI - INDIA" at the bottom. The acronym "AIKTC" is prominently displayed in the center of the emblem.

Introduction to Machine Learning

What is Machine Learning

- 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 .
- A branch of artificial intelligence, concerned with the design and development of algorithms that allow computers to evolve behaviours based on empirical data.

Why Learning

Why “Learn” ?

- Machine learning is programming computers to optimize a performance criterion using example data or past experience.
- There is no need to “learn” to calculate payroll
- Learning is used when:
 - Human expertise does not exist (navigating on Mars),
 - Humans are unable to explain their expertise (speech recognition)
 - Solution changes in time (routing on a computer network)
 - Solution needs to be adapted to particular cases (user biometrics)

What We Talk About When We Talk About “Learning”

- Learning general models from a data of particular examples
- Data is cheap and abundant (data warehouses, data marts); knowledge is expensive and scarce.
- Example in retail: Customer transactions to consumer behavior:
People who bought “Da Vinci Code” also bought “The Five People You Meet in Heaven” (www.amazon.com)
- Build a model that is *a good and useful approximation* to the data.

Types of Machine Learning

Supervised learning

Learning a mapping from a set of inputs to a target variable

Classification: target variable is discrete (e.g., spam email)

Regression: target variable is real-valued (e.g., stock market)

Unsupervised learning

No target variable provided

Clustering: grouping data into K groups

Other types of learning

Reinforcement learning: e.g., game-playing agent

Learning to rank, e.g., document ranking in Web search

And many others....

Dataset Example

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Applications of Machine Learning

Adaptive websites

Affective computing

Bioinformatics

Brain-machine interfaces

Classifying DNA sequences

Computational anatomy

Computer vision, including object recognition

Detecting credit card fraud

Game playing

Information retrieval

Internet fraud detection

Marketing

Machine perception

Medical diagnosis

Economics



Applications of Machine Learning

Natural language processing

Online advertising

Recommender systems

Robot locomotion

Search engines

Sentiment analysis (or opinion mining)

Sequence mining

Software engineering

Speech and handwriting recognition

Stock market analysis

Structural health monitoring

Syntactic pattern recognition

User behavior analytics



Steps in developing Machine Learning App

1. Collect data
2. Prepare the input data
3. Analyze the input data
4. Train the algorithm
5. Test the algorithm
6. Use it



Issues in Machine Learning

1. What algorithms can handle high-dimensional data well (and when) ?
2. How does number of features influence the accuracy?
3. How does class imbalance affect classification performance?
4. How does noise in data affect machine learning?
5. What are the challenges of online learning?
6. How can prior knowledge be incorporated into machine learning?
7. How can system alternatives be represented?



Issues in Machine Learning

- What algorithms are available for learning a concept? How well do they perform?
- How much training data is sufficient to learn a concept with high confidence?
- When is it useful to use prior knowledge?
- Are some training examples more useful than others?
- What are best tasks for a system to learn?
- What is the best way for a system to represent its knowledge?

Choosing the Right Algorithm

- How much data do you have & is it continuous?
- Is it classification or regression problem?
- Predefined variables (Labeled), unlabeled or mix?
- Data class skewed?
- What is the goal? – predict or rank?
- Result interpretation easy or hard?

[Link](https://www.datasciencecentral.com/profiles/blogs/want-to-know-how-to-choose-machine-learning-algorithm) <https://www.datasciencecentral.com/profiles/blogs/want-to-know-how-to-choose-machine-learning-algorithm>

[Link2](http://bigdata-madesimple.com/want-to-know-how-to-choose-a-machine-learning-algorithm/) <http://bigdata-madesimple.com/want-to-know-how-to-choose-a-machine-learning-algorithm/>