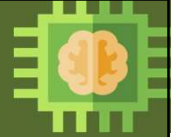


Elective Course

Course Code: CS4103

Autumn 2025-26



Lecture #33

Artificial Intelligence for Data Science

Week-9:

MACHINE LEARNING (Part I)

Introduction to the concept of learning

Course Instructor:

Dr. Monidipa Das

Assistant Professor

Department of Computational and Data Sciences

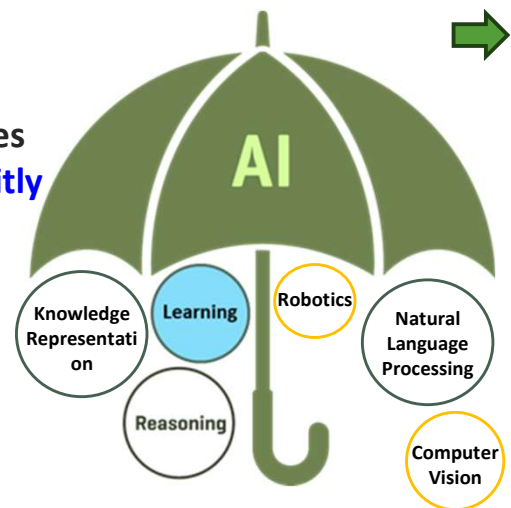
Indian Institute of Science Education and Research Kolkata, India 741246

What is Learning

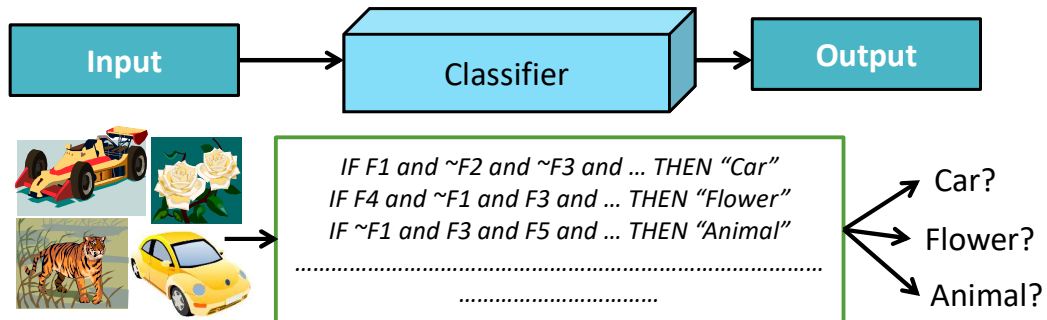


- A subfield of artificial intelligence (AI)
- Machine Learning: “the field of study that **gives computers the ability to learn without explicitly being programmed.**” [Arthur Samuel, 1950]

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*.
---[Tom M. Mitchell]



Explicit Programming



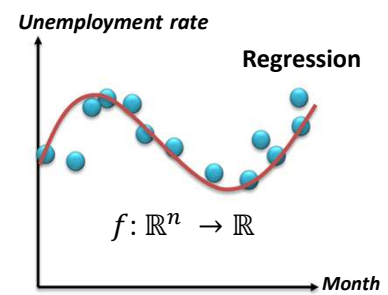
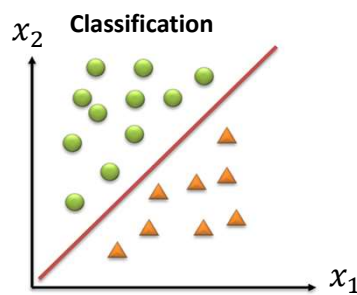
Dr. Monidipa Das, Department of CDS, IISER Kolkata

The Task, T



1. Classification
2. Regression
3. Transcription
4. Machine Translation
5. Density Estimation
6. Clustering

.....



	Regression	Classification
Outcome	Continuous	Class

Dr. Monidipa Das, Department of CDS, IISER Kolkata

The Performance Measure, P



- Need measure of performance specific to task T
E.g.
 - For **classification**:
 - **Accuracy**: Proportion of samples for which correct output is produced
 - For **regression**:
 - **Mean Squared Error (MSE)**
- Usually on data not seen before, a test set
 - 1. Separate the data into training set and test set
 - 2. Train the model with training set
 - 3. Measure the model's performance with test set

Dr. Monidipa Das, Department of CDS, IISER Kolkata

The Experience, E



- Most algorithms **experience a dataset**
 - A dataset is a collection of many examples (also called **data points**) [Example: Next to next Slide]
- **Supervised learning algorithms:**
 - Experiences a dataset associated with labels
 - Learns to predict the labels from the data
Training data: $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$
- **Unsupervised learning algorithms:**
 - Experiences a dataset containing many features
 - Learns useful properties of the structure of the dataset
Training data: $\{\mathbf{x}_i\}_{i=1}^N$
- **Reinforcement learning algorithms:**
 - Not just experience with a fixed dataset, but interact with an environment
 - Learns actions to maximize cumulative rewards

Size or length of the input x_i is commonly known as **data/input dimensionality** or **feature dimensionality**

Dr. Monidipa Das, Department of CDS, IISER Kolkata

Describing a data set



- Common way of describing a data set is with a design matrix
- Different *examples in each row*
- Each *column corresponds to a different feature*
- Iris dataset contains 150 examples with four features for each example

Anderson's Iris data (oldest set in stat/ML)



Sepal length ◀	Sepal width ◀	Petal length ◀	Petal width ◀	Species ◀
5.1	3.5	1.4	0.2	<i>I. setosa</i>
4.9	3.0	1.4	0.2	<i>I. setosa</i>
4.7	3.2	1.3	0.2	<i>I. setosa</i>
7.0	3.2	4.7	1.4	<i>I. versicolor</i>
6.4	3.2	4.5	1.5	<i>I. versicolor</i>
6.9	3.1	4.9	1.5	<i>I. versicolor</i>
6.3	3.3	6.0	2.5	<i>I. virginica</i>
5.8	2.7	5.1	1.9	<i>I. virginica</i>
7.1	3.0	5.9	2.1	<i>I. virginica</i>

$$x \in \mathbb{R}^{150 \times 4}$$

Dr. Monidipa Das, Department of CDS, IISER Kolkata

Types of Features and Types of Outputs



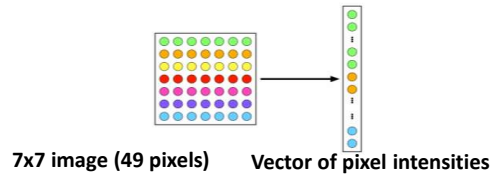
- **Real-valued:** Pixel intensity, house area, house price, rainfall amount, temperature, etc.
- **Binary:** Male/female, adult/non-adult, or any yes/no or present/absent, etc.
- **Categorical/Discrete:** Blood-group, color, etc.
- **Ordinal:** Grade (A/B/C etc.) in a course, or any other type where relative values matter

Dr. Monidipa Das, Department of CDS, IISER Kolkata

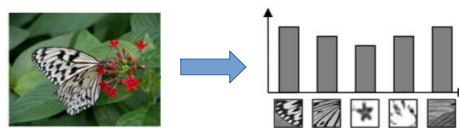
Example: Feature Extraction for Image Data



- A very simple feature extraction approach for image data is **flattening**



- Histogram** of visual patterns is another popular feature extraction method for images



Dr. Monidipa Das, Department of CDS, IISER Kolkata

Example: Feature Extraction for Text Data



- Consider some text data consisting of the following sentences:
 - John likes to watch movies
 - Mary likes movies too
 - John also likes football
- Want to construct a **feature representation** for these sentences
- Here is a **“bag-of-words”** (BoW) feature representation of these sentences

	John	likes	to	watch	movies	Mary	too	also	football
Sentence 1	1	1	1	1	1	0	0	0	0
Sentence 2	0	1	0	0	1	1	1	0	0
Sentence 3	1	1	0	0	0	0	0	1	1

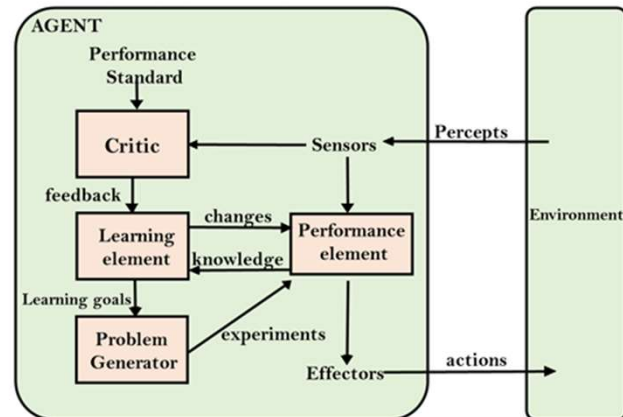
Dr. Monidipa Das, Department of CDS, IISER Kolkata

Learning as Learning of Functions



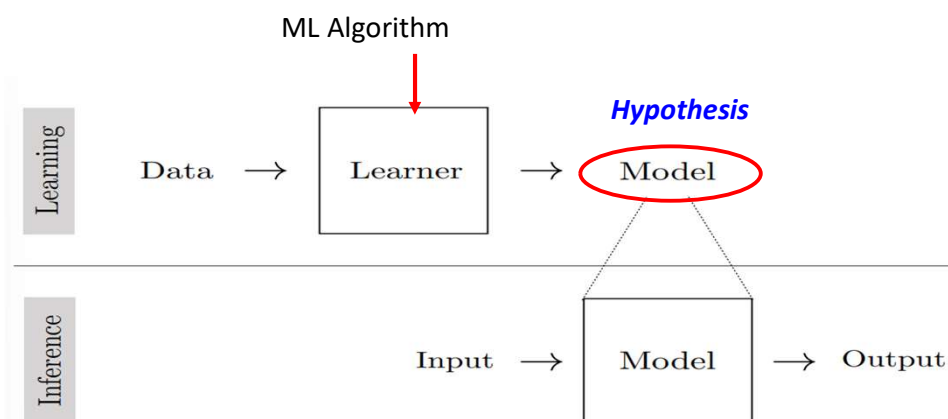
All learning can be seen as learning the representation of a function

- Numerical functions
 - Linear regression
 - **Neural networks**
 - **Support vector machines**
- Symbolic functions
 - **Decision trees**
 - Rules in PL, FOPL/FOL
- Instance-based functions
 - **Nearest-neighbor**
- Probabilistic Graphical Models
 - **Naïve Bayes**
 - Bayesian networks
 - Hidden-Markov Models (HMMs)
 - Markov networks



Dr. Monidipa Das, Department of CDS, IISER Kolkata

Machine Learning



Hypothesis: a function f that reads in low level properties (which are referred to as features) of a data point and delivers the prediction for the same. Usually denoted as h_θ

Hypothesis class: set of possible such functions

Dr. Monidipa Das, Department of CDS, IISER Kolkata

Inductive learning



- Simplest form: learn a function from examples

f is the target function

An example is a pair $(x, f(x))$

Problem: find a hypothesis h
such that $h \approx f$
given a training set of examples

Dr. Monidipa Das, Department of CDS, IISER Kolkata

Classification



- Computer program asked to specify which of k categories some input belongs to
 - Learning algorithm is asked to produce a function $f: \mathbb{R}^n \rightarrow \{1, \dots, k\}$, where n = no of input variables
 - When $y = f(x)$ model assigns input vector x to a category identified by a numeric code y
- Other variants of classification task:
 - f outputs a probability distribution over classes

Dr. Monidipa Das, Department of CDS, IISER Kolkata

Regression



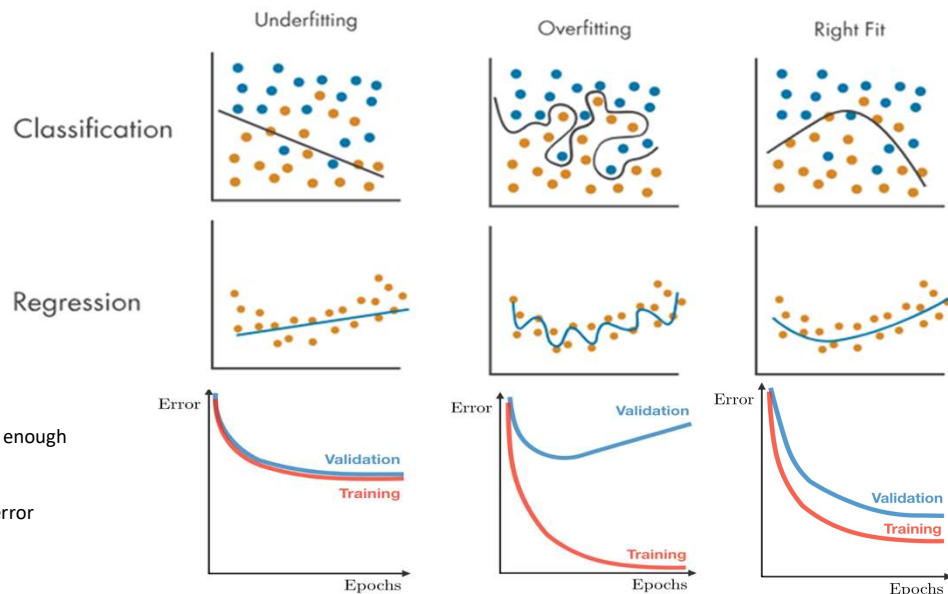
- Computer program required to predict a numerical value given some input
 - Algorithm to output $f: \mathbb{R}^n \rightarrow \mathbb{R}$
 - Task similar to classification except that format of output is different
- Ex: expected claim amount an insured person will make (used to set insurance premiums) or prediction of future prices of securities

Dr. Monidipa Das, Department of CDS, IISER Kolkata

Generalization, Underfitting and Overfitting



In ML, **generalization** is the ability to perform well on previously unobserved inputs



Underfitting: Inability to obtain low enough error rate on the training set

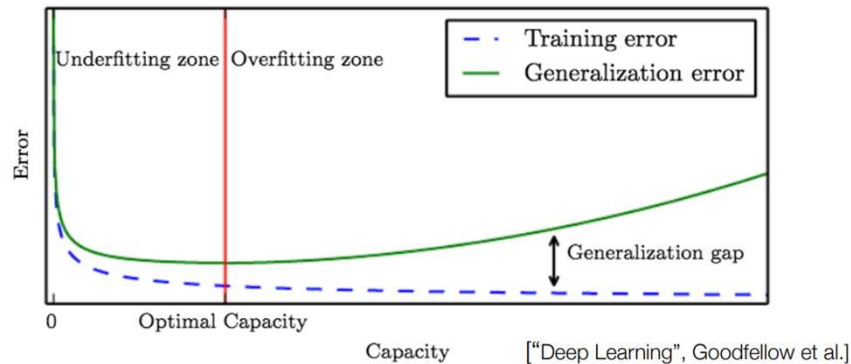
Overfitting: Gap between training error and testing error is too large

Dr. Monidipa Das, Department of CDS, IISER Kolkata

Generalization Error and Capacity



- **Generalization error definition:** Expected value of the error on a new input.
- Typically generalization error has a U-shaped curve
- **Model capacity** is ability to fit variety of functions



Dr. Monidipa Das, Department of CDS, IISER Kolkata



Questions?

Dr. Monidipa Das, Department of CDS, IISER Kolkata