Hidden Markov Model and Machine Learning

Dr. Syed Maaz Shahid

13th May 2024

Outline

- Markov Chain
- Hidden Markov Model
  - Working of HMM
  - Example scenarios
  - Applications of HMM
  - Limitations of HMM
- Machine Learning (ML) Preliminaries
  - Types of machine learning
- Federated Learning (FL)

Markov Chain

- A Markov chain is a discrete-time and discrete-valued random process in which each new sample is only dependent on the previous sample.

- Let $\{X_n\}_{n=0}^{N}$ be a sequence of random variables taking values in the countable set Ω.
- ***Def:*** X_n is a Markov chain if for all values of X_k and all n

$$P\{X_n = x_n | X_k = x_k \text{ for all } k < n\} = P\{X_n = x_n | X_{n-1} = x_{k-1}\}$$

Markov Chain

- A Markov chain tells something about the probabilities of sequences of random variables (states)

- The basic idea behind a Markov chain is to assume that X_k captures all *the relevant information* for predicting the future.

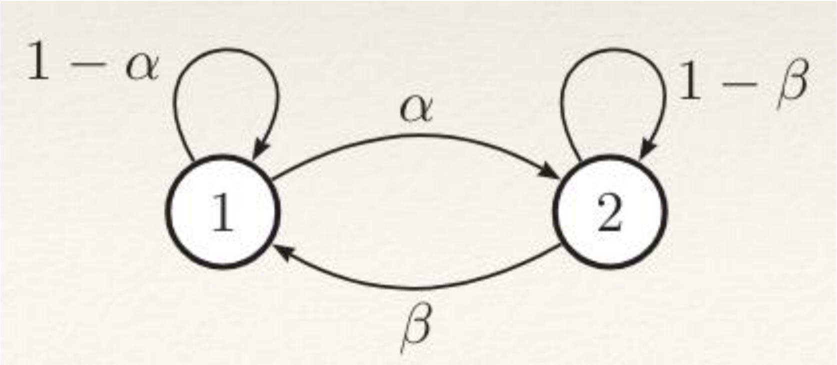


Fig: state transition diagram for Markov chain

Hidden Markov Model

- A Markov chain is useful when we need to compute a probability for a sequence of observable events.
  - What if the events we are interested in are hidden?

- A hidden Markov model (HMM) allows us to talk about both observed events and hidden events.

- For example: How do you know your wife is happy or not?
  - Determine from observable external factors

Hidden Markov Model

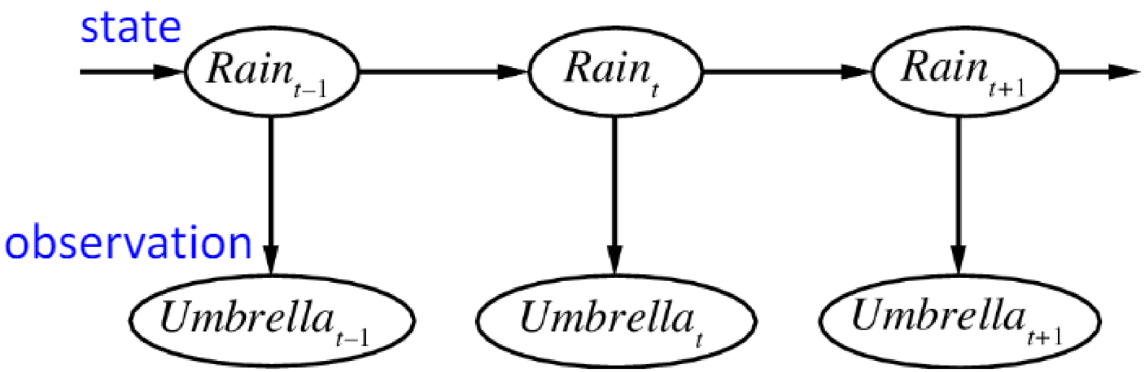
- An HMM is specified by the following components:
  - states
  - transition probability matrix
  - observation likelihoods (emission probabilities)
  - initial probability distribution over states
- How do we obtain the HMM?

Hidden Markov Model

- Example Scenario: Umbrella World (Scenario from chapter 15 of Russell & Norvig)
  - **Elspeth** Dunsany is an AI researcher.
  - **Richard** Feynman is an AI, its workstation is not connected to the internet.
  - He has noticed that Elspeth sometimes brings an **umbrella** to work.
  - He correctly infers that she is more likely to carry an umbrella on days when it **rains**.

Hidden Markov Model

- Richard proposes a hidden Markov model:
  - Rain on day $t-1, (R_{t-1})$, makes rain on day $t, (R_t)$, more likely.
  - Elspeth usually brings her umbrella (U_t) on days when it rains (R_t), but not always.



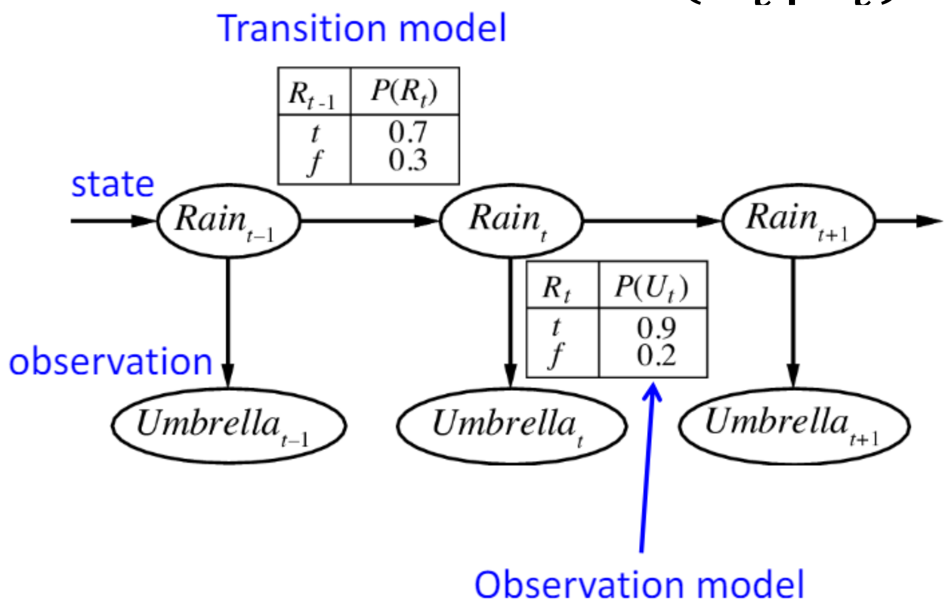
Hidden Markov Model

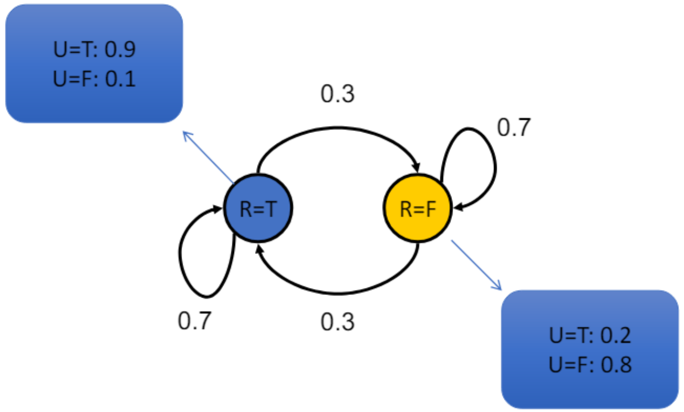
- Richard learns that the weather changes on 3 out of 10 days,

$$P(R_t|R_{t-1}) = 0.7, \qquad P(R_t|\sim R_{t-1}) = 0.3,$$

- Also, Elspeth sometimes forgets umbrella when it's raining, and sometimes brings an umbrella when it's not raining.

$$P(U_t|R_t) = 0.9, \qquad P(U_t|\sim R_t) = 0.1,$$





Transition probabilities

	R_t = T	R_t = F
R_{t-1} = T	0.7	0.3
R_{t-1} = F	0.3	0.7

Observation probabilities

	U_t = T	U_t = F
R_t = T	0.9	0.1
R_t = F	0.2	0.8

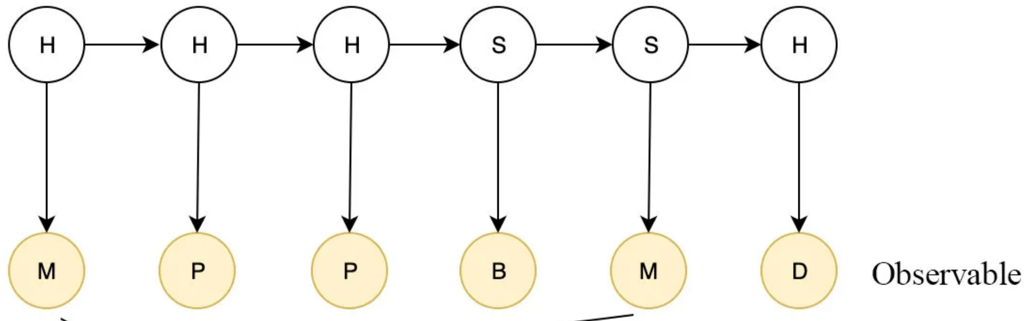
Hidden Markov Model

- The HMM is characterized by three **fundamental problems**
  - **Likelihood:** Given an HMM $\lambda = (A, B)$ (parameters) and observation sequence O, determine the likelihood probability of observed sequence $\left(P(O|\lambda)\right)$.
  - **Decoding:** Given observation sequence O and an HMM $\lambda = (A, B)$, discover the best hidden state sequence Q.
  - **Learning:** Given an observation sequence O and the set of states in the HMM, learn the HMM parameters A and B.

Hidden Markov Model

- Example scenario-2

internal state {H, S} is not observable or hard to determine



0.2 chance that I go to movie when I am happy.
0.4 chance that I go to movie when I am sad.

π

$P(x_1)$

$P(x_1 = happy) = 0.8$
$P(x_1 = sad\ ) = 0.2$

Initial state distribution
π

A

x_{t+1}

x_t	Happy	Sad
Happy	0.99	0.01
Sad	0.1	0.9

For example, $P(Happy_{t+1} | Happy_t) = 0.99$

Transition probability matrix ***A*** in Markov Process

B

$P(y_t | x_t)$:

	movie	book	party	dinning
Given being happy	0.2	0.2	0.4	0.2
Given being sad	0.4	0.3	0.1	0.2

Observation likelihoods or Emission probabilities ***B***

model λ

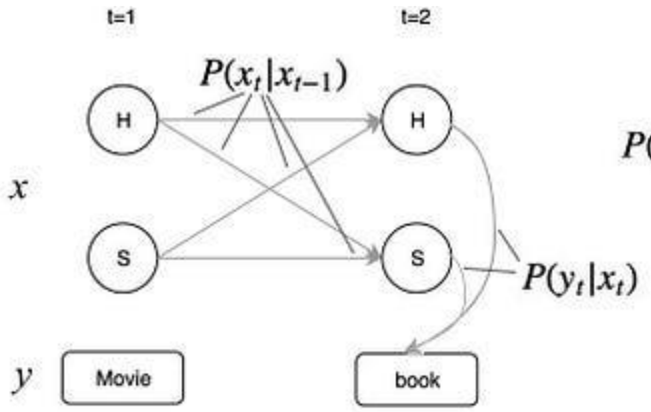
Hidden Markov Model

- The next state and the current observation solely depend on the current state only.

$$P(x_i \mid x_1, x_2, \ldots, x_{i-1}) = P(x_i \mid x_{i-1}) \qquad \text{(Markov process)}$$

$$O(o_i \mid x_1, x_2, \ldots, x_{i-1}) = P(o_i \mid x_i) \qquad \text{(Output independence)}$$

$$A + B = HMM\ model\ \lambda$$



from time 1 to t

$$P(x_t|y_{1:t}) = \frac{P(y_t|x_t)\, P(x_t|y_{1:t-1})}{\sum_{x_t} P(y_t|x_t)\, P(x_t|y_{1:t-1})}$$

$$\sum_{x_{t-1}} P(x_t|x_{t-1}) P(x_{t-1}|y_{1:t-1})$$

Hidden Markov Model

- **Likelihood** (likelihood of the observation)

calculated from emission probability calculated from transition probability

$$p(\mathrm{Y}) = \sum_{\mathrm{X}} p(\mathrm{Y}, \mathrm{X}) = \sum_{\mathrm{X}} p(\mathrm{Y} \mid \mathrm{X})\, p(\mathrm{X})$$

the observed events sum over all possible time sequences of internal states

Applications of Hidden Markov Model

- Speech Recognition
  - observations are acoustic signals, hidden states correspond to the different sounds
- Natural Language Processing
  - observations are the words in the text, hidden states are associated with the underlying grammar or structure of the text
- Bioinformatics
- Finance
  - observations are the stock prices, interest rates, or exchange rates, hidden states correspond to different economic states

Limitations of Hidden Markov Models

- Limited Modeling Capabilities
- Overfitting
- Lack of Robustness
- Computational Complexity

Assignment

- Find a paper that uses HMM to solve a problem in your relevant field.

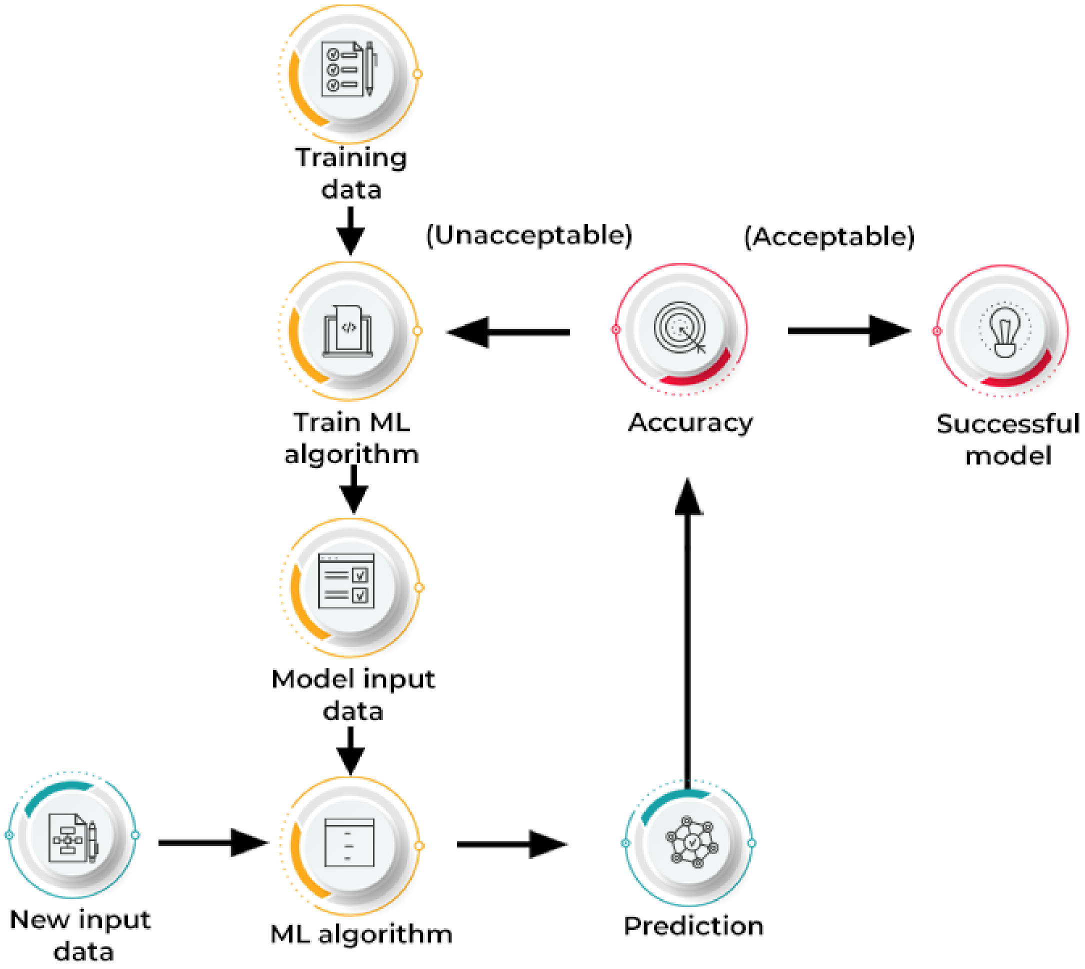
- Make a report and submit by 26 May, 20204.

What is Machine Learning

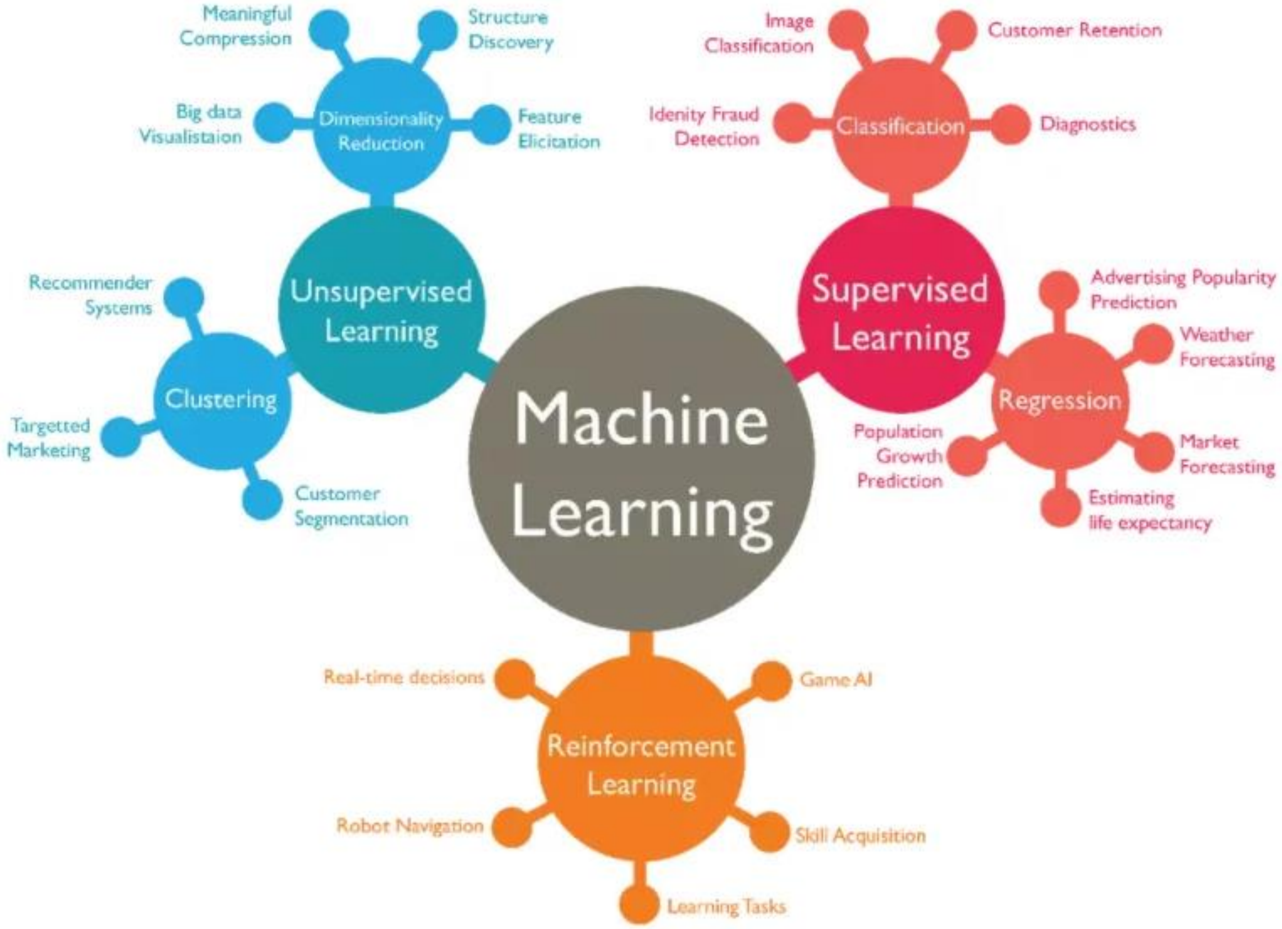
- Machine learning is a discipline of artificial intelligence (AI).
  - It provides machines with the ability to automatically learn from data and past experiences while identifying patterns to make predictions with minimal human intervention.

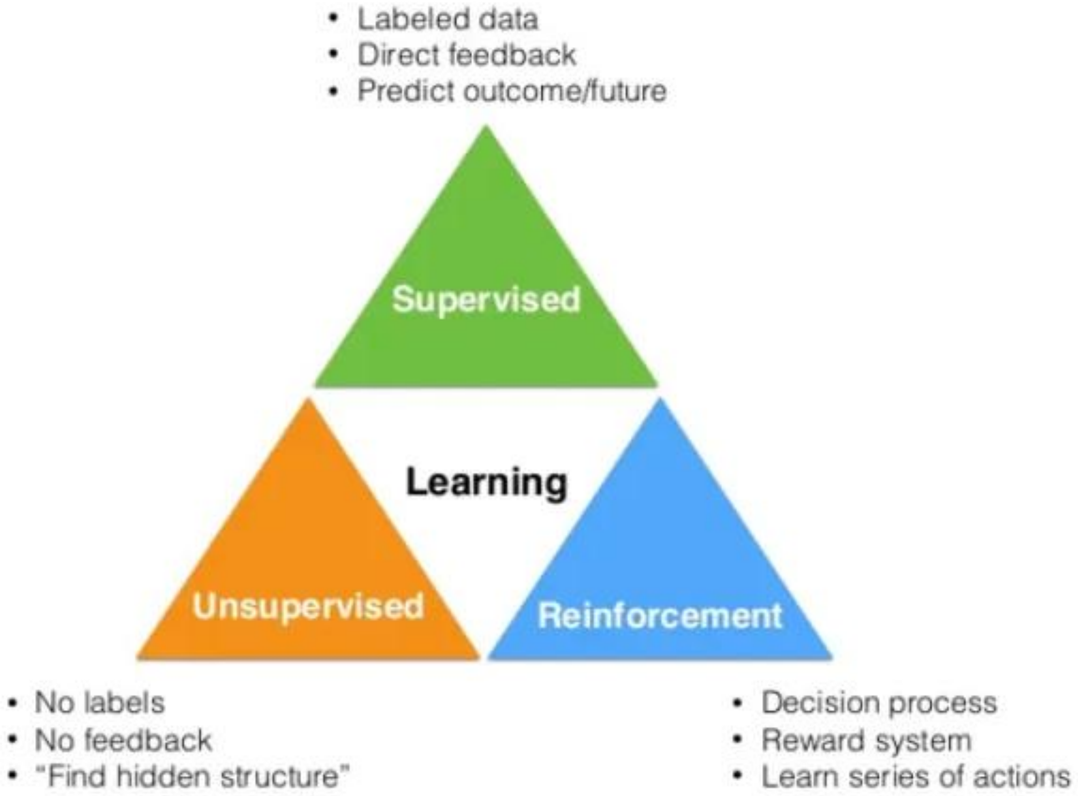
- Machine learning algorithms employ statistics to detect patterns in massive amounts of data.
  - Data could be anything: numbers, words, images, signals, or anything else.

How Does Machine Learning Work?

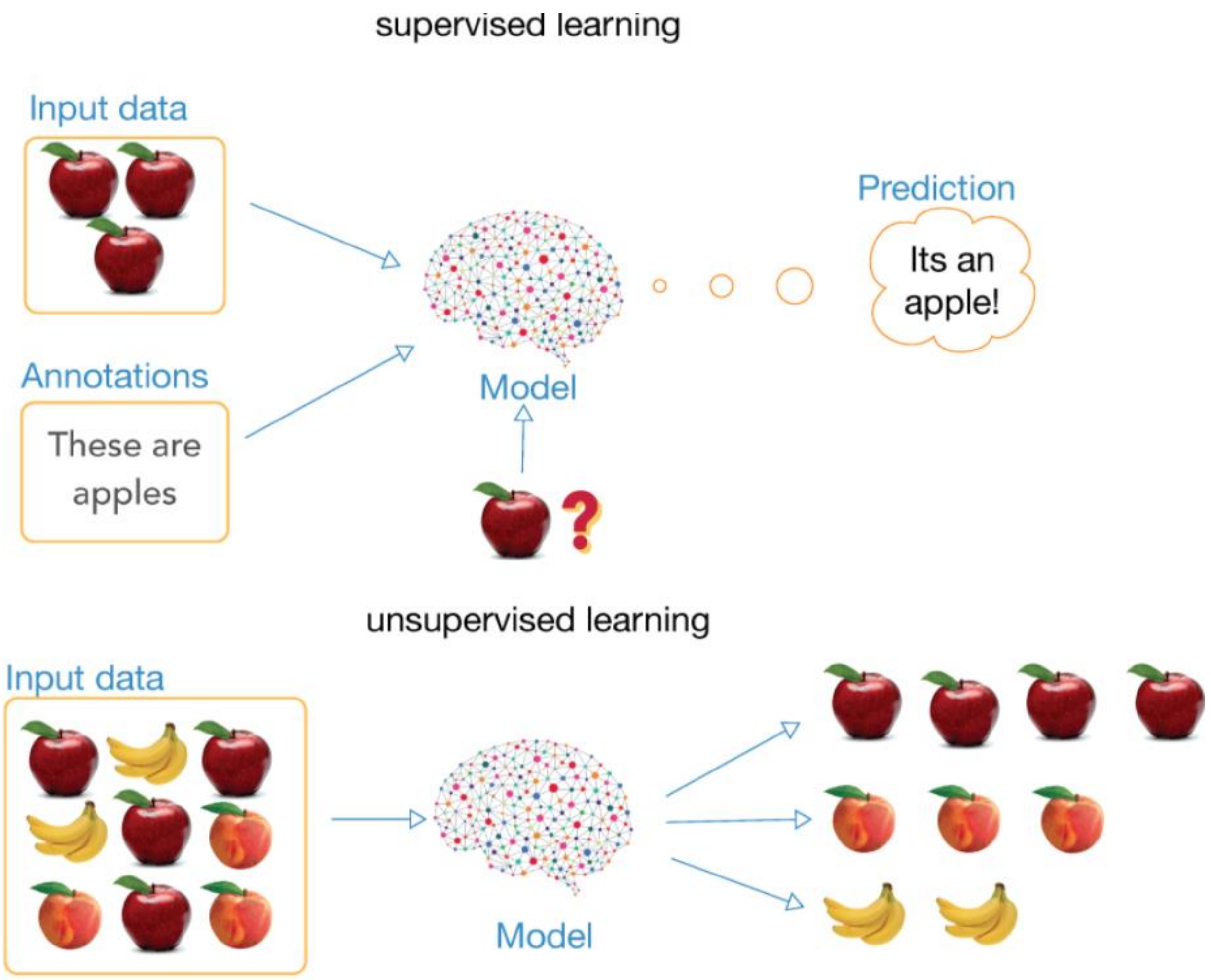


Categorization of Machine Learning



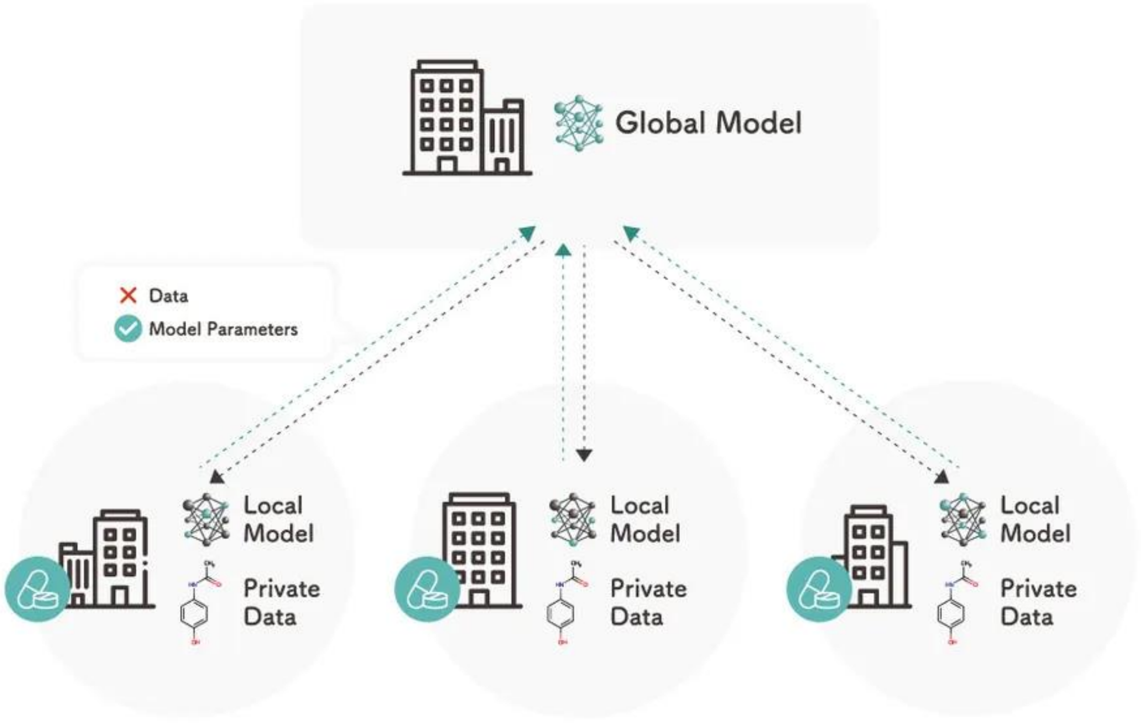


Supervised vs Unsupervised Learning

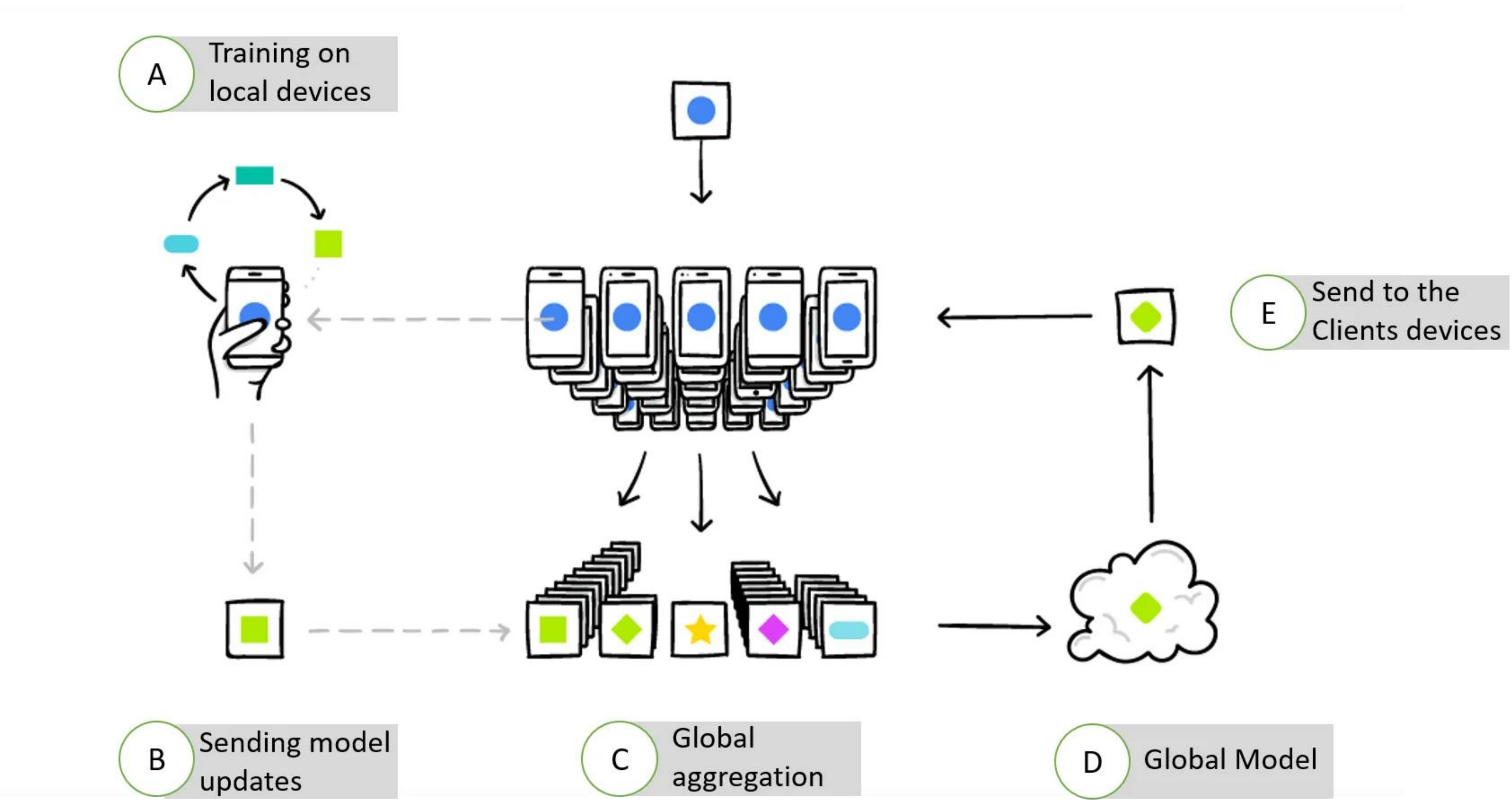


Federated Learning (FL)

- Federated Learning addresses the **challenges** of privacy, security, and data decentralization.



Federated Learning-Training Mechanism



Issues and Challenges in FL

- Communication Efficiency
- Heterogeneity of Clients
- Non-Independent and Identically Distributed (IID) Data

FL in 5G Networks

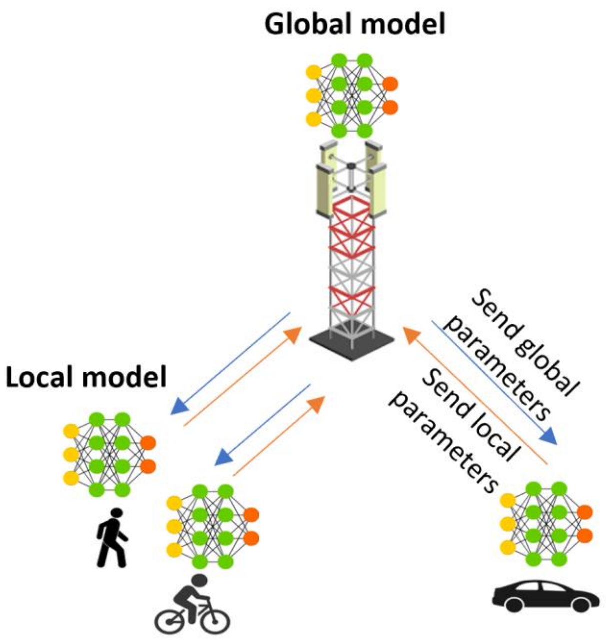


Fig: Federated learning

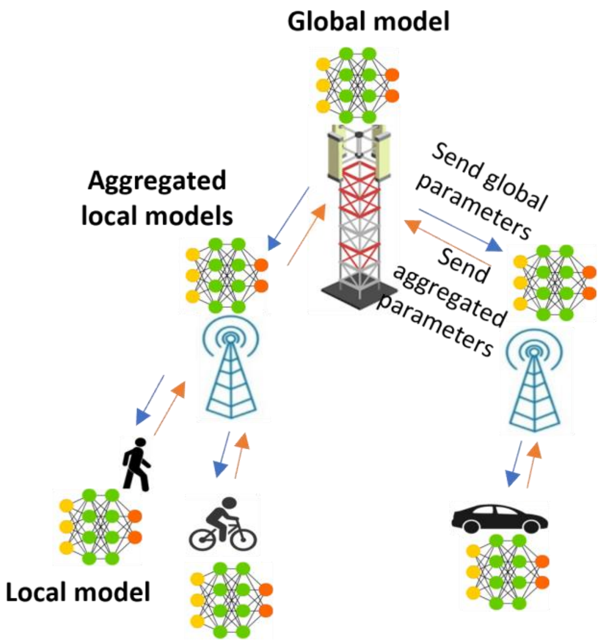


Fig: Hierarchical FL

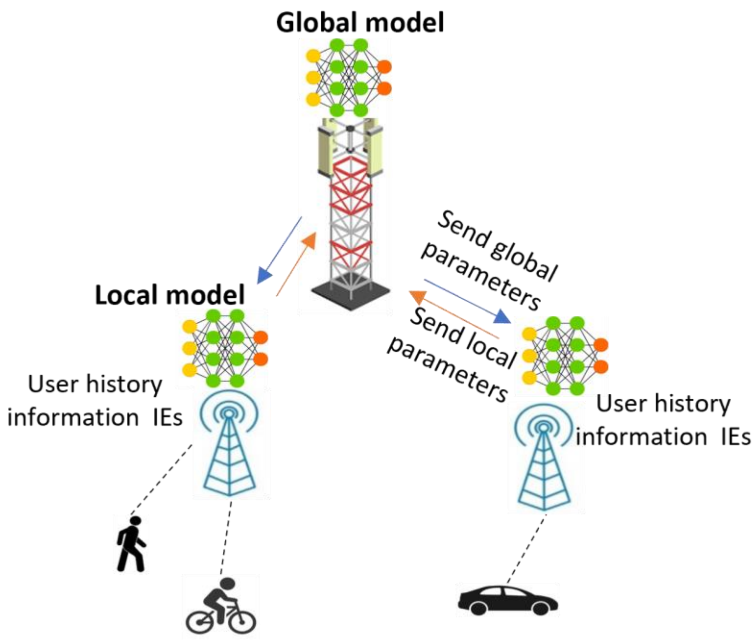


Fig: Proposed FL architecture in VTC- spring 2024