Semester: VIII Subject: Advanced AI Academic Year:2024-2025

Module 1

Probabilistic Models:

1. Gaussian Mixture Models (GMMs),
2. Hidden Markov Models (HMMs),
3. Bayesian Networks,
4. Markov Random Field (MRFs),
5. Probabilistic Graphical Models (PGM)

1. Probabilistic Graphical Models (PGM):

- Probabilistic Graphical models (PGMs) are statistical models that encode complex joint multivariate probability distributions using graphs.
- PGMs capture conditional independence relationships between interacting random variables.
- A probabilistic graphical model (PGM) is a probabilistic model for which a graph expresses the conditional dependence structure between random variables.
- There are two main branches of PGMs - namely, Bayesian networks and Markov random fields.
- To apply PGMs, we first model the problem as a Directed Acyclic Graph(DAG) Features are treated as random variables with probabilities. The nodes and dependencies represent the joint probability of the interaction of the variables in a hierarchy.
- Thus, the whole problem is modelled as a DAG with Bayesian probabilities.

- **Probabilistic Graphical models (PGMs)** are statistical models that encode complex joint multivariate probability distributions using graphs.
- PGMs capture conditional independence relationships between interacting random variables.
- This is beneficial since a lot of knowledge on graphs has been gathered over the years in various domains, especially on separating subsets, cliques and functions on graphs.
- This knowledge can be reused in PGMs.
- Furthermore, one can easily visualize PGMs and get a quick overview of the model structure.

APPLICATION OF PGM:

Knowing the graph structure of a PGM, one can solve tasks such as

-inference (computing the marginal distribution of one or more random variables) or learning (estimating the parameters of probability functions).
-One can even try to learn the structure of the graph itself, given some data.

2. GAUSSIAN MIXTURE MODEL (GMM):

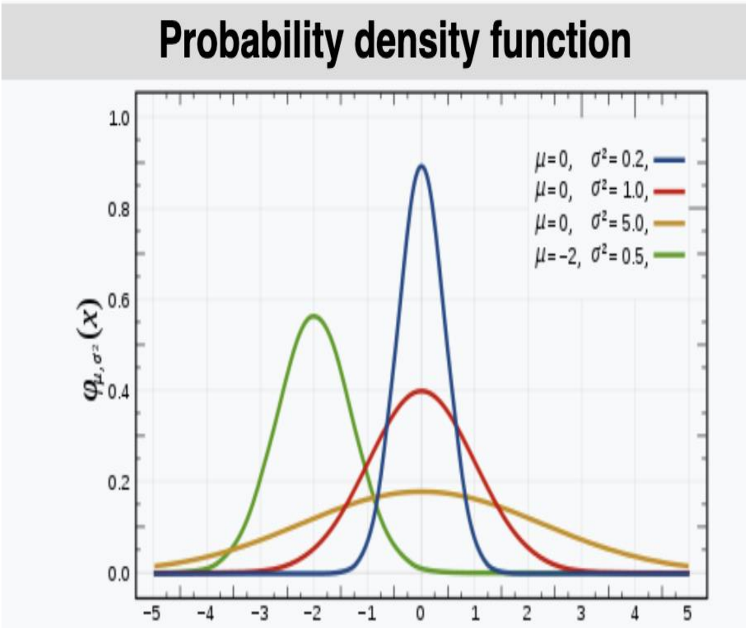
A Gaussian mixture model (GMM) is a machine learning method used to determine the probability each data point belongs to a given cluster. The model is a soft clustering method used in unsupervised learning.

A Gaussian mixture is a function that is composed of several Gaussians, each identified by $k \in \{1,\ldots, K\}$, where K is the number of clusters of our data set. Each Gaussian k in the mixture is comprised of the following parameters:

- A mean μ that defines its center.
- A covariance Σ that defines its width. This would be equivalent to the dimensions of an ellipsoid in a multivariate scenario.
- A mixing probability π that defines how big or small the Gaussian function will be.

GAUSSIAN DISTRIBUTION:

- Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean.
- In graphical form the normal distribution appears as a "bell curve"



In one dimension the probability density function of a Gaussian Distribution is given by

$$G(X|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

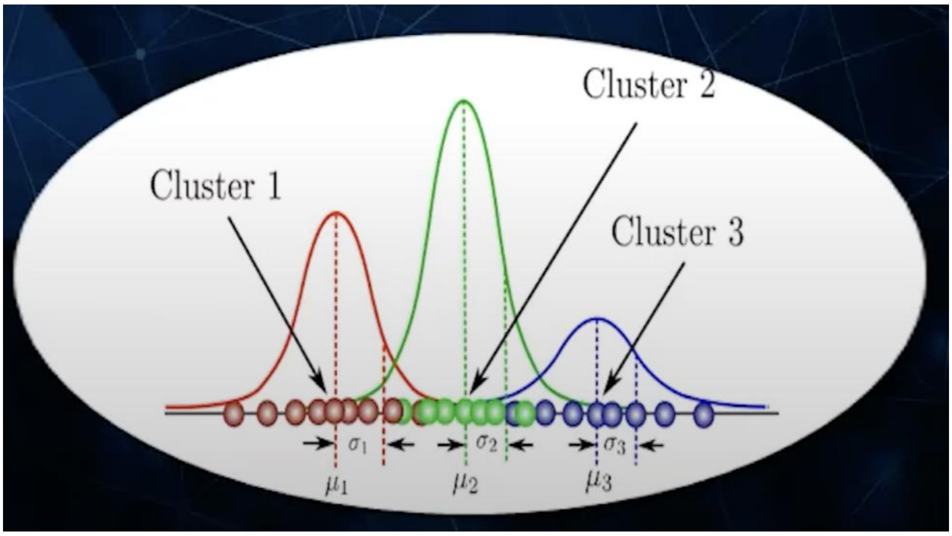
where μ and σ^2 are respectively the mean and variance of the distribution. For Multivariate (let us say d-variate) Gaussian Distribution, the probability density function is given by

$$G(X|\mu,\Sigma) = \frac{1}{\sqrt{(2\pi)|\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(X-\mu)^T \boldsymbol{\Sigma}^{-1}(X-\mu)\right)$$

Here μ is a d dimensional vector denoting the mean of the distribution and Σ is the d X d covariance matrix.

GAUSSIAN MIXTURE MODEL:

- A Gaussian Mixture Model (GMM) is a probabilistic model representing data as a mixture of multiple Gaussian distributions
- Each Gaussian distribution represents a component or cluster within the data



- The Gaussian mixture model (GMM) is a probabilistic model that assumes the data points come from a limited set of Gaussian distributions with uncertain variables. The mean and covariance matrix characterizes each individual Gaussian distribution.
- As an extension of the k-means clustering technique, a GMM takes into account the data's covariance structure and the likelihood of each point being derived from each Gaussian

distribution.

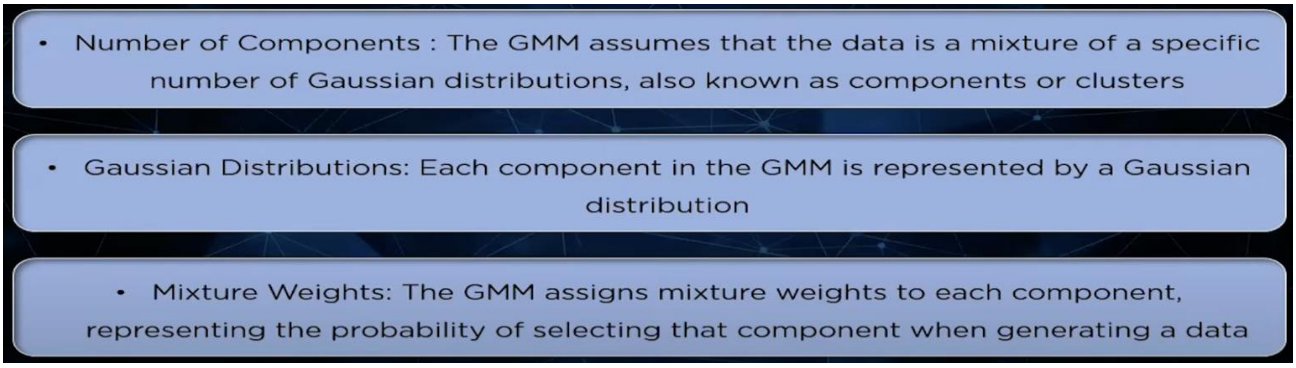
GMM ALGORITHM:

- Clustering algorithms like the Gaussian mixture models in machine learning are used to organize data by identifying commonalities and distinguishing them from one another. It may be used to classify consumers into subgroups defined by factors like demographics and buying habits.
- Each data point is given a chance of belonging to each cluster, making it a soft Gaussian mixture model clustering technique. This provides more leeway and may accommodate scenarios when data points do not naturally fall into one cluster.
- The GMM is trained using the EM algorithm, an iterative approach for determining the most likely estimations of the mixture Gaussian distribution parameters. The EM method first makes rough guesses at the parameters, then repeatedly improves those guesses until convergence is reached.

GMM WORKING:

- Initialize phase: Gaussian distributions' parameters should be initialized (means, covariances, and mixing coefficients).
- Expectation phase: Determine the likelihood that each data point was created using each of the Gaussian distributions.
- Maximization phase: Apply the probabilities found in the expectation step to re-estimate the Gaussian distribution parameters.
- Final phase: To achieve convergence of the parameters, repeat steps 2 and 3.

KEY COMPONENTS OF GMM:



3. HIDDEN MARKOV MODELS (HMM):

- A "Markov Model" process is basically one that does not have any memory -- the distribution of the next state/observation depends exclusively on the current state.
- A Markov chain is a simple Markov process, in which states can be observed directly.
- In a Hidden Markov Model, the states themselves are not directly observed, instead we assume that there are some sequences of states in a Markov chain that we cannot observe directly, and these states generate observations. Each state can have a different distribution of observations. Speech is often modeled using HMMs.
- **Hidden Markov Model** (HMM) is a statistical model that is used to describe the probabilistic relationship between a sequence of observations and a sequence of hidden states.
- It is often used in situations where the underlying system or process that generates the observations is unknown or hidden, hence it got the name "Hidden Markov Model."
- It is used to predict future observations or classify sequences, based on the underlying hidden process that generates the data.

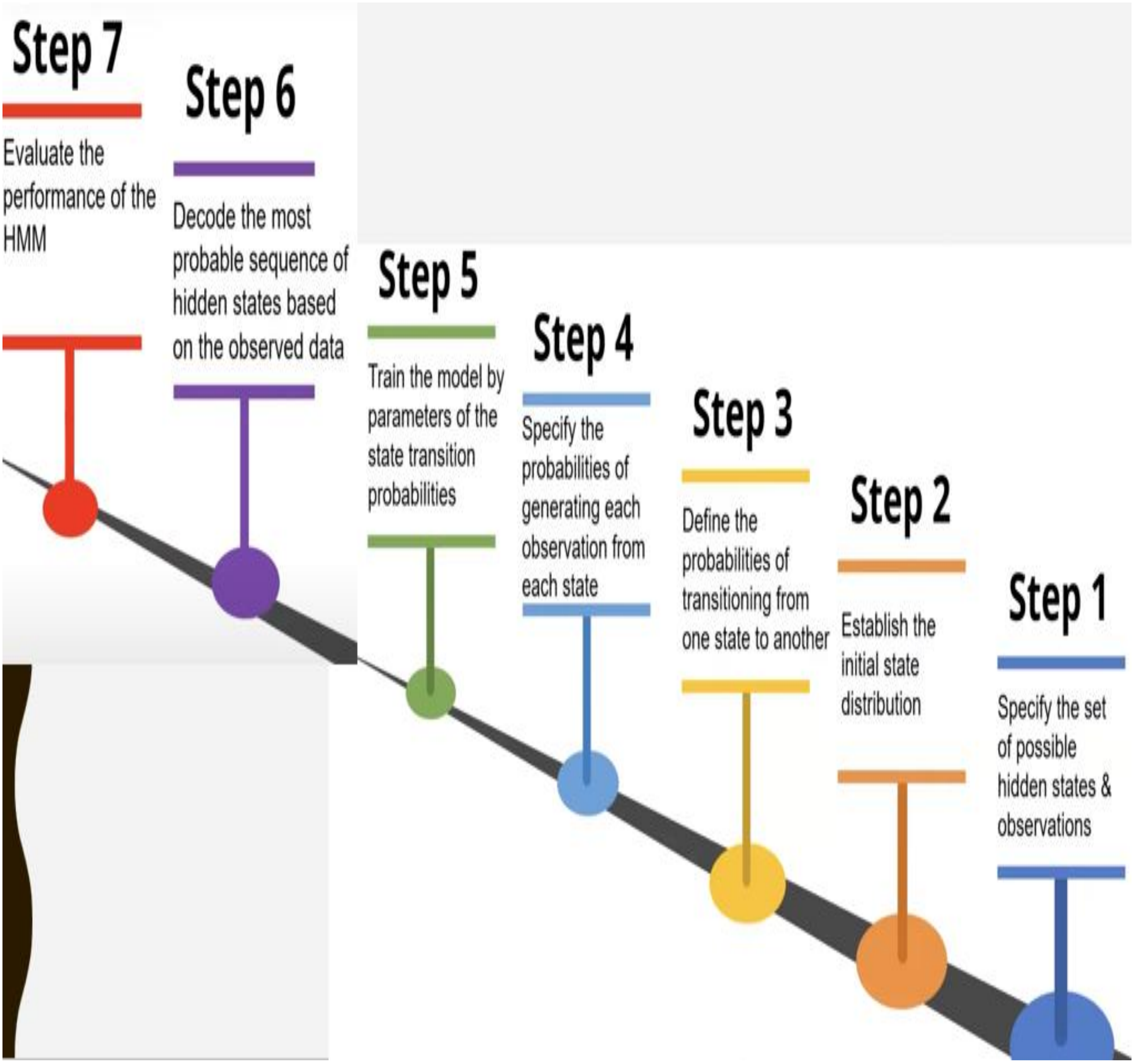
An HMM consists of two types of variables: hidden states and observations.

- The **hidden states** are the underlying variables that generate the observed data, but they are not directly observable.
- The **observations** are the variables that are measured and observed.

- The relationship between the hidden states and the observations is modeled using a probability distribution.
- The Hidden Markov Model (HMM) is the relationship between the hidden states and the observations using two sets of probabilities:
  the transition probabilities and
  the emission probabilities.

- The **transition probabilities** describe the probability of transitioning from one hidden state to another.
- The **emission probabilities** describe the probability of observing an output given a hidden state.

HMM ALGORITHM STEPS:



- The Hidden Markov Model (HMM) algorithm can be implemented using the following steps:
- **Step 1: Define the state space and observation space**
- The state space is the set of all possible hidden states, and the observation space is the set of all possible observations.
- **Step 2: Define the initial state distribution**
- This is the probability distribution over the initial state.
- **Step 3: Define the state transition probabilities**
- These are the probabilities of transitioning from one state to another. This forms the transition matrix, which describes the probability of moving from one state to another.
- **Step 4: Define the observation likelihoods:**
- These are the probabilities of generating each observation from each state. This forms the emission matrix, which describes the probability of generating each observation from each state.
- **Step 5: Train the model**
- The parameters of the state transition probabilities and the observation likelihoods are estimated using the Baum-Welch algorithm, or the forward-backward algorithm. This is done by iteratively updating the parameters until convergence.
- **Step 6: Decode the most likely sequence of hidden states**
- Given the observed data, the Viterbi algorithm is used to compute the most likely sequence of hidden states. This can be used to predict future observations, classify sequences, or detect patterns in sequential data.
- **Step 7: Evaluate the model**
- The performance of the HMM can be evaluated using various metrics,