Introduction to Machine Learning

- Prof. Balaraman Ravindran | IIT Madras

Problem Solving Session (Week-11)

Shreya Bansal

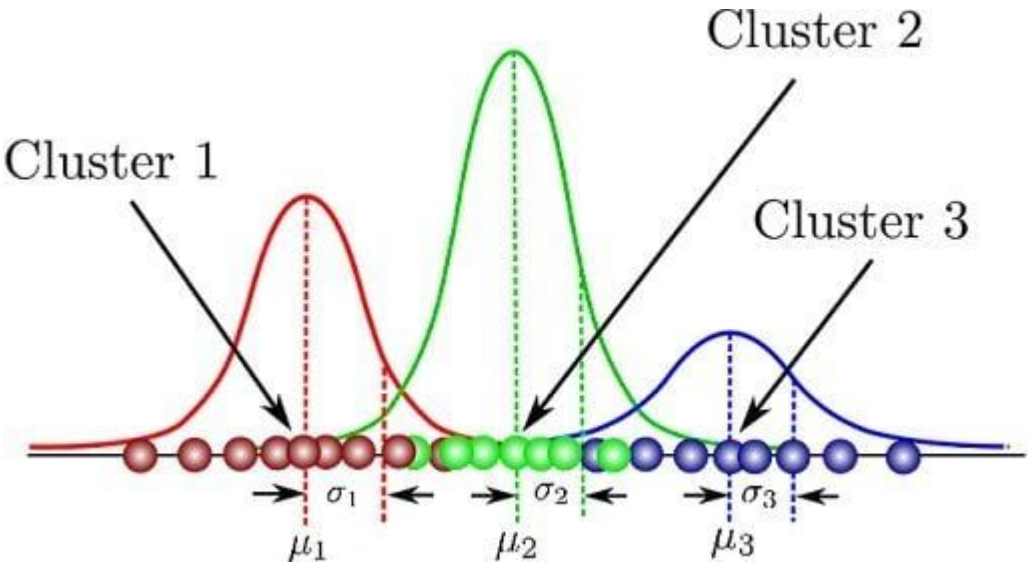
PMRF PhD Scholar
IIT Ropar

Week-11 Contents

---

1. Gaussian Mixture Model
2. Expectation Maximization

Introduction to Mixture Models



Mixture Models

- **Definition: A mixture model is a linear combination of probability distributions.**
- **Density Formula:**

$$p(x_n) = \sum_{k=1}^{K} \pi_k p(x_n|\theta_k)$$

  - π_k: Mixing weight ($\sum \pi_k = 1$).
  - $p(x|\theta_k)$: Component distribution (e.g., Gaussian).

Gaussian Mixture Models (GMMs)

- **Most common form: Components are Gaussian distributions.**
- **Gaussian Density:**

$$\mathcal{N}(x|\mu,\Sigma) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

- **Constraints:**
  - $0 \leq \pi_k \leq 1$
  - $\sum_{k=1}^{K} \pi_k = 1$

Gaussian Distribution Refresher

- **Univariate Gaussian:**

$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

- **Multivariate Gaussian (p-dimensional):**

$$\mathcal{N}(x|\mu,\Sigma) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

Introduction to Gaussian Mixture Models

- **A Gaussian Mixture Model (GMM) is a probabilistic model that assumes data is generated from a mixture of several Gaussian distributions with unknown parameters. It is widely used for:**
- **Clustering**
- **Density estimation**
- **Anomaly detection**

- **Key Features:**
- **Each Gaussian component has its own mean (μ_k) and covariance (Σ_k).**
- **Each component has a weight (π_k) representing its contribution to the mixture.**

Mathematical Formulation

A GMM with K components is defined as:

$$p(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k)$$

where:

- π_k = weight of the k-th Gaussian ($\sum_{k=1}^{K} \pi_k = 1$)
- $\mathcal{N}(x|\mu_k, \Sigma_k)$ = Gaussian PDF:

$$\mathcal{N}(x|\mu_k, \Sigma_k) = \frac{1}{(2\pi)^{d/2}|\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)\right)$$

Derivation of EM for GMM

E-Step Derivation

The responsibility $\gamma(z_{nk})$ is the posterior probability that x_n comes from the k-th Gaussian:

$$\gamma(z_{nk}) = p(z_k = 1|x_n) = \frac{p(x_n|z_k = 1)p(z_k = 1)}{\sum_j p(x_n|z_j = 1)p(z_j = 1)}$$

M-Step Derivation

Maximize the expected log-likelihood:

$$Q(\theta, \theta^{\text{old}}) = \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \left[\log \pi_k + \log \mathcal{N}(x_n|\mu_k, \Sigma_k)\right]$$

Taking derivatives w.r.t. μ_k, Σ_k, π_k and setting to zero gives the update rules.

Expectation-Maximization (EM) Algorithm

Since GMM parameters are unknown, we use the **EM algorithm** to estimate them iteratively.

EM Steps

1. **E-step (Expectation):**

   Compute the posterior probability (responsibility) of each data point belonging to each Gaussian:

$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n|\mu_j, \Sigma_j)}$$

2. **M-step (Maximization):**

   Update parameters (π_k, μ_k, Σ_k) using the responsibilities:

$$\mu_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk}) x_n}{\sum_{n=1}^{N} \gamma(z_{nk})}$$

$$\Sigma_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk})(x_n - \mu_k)(x_n - \mu_k)^T}{\sum_{n=1}^{N} \gamma(z_{nk})}$$

$$\pi_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk})}{N}$$

Introduction to EM

- EM is an iterative optimization algorithm used for maximum likelihood estimation when data involves latent (hidden) variables.
- Commonly used in:
- Gaussian Mixture Models (GMM)
- Hidden Markov Models (HMM)
- Missing data problems

Why EM is Needed

- Direct maximization of the log-likelihood is difficult due to hidden variables.
- For example, in GMM, the identity of the cluster (latent) for each data point is unknown.
- EM provides a framework to iteratively estimate the latent variables and optimize model parameters.

Basic Idea of EM

- Start with initial guesses for parameters.
- E-Step (Expectation): Compute expected value of the log-likelihood, given current parameters.
- M-Step (Maximization): Maximize the expected log-likelihood w.r.t. parameters.
- Repeat until convergence.

EM Notation

- Let: X: Observed data, Z: Latent variables, θ: Model parameters

- Log-likelihood:

$$\log p(X|\theta) = \log \sum_{Z} p(X,Z|\theta)$$

- This is intractable → EM maximizes a lower bound

E-Step (Expectation)

Compute the **posterior distribution** over latent variables using current parameters:

$$Q(\theta, \theta^{(t)}) = \mathbb{E}_{Z|X,\theta^{(t)}}[\log p(X, Z \mid \theta)]$$

This is the **expected complete-data log-likelihood**.

M-Step (Maximization)

Maximize the expected log-likelihood w.r.t. parameters:

$$\theta^{(t+1)} = \arg\max_{\theta} Q(\theta, \theta^{(t)})$$

This gives updated parameter estimates that improve the log-likelihood.

EM Algorithm Summary

Repeat until convergence:

1. **E-step:** Compute posterior $p(Z \mid X, \theta^{(t)})$
2. **M-step:** Maximize $\mathbb{E}_Z[\log p(X, Z \mid \theta)]$

EM guarantees that:

$$\log p(X \mid \theta^{(t+1)}) \geq \log p(X \mid \theta^{(t)})$$

EM - Intuition

- EM alternates between guessing the missing data (E-step) and optimizing parameters assuming the guess is correct (M-step).

- The E-step fills in the missing data.

- The M-step finds best parameters given the completed data.

Applications of EM

- Clustering (GMM)
- Missing data imputation
- Topic modeling (LDA)
- Image segmentation
- Speech recognition

Advantages and Limitations

- Advantages:
    - Handles incomplete data
    - Has theoretical guarantees
    - Simple to implement

- Limitations:
    - Can get stuck in local optima
    - Requires careful initialization
    - Convergence can be slow

Example- 1D

Problem Statement

Given the following 1D data points:

$X = \{1, 2, 6, 7\}$

Assume a **GMM with 2 components** ($K = 2$):

- Initial parameters:
  - Means: $\mu_1 = 1, \mu_2 = 6$
  - Variances: $\sigma_1^2 = 1, \sigma_2^2 = 1$
  - Mixing coefficients (priors): $\pi_1 = 0.5, \pi_2 = 0.5$

Task: Perform **one full iteration of the EM algorithm** to update the parameters.

Example- 1D

Step 1: E-Step (Compute Responsibilities)

The responsibility $\gamma(z_{ik})$ measures how much component k is responsible for data point x_i.

For each x_i, compute:

$$\gamma(z_{ik}) = \frac{\pi_k \cdot \mathcal{N}(x_i|\mu_k, \sigma_k^2)}{\sum_{j=1}^{2} \pi_j \cdot \mathcal{N}(x_i|\mu_j, \sigma_j^2)}$$

where $\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$.

Example- 1D

Compute for each data point:

1. $x_1 = 1$:
   - $\mathcal{N}(1|\mu_1 = 1, \sigma_1^2 = 1) = \frac{1}{\sqrt{2\pi}} e^{0} = 0.3989$
   - $\mathcal{N}(1|\mu_2 = 6, \sigma_2^2 = 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(1-6)^2}{2}} = 0.0015$
   - $\gamma(z_{11}) = \frac{0.5 \times 0.3989}{0.5 \times 0.3989 + 0.5 \times 0.0015} = \frac{0.1995}{0.2009} \approx 0.993$
   - $\gamma(z_{12}) = 1 - 0.993 = 0.007$
2. $x_2 = 2$:
   - $\mathcal{N}(2|\mu_1 = 1, \sigma_1^2 = 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(2-1)^2}{2}} = 0.2420$
   - $\mathcal{N}(2|\mu_2 = 6, \sigma_2^2 = 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(2-6)^2}{2}} = 0.0003$
   - $\gamma(z_{21}) = \frac{0.5 \times 0.2420}{0.5 \times 0.2420 + 0.5 \times 0.0003} \approx 0.998$
   - $\gamma(z_{22}) \approx 0.002$
3. $x_3 = 6$:
   - $\mathcal{N}(6|\mu_1 = 1, \sigma_1^2 = 1) = 0.0003$
   - $\mathcal{N}(6|\mu_2 = 6, \sigma_2^2 = 1) = 0.3989$
   - $\gamma(z_{31}) \approx 0.002$
   - $\gamma(z_{32}) \approx 0.998$
4. $x_4 = 7$:
   - $\mathcal{N}(7|\mu_1 = 1, \sigma_1^2 = 1) = 0.0001$
   - $\mathcal{N}(7|\mu_2 = 6, \sigma_2^2 = 1) = 0.2420$
   - $\gamma(z_{41}) \approx 0.0004$
   - $\gamma(z_{42}) \approx 0.9996$

Example- 1D

Responsibilities Summary:

x_i	$\gamma(z_{i1})$	$\gamma(z_{i2})$
1	0.993	0.007
2	0.998	0.002
6	0.002	0.998
7	0.0004	0.9996

Example- 1D

Step 2: M-Step (Update Parameters)

Now, update π_k, μ_k, σ_k^2 using the responsibilities.

1. Update Mixing Coefficients (Priors) π_k:

$$\pi_k = \frac{\sum_{i=1}^{N} \gamma(z_{ik})}{N}$$

- $\pi_1 = \frac{0.993+0.998+0.002+0.0004}{4} = \frac{1.9934}{4} \approx 0.498$
- $\pi_2 = \frac{0.007+0.002+0.998+0.9996}{4} = \frac{2.0066}{4} \approx 0.502$

2. Update Means μ_k:

$$\mu_k = \frac{\sum_{i=1}^{N} \gamma(z_{ik}) x_i}{\sum_{i=1}^{N} \gamma(z_{ik})}$$

- $\mu_1 = \frac{(0.993\times1)+(0.998\times2)+(0.002\times6)+(0.0004\times7)}{1.9934} \approx \frac{2.987}{1.9934} \approx 1.498$
- $\mu_2 = \frac{(0.007\times1)+(0.002\times2)+(0.998\times6)+(0.9996\times7)}{2.0066} \approx \frac{12.98}{2.0066} \approx 6.47$

3. Update Variances σ_k^2:

$$\sigma_k^2 = \frac{\sum_{i=1}^{N} \gamma(z_{ik})(x_i - \mu_k)^2}{\sum_{i=1}^{N} \gamma(z_{ik})}$$

- $\sigma_1^2 = \frac{0.993(1-1.498)^2+0.998(2-1.498)^2+0.002(6-1.498)^2+0.0004(7-1.498)^2}{1.9934} \approx 0.25$
- $\sigma_2^2 = \frac{0.007(1-6.47)^2+0.002(2-6.47)^2+0.998(6-6.47)^2+0.9996(7-6.47)^2}{2.0066} \approx 0.25$

Example- 1D

Final Updated Parameters After 1 EM Iteration:

Parameter	Component 1 ($k = 1$)	Component 2 ($k = 2$)
π_k	0.498	0.502
μ_k	1.498	6.47
σ_k^2	0.25	0.25

Example- 1D

Interpretation

- The means shifted slightly towards the data points they are most responsible for.
- The variances decreased because the model is becoming more confident in cluster assignments.
- The priors π_k adjusted slightly based on cluster responsibilities.

This process is repeated until convergence (when parameters change negligibly).

Example- 3D

— Problem Statement

Given the following **3D data points**:

$$X = \left\{ \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 5 \\ 4 \\ 6 \end{bmatrix}, \begin{bmatrix} 6 \\ 5 \\ 5 \end{bmatrix} \right\}$$

Assume:

- **Initial means**:

$$\mu_1 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \quad \mu_2 = \begin{bmatrix} 5 \\ 4 \\ 6 \end{bmatrix}$$

- **Initial covariances** (isotropic):

$$\Sigma_1 = \Sigma_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- **Initial mixing coefficients (priors)**:

$$\pi_1 = 0.5, \quad \pi_2 = 0.5$$

Task: Perform **one full EM iteration** to update the parameters.

Example- 3D

Step 1: E-Step (Compute Responsibilities)

The **responsibility** $\gamma(z_{ik})$ measures how likely point x_i belongs to cluster k:

$$\gamma(z_{ik}) = \frac{\pi_k \mathcal{N}(x_i|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_i|\mu_j, \Sigma_j)}$$

where the **multivariate Gaussian PDF** is:

$$\mathcal{N}(x|\mu, \Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$$

Example- 3D

Compute for each point:

1. **For** $x_1 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$:
   - $\mathcal{N}(x_1|\mu_1, \Sigma_1) = 0.0635$ (high, since $x_1 = \mu_1$)
   - $\mathcal{N}(x_1|\mu_2, \Sigma_2) \approx 0.0000$ (very low)
   - $\gamma(z_{11}) = \frac{0.5 \times 0.0635}{0.5 \times 0.0635 + 0.5 \times 0.0000} \approx 1.0$
   - $\gamma(z_{12}) \approx 0.0$
2. **For** $x_2 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$:
   - $\mathcal{N}(x_2|\mu_1, \Sigma_1) \approx 0.0299$
   - $\mathcal{N}(x_2|\mu_2, \Sigma_2) \approx 0.0000$
   - $\gamma(z_{21}) \approx 1.0$
   - $\gamma(z_{22}) \approx 0.0$
3. **For** $x_3 = \begin{bmatrix} 5 \\ 4 \\ 6 \end{bmatrix}$:
   - $\mathcal{N}(x_3|\mu_1, \Sigma_1) \approx 0.0000$
   - $\mathcal{N}(x_3|\mu_2, \Sigma_2) = 0.0635$ (since $x_3 = \mu_2$)
   - $\gamma(z_{31}) \approx 0.0$
   - $\gamma(z_{32}) \approx 1.0$
4. **For** $x_4 = \begin{bmatrix} 6 \\ 5 \\ 5 \end{bmatrix}$:
   - $\mathcal{N}(x_4|\mu_1, \Sigma_1) \approx 0.0000$
   - $\mathcal{N}(x_4|\mu_2, \Sigma_2) \approx 0.0299$
   - $\gamma(z_{41}) \approx 0.0$
   - $\gamma(z_{42}) \approx 1.0$

Example- 3D

Responsibilities Summary:

x_i	$\gamma(z_{i1})$	$\gamma(z_{i2})$
$[1, 0, 2]^T$	**1.0**	**0.0**
$[2, 1, 1]^T$	**1.0**	**0.0**
$[5, 4, 6]^T$	**0.0**	**1.0**
$[6, 5, 5]^T$	**0.0**	**1.0**

Example- 3D

Step 2: M-Step (Update Parameters)

Now, update π_k, μ_k, Σ_k using the responsibilities.

1. Update Mixing Coefficients (Priors) π_k:

$$\pi_k = \frac{\sum_{i=1}^{N} \gamma(z_{ik})}{N}$$

- $\pi_1 = \frac{1.0+1.0+0.0+0.0}{4} = 0.5$
- $\pi_2 = \frac{0.0+0.0+1.0+1.0}{4} = 0.5$

2. Update Means μ_k:

$$\mu_k = \frac{\sum_{i=1}^{N} \gamma(z_{ik}) x_i}{\sum_{i=1}^{N} \gamma(z_{ik})}$$

- $\mu_1 = \frac{1.0\times[1,0,2]^T + 1.0\times[2,1,1]^T}{2} = \frac{[3,1,3]^T}{2} = [1.5, 0.5, 1.5]^T$
- $\mu_2 = \frac{1.0\times[5,4,6]^T + 1.0\times[6,5,5]^T}{2} = \frac{[11,9,11]^T}{2} = [5.5, 4.5, 5.5]^T$

Example- 3D

- **3. Update Covariance Matrices Σ_k:**

$$\Sigma_k = \frac{\sum_{i=1}^{N} \gamma(z_{ik})(x_i - \mu_k)(x_i - \mu_k)^T}{\sum_{i=1}^{N} \gamma(z_{ik})}$$

  - **For Σ_1:**

$$\Sigma_1 = \frac{1.0([1,0,2]^T - [1.5,0.5,1.5]^T)([1,0,2]^T - [1.5,0.5,1.5]^T)^T + 1.0([2,1,1]^T - [1.5,0.}{2}$$

$$= \frac{1.0\begin{bmatrix} 0.25 & 0.25 & -0.25 \\ 0.25 & 0.25 & -0.25 \\ -0.25 & -0.25 & 0.25 \end{bmatrix} + 1.0\begin{bmatrix} 0.25 & -0.25 & -0.25 \\ -0.25 & 0.25 & 0.25 \\ -0.25 & 0.25 & 0.25 \end{bmatrix}}{2} = \begin{bmatrix} 0.25 & 0.0 & -0. \\ 0.0 & 0.25 & 0.(\\ -0.25 & 0.0 & 0.2 \end{bmatrix}$$

  - **For Σ_2:**

$$\Sigma_2 = \frac{1.0([5,4,6]^T - [5.5,4.5,5.5]^T)([5,4,6]^T - [5.5,4.5,5.5]^T)^T + 1.0([6,5,5]^T - [5.5,4.}{2}$$

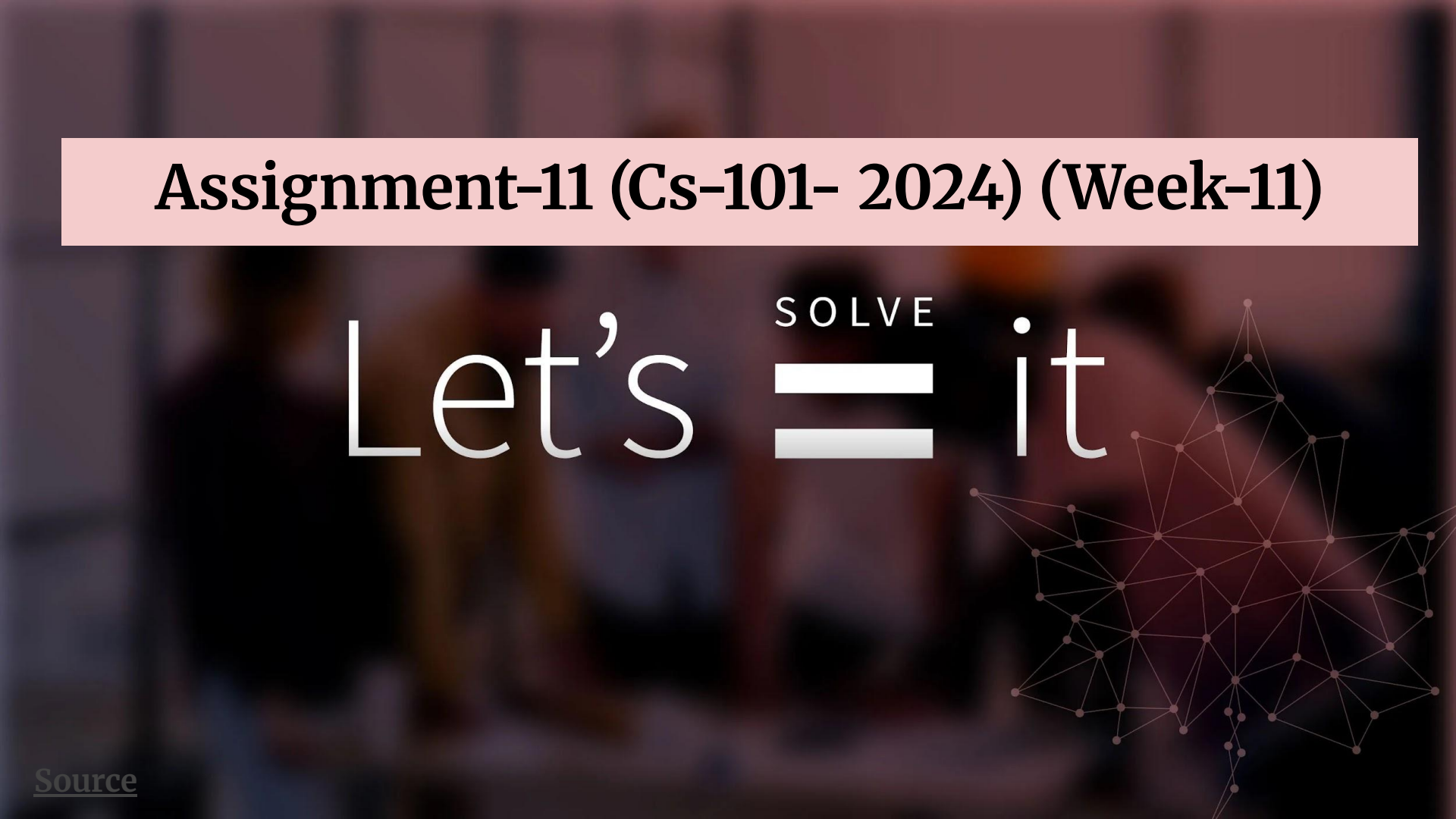
$$= \begin{bmatrix} 0.25 & 0.25 & -0.25 \\ 0.25 & 0.25 & -0.25 \\ -0.25 & -0.25 & 0.25 \end{bmatrix}$$

Example- 3D

Final Updated Parameters After 1 EM Iteration:

Parameter	Cluster 1 ($k = 1$)	Cluster 2 ($k = 2$)
π_k	0.5	0.5
μ_k	$[1.5, 0.5, 1.5]^T$	$[5.5, 4.5, 5.5]^T$
Σ_k	$\begin{bmatrix} 0.25 & 0.0 & -0.25 \\ 0.0 & 0.25 & 0.0 \\ -0.25 & 0.0 & 0.25 \end{bmatrix}$	$\begin{bmatrix} 0.25 & 0.25 & -0.25 \\ 0.25 & 0.25 & -0.25 \\ -0.25 & -0.25 & 0.25 \end{bmatrix}$

Assignment-11 (Cs-101- 2024) (Week-11)



Source

Question-1

01:00

What constraint must be satisfied by the mixing coefficients (π_k) in a GMM?

a) $\pi_k > 0 \ \forall \ k$

b) $\sum_k \pi_k = 1$

c) $\pi_k < 1 \ \forall \ k$

d) $\sum_k \pi_k = 0$

Question-1- Correct answer

What constraint must be satisfied by the mixing coefficients (π_k) in a GMM?

a) $\pi_k > 0 \ \forall \ k$

b) $\Sigma_k \pi_k = 1$

c) $\pi_k < 1 \ \forall \ k$

d) $\Sigma_k \pi_k = 0$

Correct options: (b) - Summation of priors probability=1

Question-2

01:00

The EM algorithm is guaranteed to decrease the value of its objective function on any iteration

a) True

b) False

Question-2- Correct answer

The EM algorithm is guaranteed to decrease the value of its objective function on any iteration

a) True

b) False

Correct options: (b).

Question-3

03:00

Why might the EM algorithm for GMMs converge to a local maximum rather than the global maximum of the likelihood function?

a) The algorithm is not guaranteed to increase the likelihood at each iteration
b) The likelihood function is non-convex
c) The responsibilities are incorrectly calculated
d) The number of components K is too small

Question-3 - Correct answer

Why might the EM algorithm for GMMs converge to a local maximum rather than the global maximum of the likelihood function?

a) The algorithm is not guaranteed to increase the likelihood at each iteration
b) The likelihood function is non-convex
c) The responsibilities are incorrectly calculated
d) The number of components K is too small

Correct options: (b)

Question-4

03:00

What does soft clustering mean in GMMs?

a) There may be samples that are outside of any cluster boundary.
b) The updates during maximum likelihood are taken in small steps, to guarantee convergence.
c) It restricts the underlying distribution to be gaussian.
d) Samples are assigned probabilities of belonging to a cluster

Question-4 - Correct answer

What does soft clustering mean in GMMs?

a) There may be samples that are outside of any cluster boundary.
b) The updates during maximum likelihood are taken in small steps, to guarantee convergence.
c) It restricts the underlying distribution to be gaussian.
d) Samples are assigned probabilities of belonging to a cluster

Correct options: (d)

Question-5

01:00

KNN is a special case of GMM with the following properties: (Multiple Correct)

(a) $\gamma_i = \frac{i}{(2\pi\epsilon)^{1/2}} e^{-\frac{1}{2\epsilon}}$

(b) Covariance $= \epsilon\mathbb{I}$

(c) $\mu_i = \mu_j \forall i, j$

(d) $\pi_k = \frac{1}{k}$

Question-5 - Correct answer

KNN is a special case of GMM with the following properties: (Multiple Correct)

(a) $\gamma_i = \frac{i}{(2\pi\epsilon)^{1/2}} e^{-\frac{1}{2\epsilon}}$

(b) Covariance $= \epsilon\mathbb{I}$

(c) $\mu_i = \mu_j \forall i, j$

(d) $\pi_k = \frac{1}{k}$

Correct options: (b) (d)

Question-6

03:00

We apply the Expectation Maximization algorithm to f(D, Z, θ) where D denotes the data, Z denotes the hidden variables and θ the variables we seek to optimize. Which of the following are correct?

a) EM will always return the same solution which may not be optimal

b) EM will always return the same solution which must be optimal

c) The solution depends on the initialization

Question-6 - Correct answer

We apply the Expectation Maximization algorithm to f(D, Z, θ) where D denotes the data, Z denotes the hidden variables and θ the variables we seek to optimize. Which of the following are correct?

a) EM will always return the same solution which may not be optimal

b) EM will always return the same solution which must be optimal

c) The solution depends on the initialization

Correct options: (c)

Question-7

03:00

True or False: Iterating between the E-step and M-step of EM algorithms always converges to a local optimum of the likelihood.

a) True

b) False

Question-7 - Correct answer

True or False: Iterating between the E-step and M-step of EM algorithms always converges to a local optimum of the likelihood.

a) True

b) False

Correct options: (a)

Question-8

03:00

The number of parameters needed to specify a Gaussian Mixture Model with 4 clusters, data of dimension 5, and diagonal covariances is:

a) Lesser than 21
b) Between 21 and 30
c) Between 31 and 40
d) Between 41 and 50

Question-8

03:00

The number of parameters needed to specify a Gaussian Mixture Model with 4 clusters, data of dimension 5, and diagonal covariances is:

---

1. Mixing Coefficients (Priors) π_k:

- There are $K = 4$ mixing coefficients, but they must sum to 1, so only $K - 1 = 3$ are free parameters.

2. Means μ_k:

- Each cluster has a mean vector of dimension $D = 5$.
- Total parameters: $K \times D = 4 \times 5 = 20$.

3. Diagonal Covariance Matrices Σ_k:

- Since the covariance is diagonal, each cluster has $D = 5$ variance parameters (one per dimension).
- Total parameters: $K \times D = 4 \times 5 = 20$.

Question-8 - Correct answer

The number of parameters needed to specify a Gaussian Mixture Model with 4 clusters, data of dimension 5, and diagonal covariances is:

a) Lesser than 21

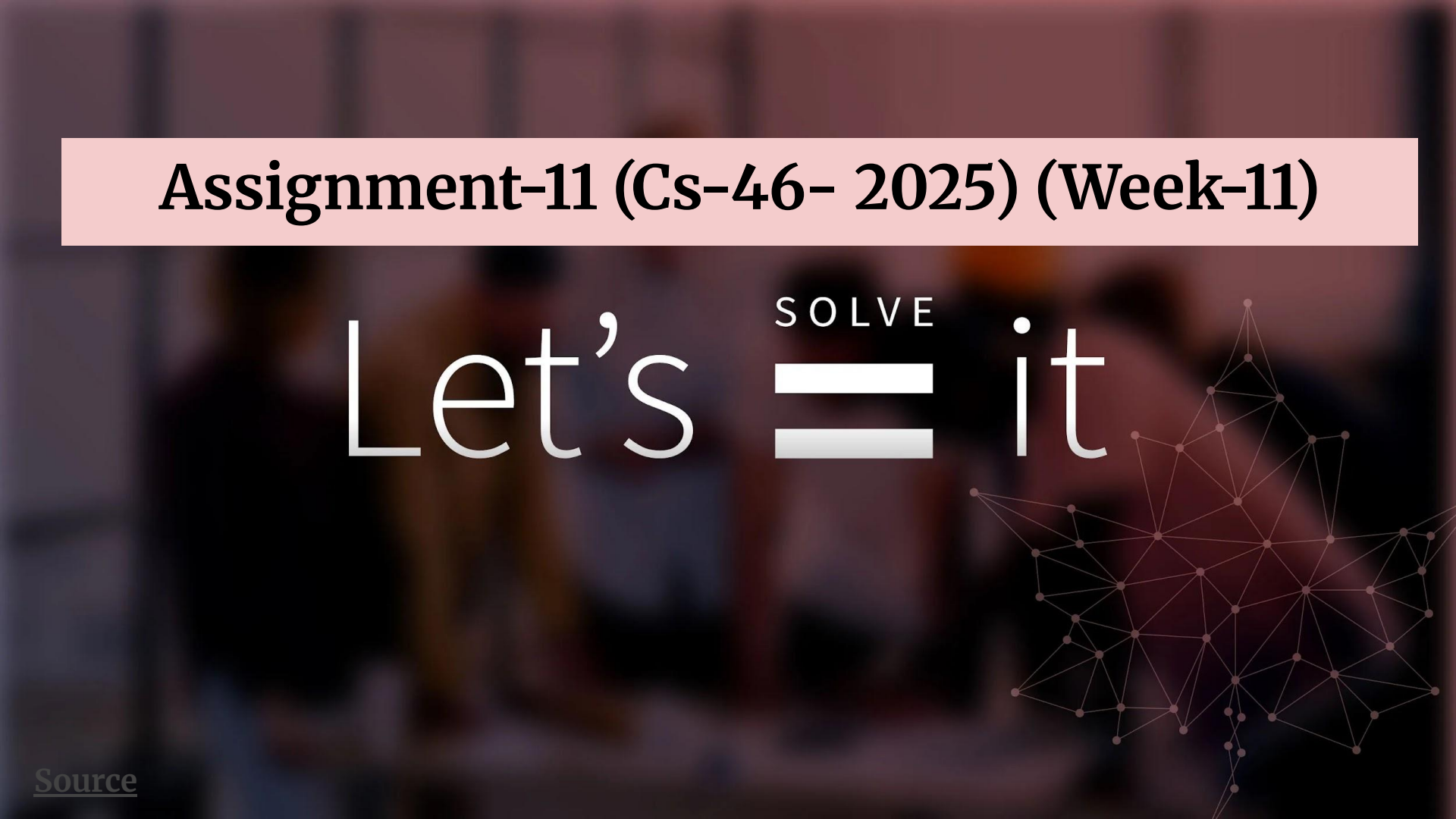
b) Between 21 and 30

c) Between 31 and 40

d) Between 41 and 50

Correct options: (d)

Assignment-11 (Cs-46- 2025) (Week-11)



Source

Question-1

01:00

Which of the following is/are estimated by the Expectation Maximization (EM) algorithm for a Gaussian Mixture Model (GMM)?

a) K (number of components)
b) πk (mixing coefficient of each component)
c) μk (mean vector of each component)
d) Σk (covariance matrix of each component)
e) None of the above

Question-1- Correct answer

Which of the following is/are estimated by the Expectation Maximization (EM) algorithm for a Gaussian Mixture Model (GMM)?

a) K (number of components)
b) πk (mixing coefficient of each component)
c) μk (mean vector of each component)
d) Σk (covariance matrix of each component)
e) None of the above

Correct options: (b) (c) (d)

Question-2

01:00

Which of the following is/are true about the responsibility terms in GMMs? Assume the standard notation used in the lectures.

a) $\Sigma k \gamma(znk)=1 \forall n$

b) $\Sigma n \gamma(znk)=1 \forall k$

c) $\gamma(znk) \in \{0,1\} \ \forall n,k$

d) $\gamma(znk) \in [0,1] \ \forall n,k$

e) $\pi j > \pi k \Rightarrow \gamma(znj) > \gamma(znk) \forall n$

Question-2- Correct answer

Which of the following is/are true about the responsibility terms in GMMs? Assume the standard notation used in the lectures.

a) $\Sigma_k \gamma(z_{nk}) = 1 \; \forall n$

b) $\Sigma_n \gamma(z_{nk}) = 1 \; \forall k$

c) $\gamma(z_{nk}) \in \{0,1\} \; \forall n,k$

d) $\gamma(z_{nk}) \in [0,1] \; \forall n,k$

e) $\pi_j > \pi_k \Rightarrow \gamma(z_{nj}) > \gamma(z_{nk}) \; \forall n$

Correct options: (a) (d)

Question-3

03:00

What is the update equation for μk in the EM algorithm for GMM?

a) $\mu(m)k = \sum Nn=1 \gamma(znk) | v(m) xn \sum Nn=1 \gamma(znk) | v(m-1)$

b) $\mu(m)k = \sum Nn=1 \gamma(znk) | v(m-1) xn \sum Nn=1 \gamma(znk) | v(m-1)$

c) $\mu(m)k = \sum Nn=1 \gamma(znk) | v(m-1) xnN$

d) $\mu(m)k = \sum Nn=1 \gamma(znk) | v(m) xnN$

Question-3 - Correct answer

What is the update equation for μk in the EM algorithm for GMM?

a) $\mu(m)k=\sum Nn=1\gamma(znk)|v(m)xn\sum Nn=1\gamma(znk)|v(m-1)$

b) $\mu(m)k=\sum Nn=1\gamma(znk)|v(m-1)xn\sum Nn=1\gamma(znk)|v(m-1)$

c) $\mu(m)k=\sum Nn=1\gamma(znk)|v(m-1)xnN$

d) $\mu(m)k=\sum Nn=1\gamma(znk)|v(m)xnN$

Correct options: (b)

Question-4

03:00

Select the correct statement(s) about the EM algorithm for GMMs.

a) In the mth iteration, the γ(znk) values are computed using the paramater estimates computed in the same iteration.
b) In the mth iteration, the γ(znk) values are computed using the paramater estimates computed in the (m−1)th iteration.
c) The Σk parameter estimates are computed during the E step.
d) The πk parameter estimates are computed during the M step.

Question-4 - Correct answer

Select the correct statement(s) about the EM algorithm for GMMs.

a) In the mth iteration, the $\gamma(znk)$ values are computed using the paramater estimates computed in the same iteration.
b) In the mth iteration, the $\gamma(znk)$ values are computed using the paramater estimates computed in the (m-1)th iteration.
c) The Σk parameter estimates are computed during the E step.
d) The πk parameter estimates are computed during the M step.

Correct options: (b)(d)

Question-5-7

01:00

For questions 5 to 7, use the following data consisting of 8 points (xi,yi)

x_i	1.0	1.8	0.9	8.2	9.1	7.8	7.6	8.8
y_i	1.5	1.2	1.6	7.3	7.2	9.5	8.0	7.5

Question-5

03:00

Fit a GMM with 2 components for this data. What are the mixing coefficients of the learned components? (Note: Use the sklearn implementation of GMM with random state = 0. Do not change the other default parameters).

a) (0.791, 0.209)

b) (0.538, 0.462)

c) (0.714, 0.286)

d) (0.625, 0.375)

Question-5 - Correct answer

Fit a GMM with 2 components for this data. What are the mixing coefficients of the learned components? (Note: Use the sklearn implementation of GMM with random state = 0. Do not change the other default parameters).

a) (0.791, 0.209)

b) (0.538, 0.462)

c) (0.714, 0.286)

d) (0.625, 0.375)

Correct options: (d)

Question-6

03:00

Using the model trained in question 5, compute the log-likelihood of the following points. Which of these points has the highest likelihood of being sampled from the model?

a) (2.0, 0.5)

b) (-1.0, -0.5)

c) (7.5, 8.0)

d) (5.0, 5.5)

Question-6 - Correct answer

Using the model trained in question 5, compute the log-likelihood of the following points. Which of these points has the highest likelihood of being sampled from the model?

a) (2.0, 0.5)
b) (-1.0, -0.5)
c) (7.5, 8.0)
d) (5.0, 5.5)

Correct options: (c)

Question-7

03:00

Let Model A be the GMM with 2 components that was trained in question 5. Using the same data from question 5, estimate a GMM with 3 components (Model B). (Note: Use the sklearn implementation of GMM with random state = 0 and all the other default parameters.)
Select the pair(s) of points that have the same label in Model A but different labels in Model B.

a) (1.0, 1.5) and (0.9, 1.6)
b) (1.8, 1.2) and (0.9, 1.6)
c) (7.8, 9.5) and (8.8, 7.5)
d) (7.8, 9.5) and (7.6, 8.0)
e) (8.2, 7.3) and (7.6, 8.0)

Question-7 - Correct answer

Let Model A be the GMM with 2 components that was trained in question 5. Using the same data from question 5, estimate a GMM with 3 components (Model B). (Note: Use the sklearn implementation of GMM with random state = 0 and all the other default parameters.)
Select the pair(s) of points that have the same label in Model A but different labels in Model B.

a) (1.0, 1.5) and (0.9, 1.6)
b) (1.8, 1.2) and (0.9, 1.6)
c) (7.8, 9.5) and (8.8, 7.5)
d) (7.8, 9.5) and (7.6, 8.0)
e) (8.2, 7.3) and (7.6, 8.0)

Correct options: (c) (e)

Question-8

03:00

Consider the following two statements.
Statement A: In a GMM with two or more components, the likelihood can attain arbitrarily high values.
Statement B: The likelihood increases monotonically with each iteration of EM.

a) Both the statements are correct and Statement B is the correct explanation for Statement A.
b) Both the statements are correct, but Statement B is not the correct explanation for Statement A.
c) Statement A is correct and Statement B is incorrect.
d) Statement A is incorrect and Statement B is correct.
e) Both the statements are incorrect.

Question-8 - Correct answer

Consider the following two statements.
Statement A: In a GMM with two or more components, the likelihood can attain arbitrarily high values.
Statement B: The likelihood increases monotonically with each iteration of EM.

a) Both the statements are correct and Statement B is the correct explanation for Statement A.
b) Both the statements are correct, but Statement B is not the correct explanation for Statement A.
c) Statement A is correct and Statement B is incorrect.
d) Statement A is incorrect and Statement B is correct.
e) Both the statements are incorrect.

Correct options: (b)


THANK YOU

Suggestions and Feedback

Next Session:

Tuesday:
15-Apr-2025
6:00 - 8:00 PM