

# Finite Mixture Model: Gaussian

RECAP

# SESSION III:

- Representing knowledge through graphical models
- Mixture Model:
  - Introduction
  - Multinomial Mixture Model
    - Known parameters
    - Unknown parameters
      - Posterior
      - Fully Collapsed

# OUTLINE:

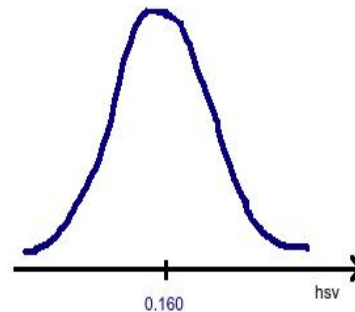
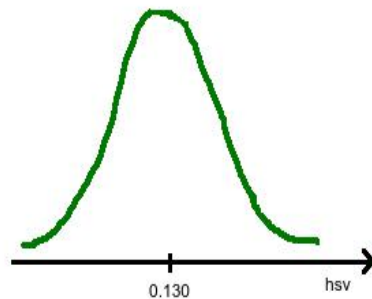
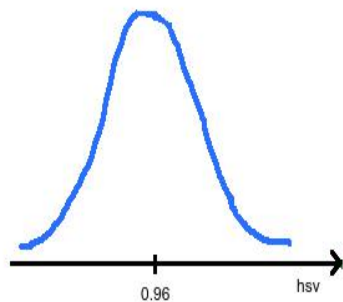
- Gaussian Mixture Model
  - Introduction
  - Known parameters
  - Unknown parameters
    - MLE
    - Posterior
    - Fully Collapsed

What is the picture representing (sky, river or forest)?



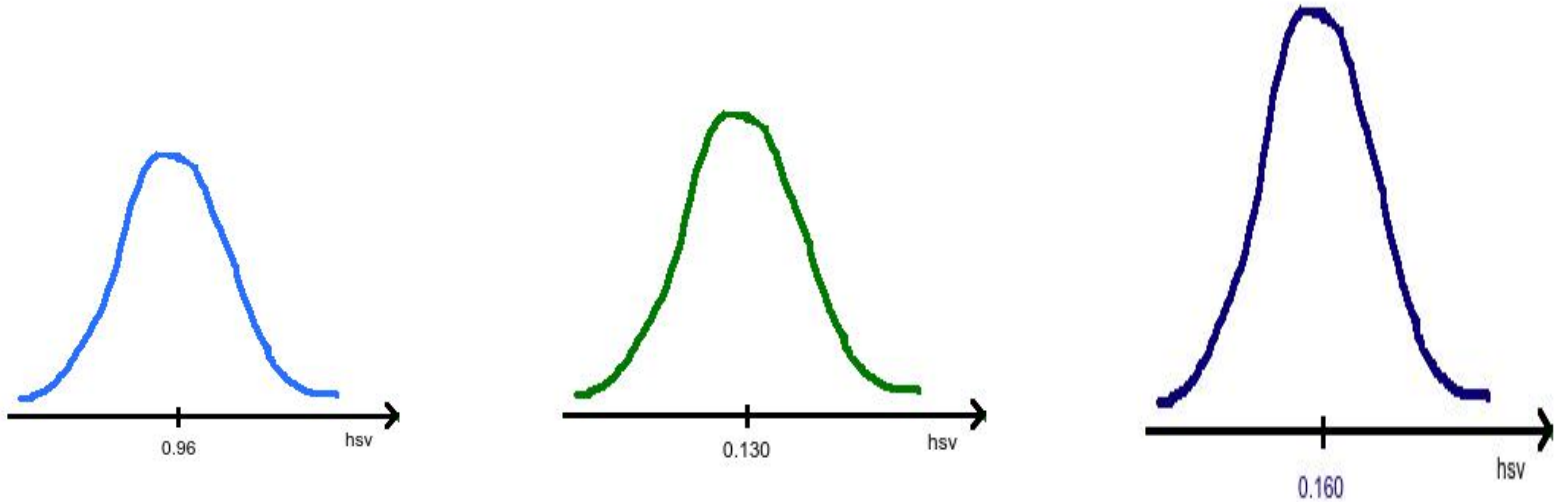
Picture from: Unsplash

# HSV color distribution



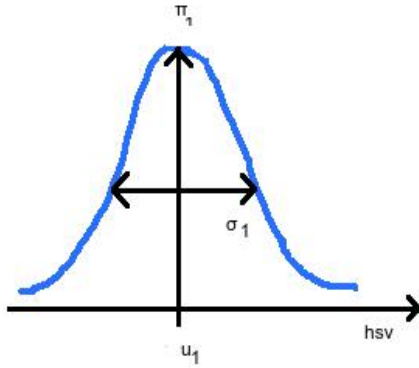
# Weight assignment

$$\Pi_k = [0.2, 0.3, 0.5]$$

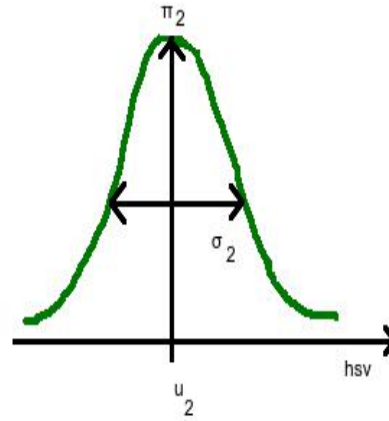


- Each value of  $\Pi_k$  lies between 0 and 1.
- The sum of all the values of  $\Pi_k$  must be equal to 1.

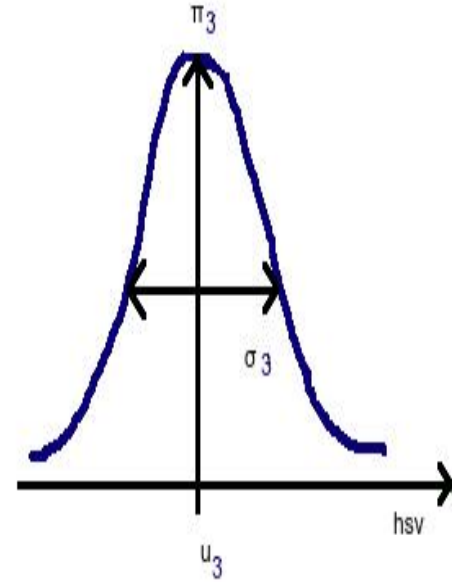
# Weight assignment:



$$N(x | \mu_1, \sigma_1)$$



$$N(x | \mu_2, \sigma_2)$$



$$N(x | \mu_3, \sigma_3)$$



# Univariate Gaussian Mixture Models

Combining all the distributions, we get:

$$p(x|\theta) = \pi_1 N(x|\mu_1, \sigma_1) + \pi_2 N(x|\mu_2, \sigma_2) + \pi_3 N(x|\mu_3, \sigma_3)$$

Here,  $\theta = \{\pi_k, \mu_k, \sigma_k\}$  are the parameters.

Therefore,  $p(x|\theta) = \sum_{k=1}^K \pi_k N(x|\mu_k, \sigma_k)$

[Code](#)

# Gaussian Mixture Distribution

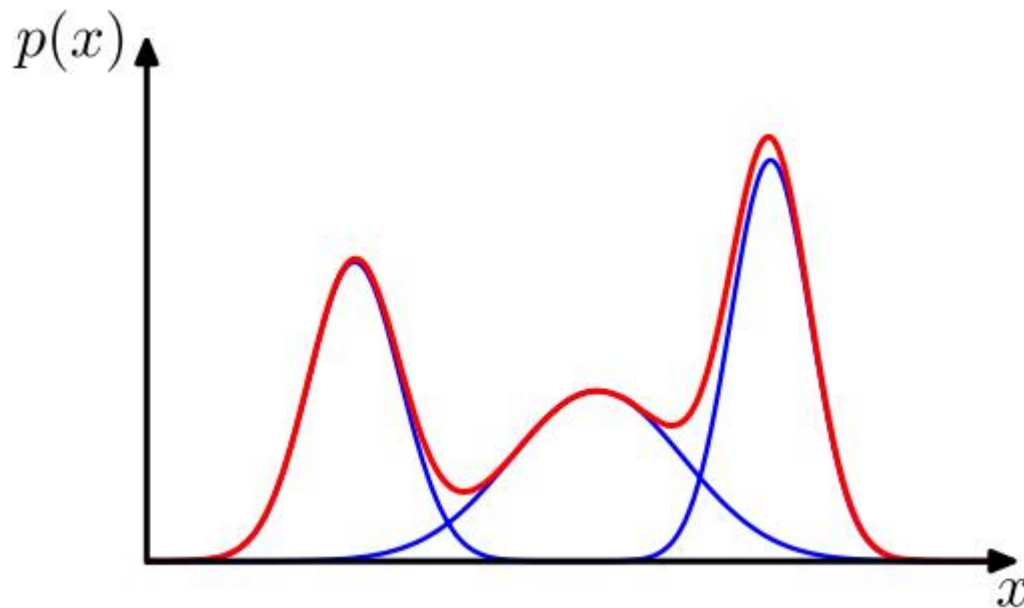


Figure 8: Illustration of Gaussian mixture distribution in one dimension showing three Gaussians (each scaled by a coefficient) in blue and their sum in red

# Gaussian mixture models (2D)

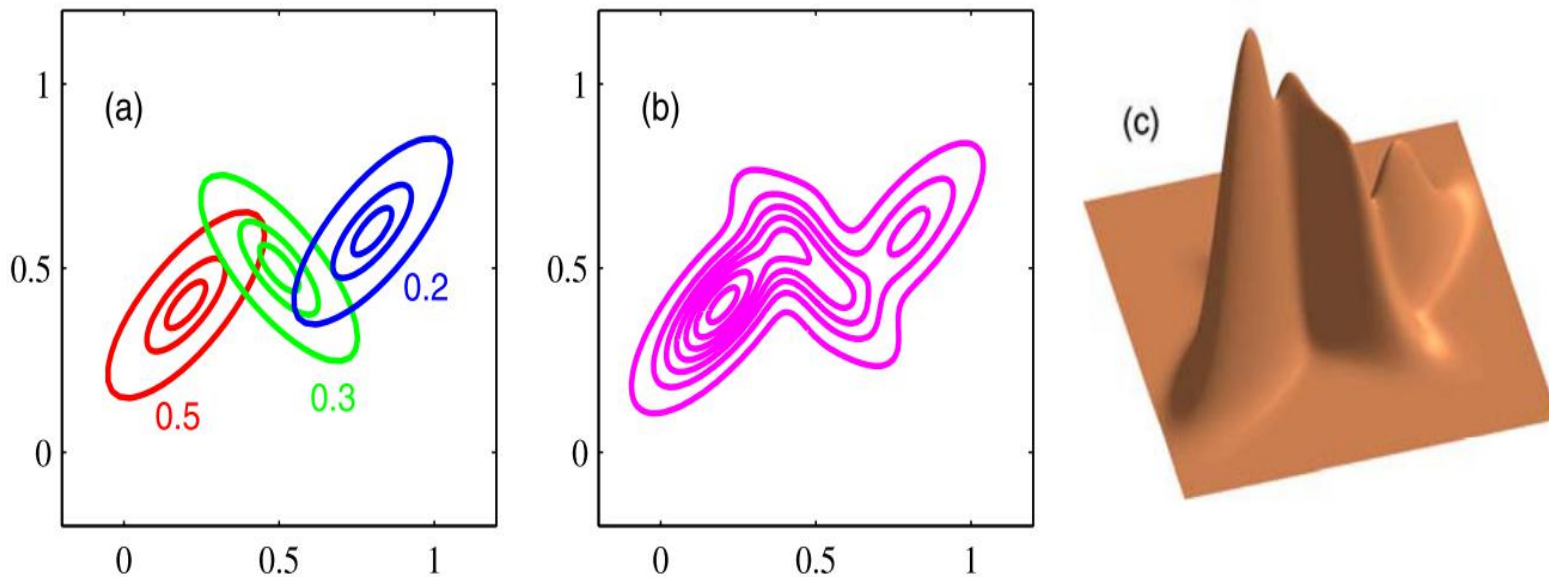


Figure 9: Illustration of a mixture of 3 Gaussians in a two-dimensional space. (a) Contours of constant density for each of the mixture components, in which the 3 components are denoted red, blue and green, and the values of the mixing coefficients are shown below each component. (b) Contours of the marginal probability density  $p(x)$  of the mixture distribution. (c) A surface plot of the distribution  $p(x)$ .

# Problem Statement:

Suppose there are three normally distributed mixture components with the following parameters.

```
params = {0: { $\pi_1$ : 0.3,  $\mu_1$ :12,  $\sigma_1$ :2},  
          1: { $\pi_2$ : 0.2,  $\mu_2$ :0,  $\sigma_2$ :1},  
          2: { $\pi_3$ : 0.5,  $\mu_3$ :20 ,  $\sigma_3$ :2}}
```

# Cases

1. We know the value of ' $\mu$ ' and ' $\sigma$ '.
2. We do not know the value of ' $\mu$ ' and ' $\sigma$ '.

# Case 1 (' $\theta$ ' known)

- The task is to determine which mixture component was used for generating certain data  $x_i$ .

*Algorithm:*

1. Randomly assign values to cluster, i.e,  $z_i = k$  with uniform probability.

2. For each  $i$ : i) Remove data point  $(x_i, z_i)$

ii) Compute normal pdf,  $p_k(x_i|\theta_k)$ , for each cluster  $k$ .

$$p_k(x_i|u_k, \sigma_k) = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}}$$

iii) Count the number of data points,  $c_k$ , in class  $k$ , given by:

$$c_k = |\{z_i = k \mid z_i \in Z\}|$$

iv) Estimate Gaussian parameters  $\theta_k = (\mu_k, \sigma_k)$  for each cluster using MLE:

$$\mu_k = \frac{1}{c_k} \sum_{x \in X_k} x, \text{ where } |X_k| = c_k$$

$$\sigma_k^2 = \frac{1}{c_k} \sum_{x \in X_k} (x - \mu_k)^2$$

Alternatively, use the posterior distribution to sample  $\theta$ , i. e.,

$$p(\theta_k | X, Z, \beta) = p(\theta_k | \{x_i | z_i = k \in Z\}, \beta_k)$$

v) Use Dirichlet categorical conditional distribution to compute the probability of observing class  $k$  having already observed counts,

$$p(z = k | c, \alpha) = \frac{c_k + \alpha_k}{C + A - 1}$$

vi) Add data point  $(x_i, z_i)$  back by sampling  $z_i = k$ , using,

$$z_i = k \sim p(z = k|c, \alpha) p_k(x_i|\theta_k)$$

where,  $p_k(x_i|\theta_k)$  and  $p(z = k|c, \alpha)$  are obtained from equation (ii) and (iv) respectively.

Mixture proportions can be estimated at the end.



## Case 2 (' $\theta$ ' unknown)(Finding parameters with MLE)

- The task is to determine which mixture component was used for generating certain data  $x_i$ .

*Algorithm:*

1. Randomly assign values to cluster, i.e,  $z_i = k$  with uniform probability.
2. For each  $i$ :
  - i) Remove data point  $(x_i, z_i)$
  - ii) Count the number of data points,  $c_k$ , in class  $k$ , given by:

$$c_k = |\{z_i = k \mid z_i \in Z\}|$$

iii) Compute normal pdf ,  $p_k(x_i|\theta_k)$ , for each cluster  $k$ .

$$p_k(x_i|u_k, \sigma_k) = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}}$$

iv) Use Dirichlet categorical conditional distribution to compute the probability of observing class  $k$  having already observed counts,

$$p(z = k|c, \alpha) = \frac{c_k + \alpha_k}{C + A - 1}$$

v) Add data point  $(x_i, z_i)$  back by sampling  $z_i = k$ , using,

$$z_i = k \sim p(z = k|c, \alpha) p_k(x_i|\theta_k)$$

where,  $p_k(x_i|\theta_k)$  and  $p(z = k|c, \alpha)$  are obtained from equation (iii) and (v) respectively.

Mixture proportions can be estimated at the end.

## Case 2 (' $\theta$ ' unknown)(Finding parameters with posterior Normal Gamma Distribution)

- The task is to determine which mixture component was used for generating certain data  $x_i$ .

*Algorithm:*

1. Randomly assign values to cluster, i.e,  $z_i = k$  with uniform probability.
2. For each  $i$ :
  - i) Remove data point  $(x_i, z_i)$
  - ii) Count the number of data points,  $c_k$ , in class  $k$ , given by:
$$c_k = |\{z_i = k \mid z_i \in Z\}|$$

iii) Sample ' $\mu$ ' and ' $\sigma$ ' with the posterior Normal Gamma Distribution:

$$\mu_{zk}, \sigma_{zk} \sim \text{NormalGamma}(\mu, \sigma \mid M, C, A, B)$$

$$M = \frac{c_m + \sum_{i=1}^n x_i}{c+n}, \quad C = c+n, \quad A = a+(n/2),$$

$$B = b+(1/2) * (cm^2 - CM^2 + \sum_i^n x_i^2)$$

Here,  $M, C, A, B$  are the parameters of Normal Gamma and  $n$  is the total datapoints in the cluster

iv) Compute normal pdf,  $p_k(x_i|\theta_k)$ , for each cluster  $k$ .

$$p_k(x_i|u_k, \sigma_k) = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}}$$

v) Use Dirichlet categorical conditional distribution to compute the probability of observing class  $k$  having already observed counts,

$$p(z = k|c, \alpha) = \frac{c_k + \alpha_k}{C + A - 1}$$

vi) Add data point  $(x_i, z_i)$  back by sampling  $z_i = k$ , using,

$$z_i = k \sim p(z = k|c, \alpha) p_k(x_i|\theta_k)$$

where,  $p_k(x_i|\theta_k)$  and  $p(z = k|c, \alpha)$  are obtained from equation (iii) and (v) respectively.

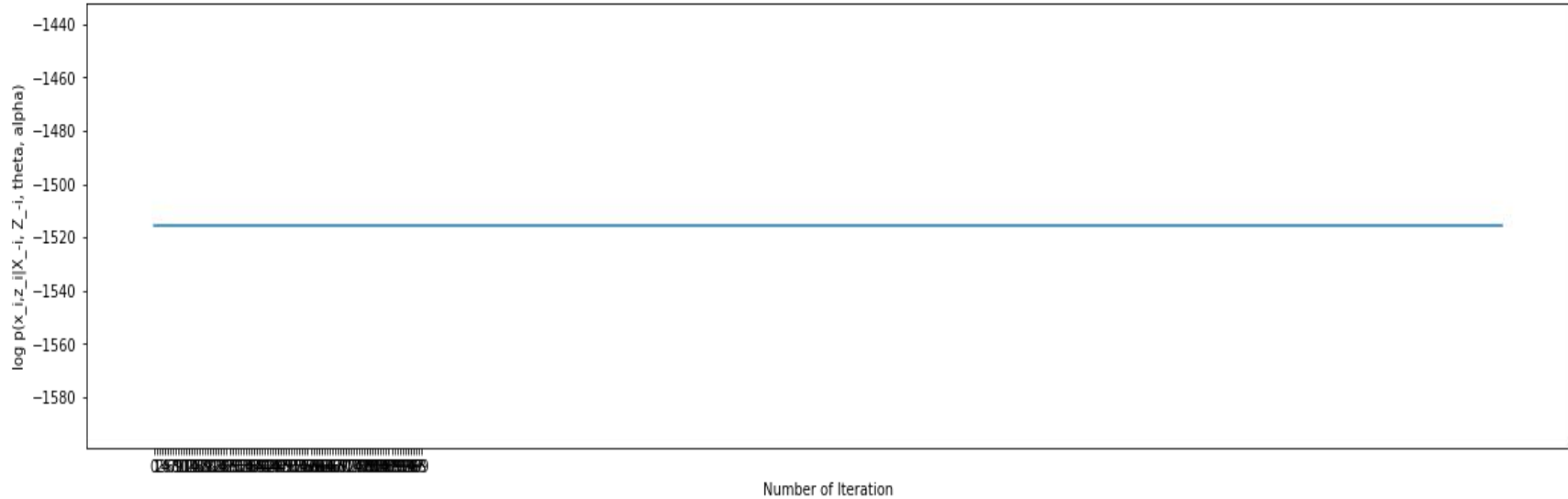
Mixture proportions can be estimated at the end.

- Let us consider the 3 mixture components with the following parameters:

```
params = {  
    0: { $\pi_1$ : 0.3,  $\mu_1$ :12,  $\sigma_1$ : 2},  
    1: { $\pi_2$ : 0.2,  $\mu_2$ :0,  $\sigma_2$ : 1},  
    2: { $\pi_3$ : 0.5,  $\mu_3$ :20,  $\sigma_3$ : 2}  
}
```

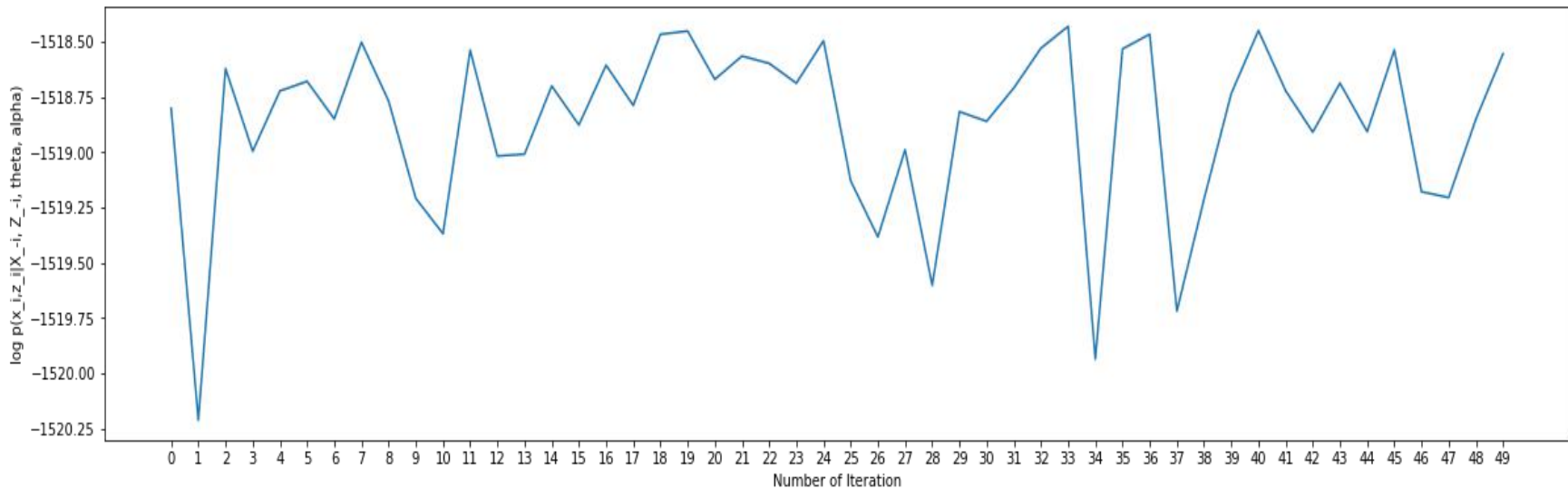
After running the above algorithms in both Case 1 and Case 2 for 50 iterations, the convergence graph looked as:

### 1. Known parameters (after 50 iterations)

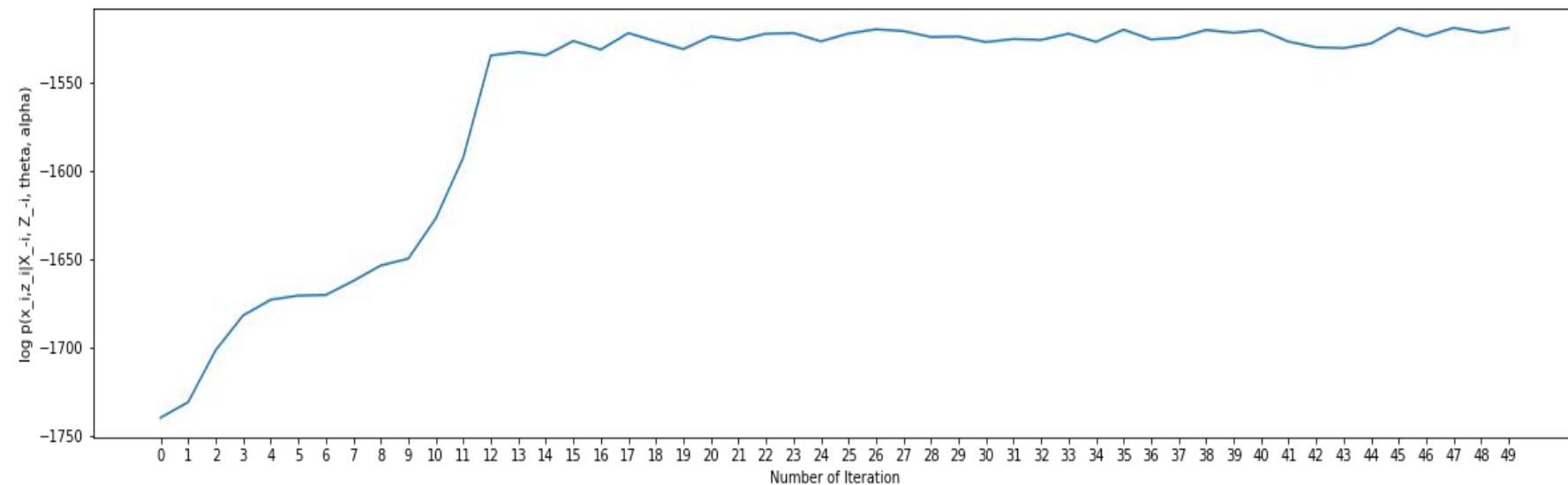




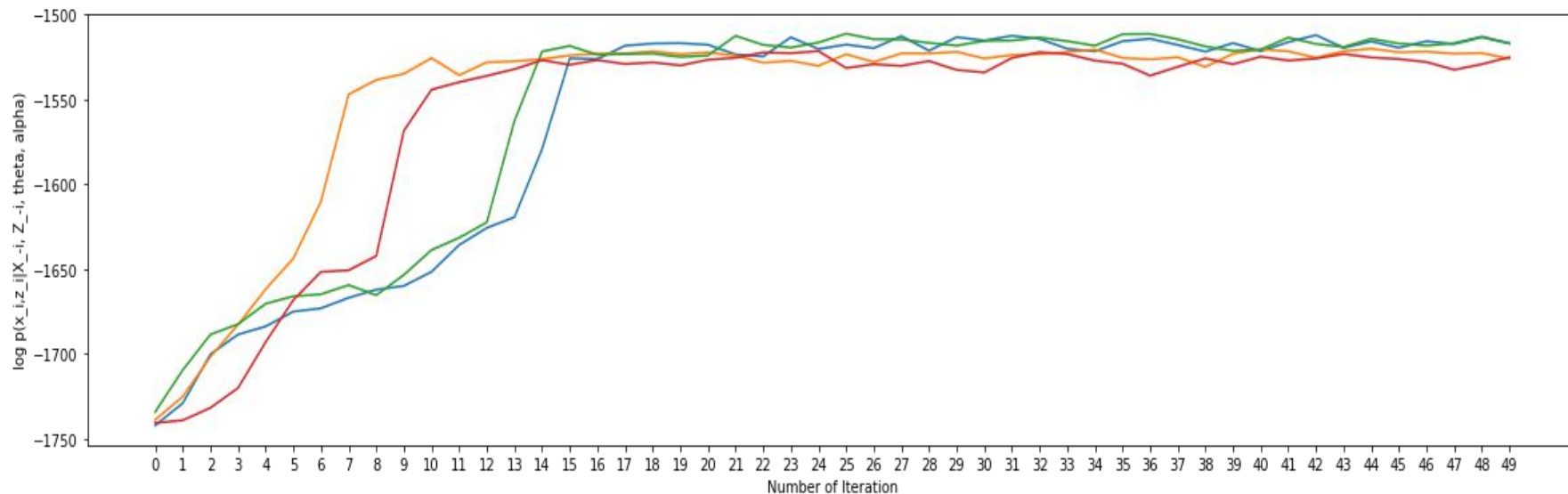
## 2. Finding parameters with MLE (after 50 iterations)



### 3. Finding parameters with normal gamma (after 50 iterations)



### 3. Finding parameters with normal gamma (after 50 iterations) [Multiple dataset]



# Sample code:

- Finite mixtures with gaussian prior: (Known and unknown parameters) (Unknown ' $\mu$ ' and ' $\sigma$ ' are found by taking average and finding variance respectively.)

[https://colab.research.google.com/drive/17xHqCrklyWhAZ-1\\_tK\\_9lEotoZRtfmJh](https://colab.research.google.com/drive/17xHqCrklyWhAZ-1_tK_9lEotoZRtfmJh)

- Finite mixtures with gaussian prior: (Unknown parameters are found using samples from normal gamma)

<https://colab.research.google.com/drive/1iStXsgTKQfrHTn9kfAlKGfLJz1ZCneSL>

# Fully Collapsed Gibbs Sampling

- **Gaussian Mixture Model**

The within class gaussian parameters  $\mu_k, \lambda_k$  can be integrated out using normal gamma distribution.

Thus, instead of

$$p(x_i | \mu_i, \lambda_i)$$

we need to compute

$$\frac{p(X_k | m, c, a, b)}{p(X_{k,-i} | m, c, a, b)}$$

# Fully Collapsed Gibbs Sampling

- **Gaussian Mixture Model**

Using NormalGamma as conjugate prior for Normal distribution

$$\frac{p(X_k | m, c, a, b)}{p(X_{k,-i} | m, c, a, b)} = \frac{\frac{1}{\frac{B'^A}{\tau(A')} \sqrt{\frac{C'}{2\pi}}}}{\frac{1}{\frac{B^A}{\tau(A)} \sqrt{\frac{C}{2\pi}}}} = \sqrt{\frac{C}{C'}} \frac{B^A}{B'^A} \frac{\tau(A')}{\tau(A)}$$

$$\begin{aligned} \text{Here, } \frac{\tau(A')}{\tau(A)} &= \frac{\tau(a + \frac{n+1}{2})}{\tau(a + \frac{n}{2})} \\ &= \frac{\tau(A + \frac{1}{2})}{\tau(A)} \end{aligned}$$

# Fully Collapsed Gibbs Sampling

- **Gaussian Mixture Model**

Multiplying both sides of equation (3) by  $\tau(1/2)$ ,

$$\frac{\tau(A')}{\tau(A)} = \frac{\tau(a + \frac{n+1}{2})}{\tau(a + \frac{n}{2})} * \frac{\tau(1/2)}{\tau(1/2)}$$

As  $\text{Beta}(x, y) = \frac{\tau(x)\tau(y)}{\tau(x+y)}$ , we get,

$$\frac{\tau(A')}{\tau(A)} = \frac{\tau(1/2)}{\text{Beta}(x, y)}$$

# Fully Collapsed Sampler

- **Algorithm: Gaussian Mixture Model**

1. Randomly assign values to cluster, i.e,  $z_i = k$  with uniform probability.

2. For each cluster  $k$ , compute  $\sqrt{\frac{C}{C'}} \frac{B^A}{B'^{A'}} \frac{\tau(1/2)}{\text{Beta}(x,y)}$

3. Add data point  $(x_i, z_i)$  back by sampling  $z_i = k$ , using:

$$z_i = k \sim \frac{c_k + A/k}{N+A-1} \left( \sqrt{\frac{C}{C'}} \frac{B^A}{B'^{A'}} \frac{\tau(1/2)}{\text{Beta}(x,y)} \right)$$

Mixture proportions can be estimated at the end.



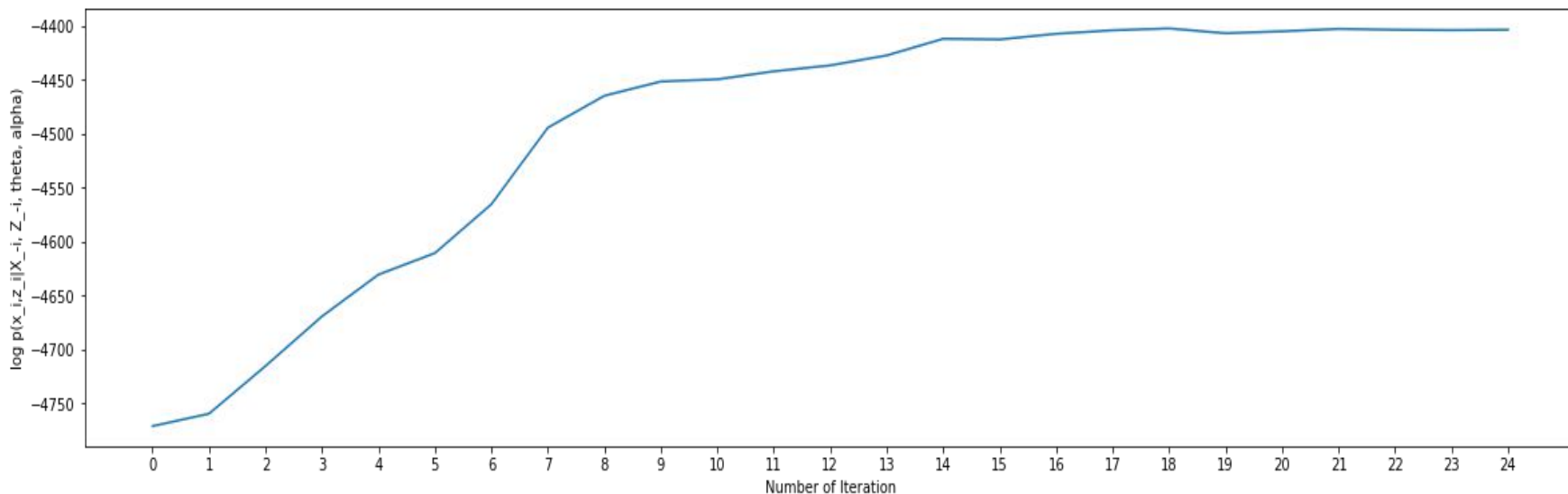
- Suppose we have three mixture components with the following parameters:

```
param = {0: {'pi': 0.3, 'mu':12, 'sigma':2},  
         1: {'pi': 0.2, 'mu':5, 'sigma':5},  
         2: {'pi': 0.5, 'mu':20 , 'sigma':2}  
}
```

```
total_iter = 1000
```

# Fully collapsed sampler (after 50 iterations)

(As algorithm converge at 24 iterations, iterations after 24 are skipped)



# Sample code

- Fully collapsed sampler (Gaussian Mixture Model):

<https://colab.research.google.com/drive/1Kcp0ZrL-XCpOAGFVoseObFTe6QpX48dq#scrollTo=cs653ksvm9Gz>

# Takeaways:

- We learned about Gaussian Mixture Model.
- We learned the way to model gaussian mixture model using gibbs sampling in the cases when parameters are known and unknown.

# References:

- Suresh Manandhar. *Bayesian ML : Posterior Distributions and Mixture Models*  
*Continuous Probability Density Function*
- Christopher Bishop. *Pattern Recognition and Machine Learning*

Thank you