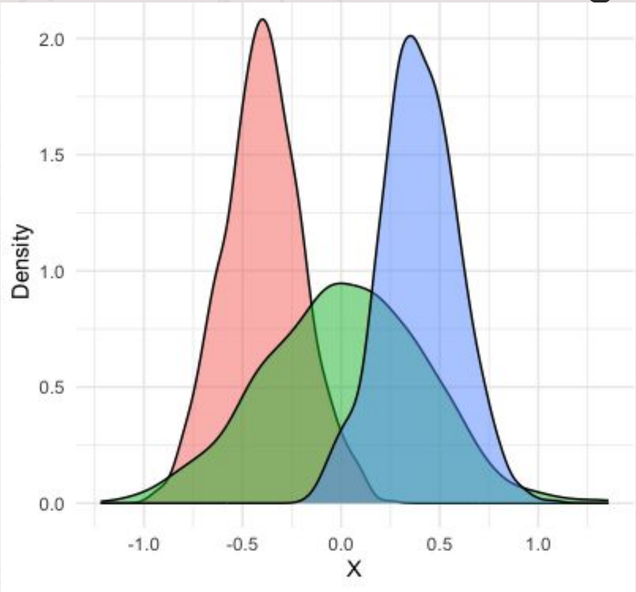
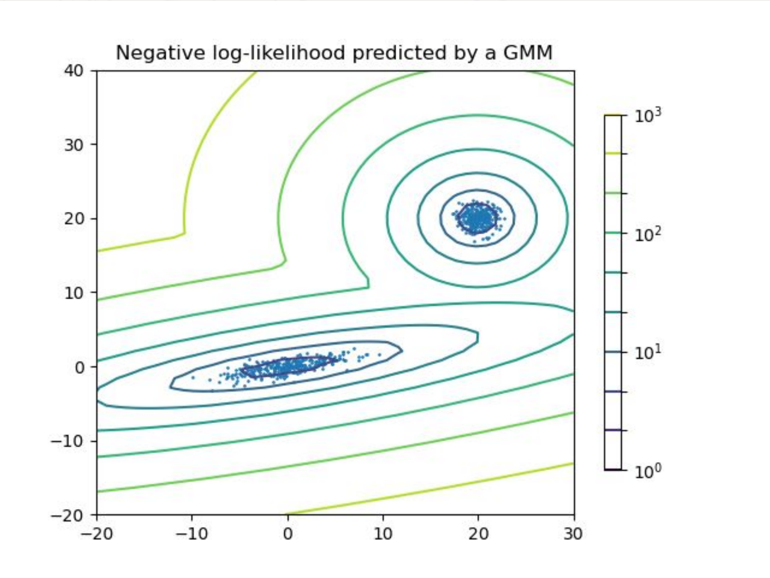
Gaussian Mixture Models & The Expectation-Maximization Algorithm

Liam Donovan & Makar Pronin

The Multimodal Problem

- Sometimes sampling leads to multimodal normal distributions (CLT)
  - Some with higher peaks than others





The Gaussian Mixture Model (GMM)

- Let θ be the set of parameters for each distribution

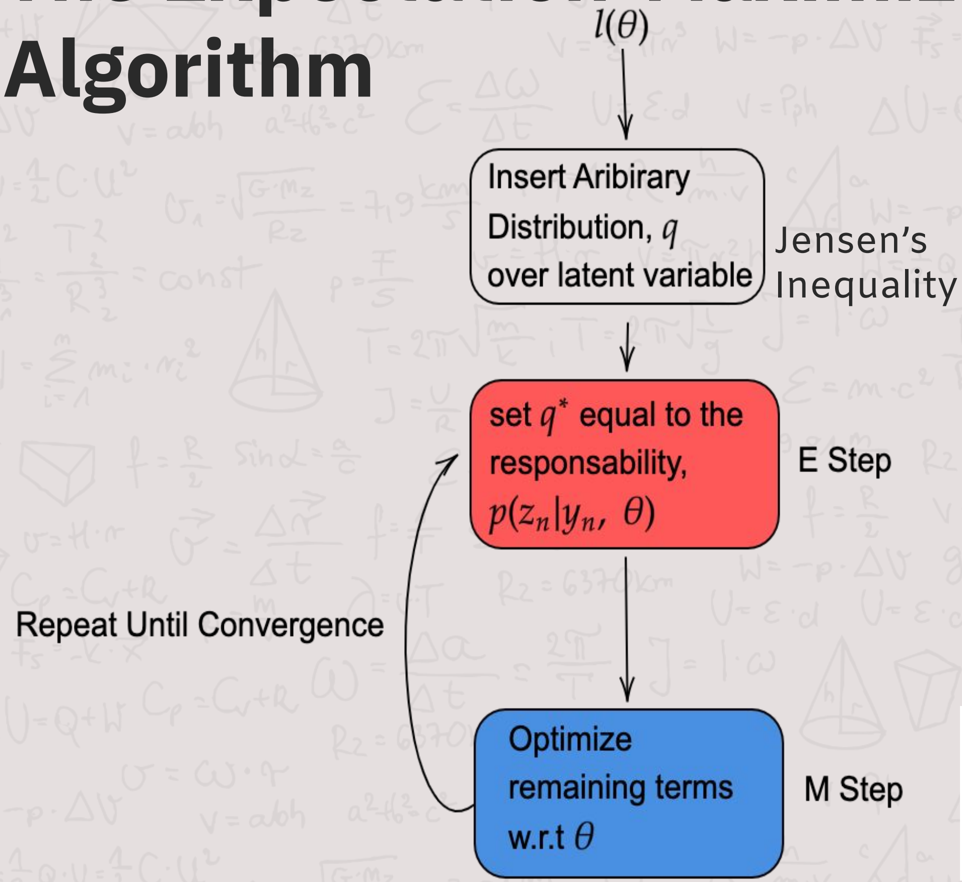
$$p(\mathbf{y}|\boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{y}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

- To satisfy the unit measure (area under curve is 1)

$$\sum_{k=1}^{K} \pi_k = 1$$

- Weighted Sum of Gaussians
- Try to approximate θ

The Expectation-Maximization (EM) Algorithm



$$
\begin{aligned}
l(\theta) &= \sum_{n=1}^{N} \log(p(y_n|\theta)) \\
&= \sum_{n=1}^{N} \log\left(\sum_{z_n=1}^{K} p(y_n, z_n|\theta)\right)
\end{aligned}
$$

$$
l(\theta) \geq -D_{KL}(q_n \,||\, p(z_n|y_n, \theta)) + \sum_{n=1}^{N} \log(p(y_n|\theta))
$$

$$
l(\theta) \geq \sum_{n=1}^{N} \mathbb{E}_{q_n}(\log(p(y_n, z_n|\theta)) + \mathbb{H}(q_n)
$$

The Expectation-Maximization (EM) Algorithm

The GMM Case

2. **E step**. Evaluate the responsibilities using the current parameter values

$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}. \tag{9.23}$$

3. **M step**. Re-estimate the parameters using the current responsibilities

$$\boldsymbol{\mu}_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \mathbf{x}_n \tag{9.24}$$

$$\boldsymbol{\Sigma}_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \left(\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{new}}\right) \left(\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{new}}\right)^{\mathrm{T}} \tag{9.25}$$

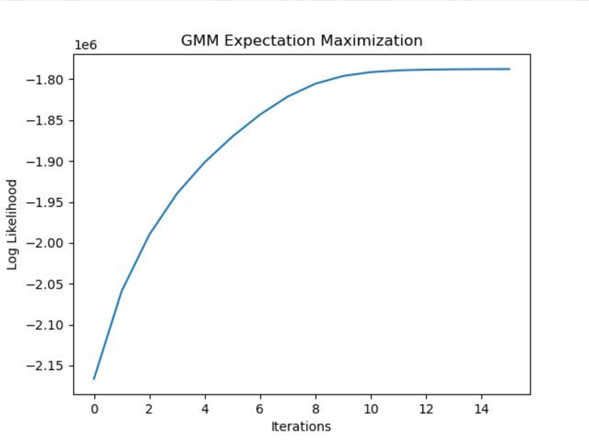
$$\pi_k^{\text{new}} = \frac{N_k}{N} \tag{9.26}$$

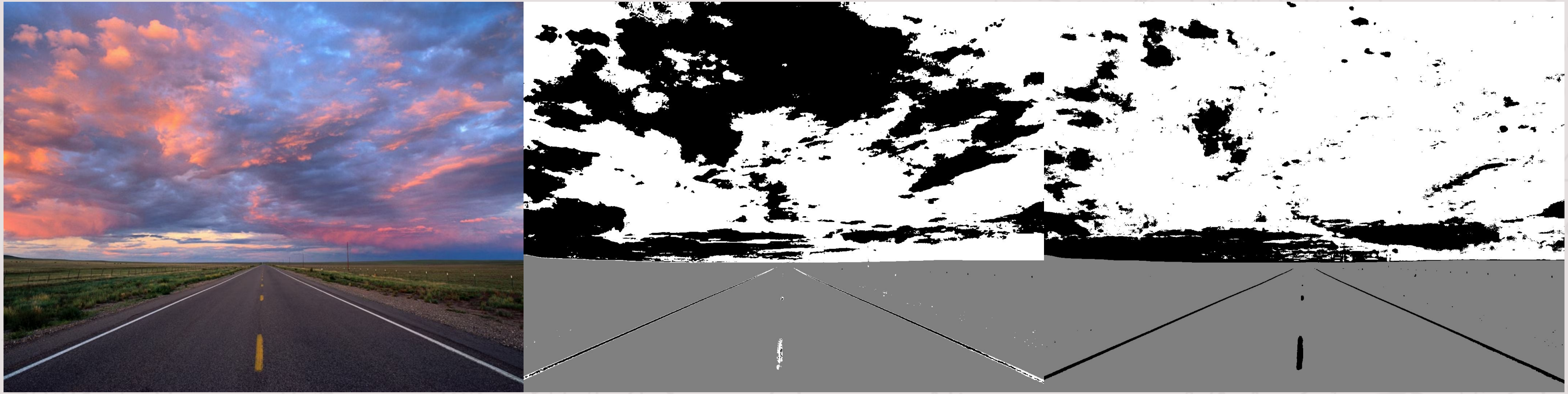
where

$$N_k = \sum_{n=1}^{N} \gamma(z_{nk}). \tag{9.27}$$

Image Segmentation

- Useful for image compression, especially in training computer vision.
- Clusters take into account only RGB values, without pixel locations.



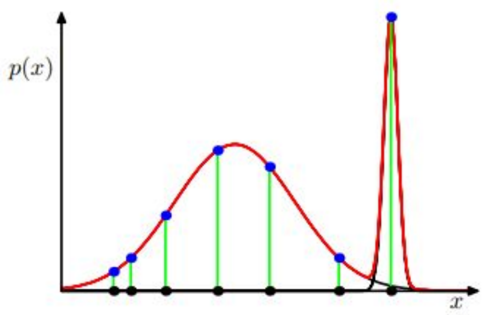


Problems with MLE

$$L(\theta|X) = \prod_{i=1}^{N} \sum_{k=1}^{K} \pi_k N(x_i|\mu_k, \Sigma_k)$$

$$N(x|\mu, \Sigma) = \frac{1}{(2\pi)^{d/2}\sqrt{|\Sigma|}} e^{\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)}$$

- What happens when x=μ? Exponent term = 1, Σ (covariance) approaches 0. In other words, we get parameter collapse when the whole Gaussian cluster fits only one data sample.



- In reality, our algorithm could not even converge in a reasonable time with Quasi-Newton methods.

References

Bishop, C. M. (2006). *Pattern recognition and machine learning by Christopher M. Bishop*. Springer Science+Business Media, LLC.

Lai, M., & Lai, T. (2019, September). *LAITUAN245/Image-segmentation-GMM: Image segmentation using gaussian mixture models*. GitHub. https://github.com/laituan245/image-segmentation-GMM

Murphy, K. P. (2022). *Probabilistic machine learning: An introduction*. The MIT Press.