Given datapoint x_n, what is the prob. that it came from Gaus k?

$$p(z_{nk}=1 \mid x_n)$$

The probability of observing a point that comes from Gaus k:

$$\pi_k = p(z_k = 1)$$

z = set of all possible latent variables for point x_n

$$\mathbf{z} = \{z_1, \ldots, z_K\}$$

The probability of observing our datapoint given it came from Gaus K

$$\rightarrow \quad p(x_n \mid z) = \prod_{k=1}^{K} \mathcal{N}(x_n \mid \mu_k, \Sigma_k)^{z_k}$$

Probability of observing our datapoint given it came from Gaus k is just the Gaussian itself

$$p(x_n | z) = \mathcal{N}(x_n | \mu_3, \Sigma_3)$$

<u>Goal</u>: determine probability of z given datapoint x_n

Chain rule: $P(A \cap B) = P(B|A)\,P(A)$

So the probability of x_n and z:

$$p(x_n, z) = p(x_n | z)\, p(z)$$

Marginalization: to get $p(x_n)$

$$p(x_n) = \left[\sum_{k=1}^{K} p(x_n | z)\, p(z)\right] = \left[\sum_{k=1}^{K} \pi_k \cdot \mathcal{N}(x_n | \mu_k, \Sigma_k)\right]$$

To determine optimal values for these, determine max likeliho which is the <u>joint probability of all observations x_n</u>:

$$p(\mathbf{X}) = \left[\prod_{n=1}^{N} p(x_n)\right] = \left[\prod_{n=1}^{N} \sum_{k=1}^{K} \pi_k \cdot \mathcal{N}(x_n | \mu_k, \Sigma_k)\right]$$

Apply log:

$$\ln p(X) = \sum_{n=1}^{N} \ln \sum_{k=1}^{K} \pi_k \cdot \mathcal{N}(x_n | \mu_k, \Sigma_k)$$

...which is the joint probability of all observations & latent variables.

$$\ln p(X, Z|\theta^*) = \sum_{n=1}^{N}\sum_{k=1}^{K} z_{nk}\left[\ln \pi_k + \ln N(x_n|\mu_k, \Sigma_k)\right]$$

Estimate parameters by maximizing Q with respect to parameters

Replace $\ln p(X, Z|\theta^*)$ in above equation:

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

$\boxed{\text{Step 3:}}$ "Maximization Step," Find the revised parameters, θ^*, using:

$$\theta^* = \arg\max_{\theta} Q(\theta^*,\theta)$$

where $Q(\theta^*,\theta)$ is taken from above Step 2.

Also, include a Lagrange multiplier to take into account that π values sum up to 1:

$$Q(\theta^*,\theta) = \sum_{n=1}^{N}\sum_{k=1}^{K} \gamma(z_{nk})\left[\ln \pi_k + \ln N(x_n|\mu_k, \Sigma_k)\right] - \lambda\left(\sum_{k=1}^{K}\pi_k - 1\right)$$

$\boxed{\text{Derivative of Q:}}$

$$\frac{d\,Q(\theta^*,\theta)}{d\pi_k} = \sum_{n=1}^{N}\frac{\gamma(z_{nk})}{\pi_k} - \lambda = 0$$

$$\sum_{n=1}^{N}\gamma(z_{nk}) = \pi_k\lambda \quad\Rightarrow\quad \sum_{k=1}^{K}\sum_{n=1}^{N}\gamma(z_{nk}) = \sum_{k=1}^{K}\pi_k\lambda$$

Since $\sum_{k=1}^{K} \pi_k = 1$, and the sum of probabilities γ over k will also give us 1, then $\lambda = N$

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

Then differentiate Q with respect to μ and Σ, equate the derivative to 0, and solve for parameters with log-likelihood equation:

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

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