Isabel Restrepo
Brown University

1 Gaussian Mixtures Models (GMMs)

The distribution of a mixture of gaussians is given by:

$$p(x) = \sum_{i=1}^{k} \pi_i N(x|\mu_i, \sigma_i) \tag{1}$$

Here the random variable has been written explicitly.

2 Distance Measures Between Distributions

There are several distance measures used across literature to compare probability distribution functions. Denote $d(p_1, p_2)$ the distance measure between two distributions. Then this measures can be described by the following properties:

- Must be non-negative
- Satisfies the reflexivity property: $d(p_1, p_1) = 0$
- May be symmetric $d(p_1, p_2) = d(p_1, p_2)$
- May satisfy the triangle inequality : $d(p_1, p_2) + d(p_2, p_3) \geq d(p_1, p_3)$
- May satisfy a relaxed version of the triangle inequality:
  $d(p_1, p_2) + d(p_2, p_3) \geq c * d(p_1, p_3), c < 1$

2.1 A Rogue's Gallery

- Kullback-Leiber Distance:
  The KL distance is the relative entropy between the distributions. For continuous random variables, it is given by

  $$d_K L(p_1, p_2) = \int p_1(x) \log \frac{p_1(x)}{p_2(x)} dx \tag{2}$$

  This distance is not symmetric and does not satisfies the triangle inequality. It can be computationally expensive to use , but asymptotically it is the optimal distance measure .

- L-p Distance:
  For continuos random variables it is given by:

$$d_L(p_1, p_2) = \left[\int (p_1(x) - p_2(x))^p\right]^{1/p} dx \tag{3}$$

  commonly used for p=2

- Bhattacharya Distance

$$d_B(p_1, p_2) = -\ln\left(\int \sqrt{p_1(x)p_2(x)}dx\right) \tag{4}$$

  This measure is symmetric and satisfies triangle inequality in certain configurations

3 Analytical Derivation for GMMs

The ideal KL distance does not have an analytical solution for Gaussian Mixture Models with more that one component. An approximation can be achieved through stochastic methods. Here, the L2 distance is explored because for GMM a deterministic expression exists.

3.1 L2 Distance

3.1.1 Notation

Let

$$p_1(x) = \sum_{i=1}^{K_1} w_i G_i \quad p_2(x) = \sum_{i=1}^{K_2} v_i F_i \tag{5}$$

$$\text{where } G_i \sim N(\mu_{G_i}, \sigma_{G_i}) \text{ and } F_i \sim N(\mu_{F_i}, \sigma_{F_i}) \tag{6}$$

3.1.2 The distance

$$\begin{aligned} d_L(p_1, p_2) &= \left[\int (p_1(x) - p_2)^2 dx\right]^{1/2} \qquad (7) \\ &= \left[\int \left(\sum_{i=1}^{K_1} w_i G_i - \sum_{i=1}^{K_2} v_i F_i\right)^2 dx\right]^{1/2} \qquad (8) \end{aligned}$$

For a moment let's ignore the outer most exponent and let's expand the integral

$$\begin{aligned}
\int \left(\sum_{i=1}^{K_1} w_i G_i - \sum_{i=1}^{K_2} v_i F_i \right)^2 dx &= \int \left(\left(\sum_{i=1}^{K_1} w_i G_i \right)^2 - 2\sum_{i=1}^{K_1} w_i G_i \sum_{i=1}^{K_2} v_i F_i + \left(\sum_{i=1}^{K_2} v_i F_i \right)^2 \right) dx \\
&= \int \Big(\sum_{i=1}^{K_1} {w_i}^2 {G_i}^2 + 2 \sum_{i=1}^{K_1} \sum_{j=i+1}^{K_1} w_i w_j G_i G_j \\
&\quad - 2 \sum_{i=1}^{K_1} \sum_{j=1}^{K_2} w_i v_j G_i F_j + \sum_{i=1}^{K_2} {v_i}^2 {F_i}^2 \\
&\quad + 2 \sum_{i=1}^{K_2} \sum_{j=i+1}^{K_2} v_i v_j F_i F_j \Big) dx \\
&= \sum_{i=1}^{K_1} {w_i}^2 \int {G_i}^2 dx + 2\sum_{i=1}^{K_1} \sum_{j=i+1}^{K_1} w_i w_j \int G_i G_j dx \\
&\quad -2 \sum_{i=1}^{K_1} \sum_{j=1}^{K_2} w_i v_j \int G_i F_j dx + \sum_{i=1}^{K_2} {v_i}^2 \int {F_i}^2 dx + \\
&\quad 2\sum_{i=1}^{K_2} \sum_{j=i+1}^{K_2} v_i v_j \int F_i F_j dx
\end{aligned} \tag{9}$$

3.1.3 The Gaussian Product Rule

The previously distance measure contains $O(K_1^2 + K_2^2)$ terms, each term consisting of an infinite integral over a product of gaussian. Luckily products of gaussians are well defined and can it is explained by the following rule.

The product of two gaussians is an unormalized gaussian

Demonstration of the Gaussian product rule :

The product of two gaussian is an unormalized gaussian. i.e

$$N(\mu_1, \sigma_1^2) \cdot N(\mu_2, \sigma_2^2) = z_3 N(\mu_3, \sigma_3) \tag{10}$$

This product takes the form

$$\frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-a(x-\mu_1)^2} \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-b(x-\mu_2)^2} \tag{11}$$

where $a = \frac{1}{2\sigma_1^2}$ and $b = \frac{1}{2\sigma_2^2}$

For a moment let's worry only about the terms of the exponential and try to rearrange them to satisfy (10).

$$
\begin{aligned}
a(x-\mu_1)^2 + b(x-\mu_2)^2 &= ax^2 - 2ax\mu_1 + a\mu_1^2 + bx^2 - 2bx\mu_2 + b\mu_2^2 \\
&= x^2(a+b) - 2x(a\mu_1 + b\mu_2) + a\mu_1^2 + b\mu_2^2 \\
&= (a+b)\left[x^2 - \frac{2x(a\mu_1+b\mu_2)}{a+b} + \frac{a\mu_1^2+b\mu_2^2}{a+b}\right] \\
&= (a+b)\left[x^2 - \frac{2x(a\mu_1+b\mu_2)}{a+b} + \frac{(a+b)}{(a+b)}\frac{a\mu_1^2+b\mu_2^2}{a+b}\right] \\
&= (a+b)\left[x^2 - \frac{2x(a\mu_1+b\mu_2)}{a+b} + \frac{a^2\mu_1^2+ab\mu_1^2+ab\mu_2^2+b^2\mu_2^2}{(a+b)^2}\right] \qquad (12) \\
&= (a+b)\left[x^2 - \frac{2x(a\mu_1+b\mu_2)}{a+b} + \frac{a^2\mu_1^2+2ab\mu_1\mu_2+b^2\mu_2^2}{(a+b)^2}\right] \\
&\quad + \left[\frac{-2ab\mu_1\mu_2+ab\mu_1^2+ab\mu_2^2}{a+b}\right] \\
&= (a+b)\left(x - \frac{a\mu_1+b\mu_2}{a+b}\right)^2 + \frac{ab}{a+b}(\mu_1-\mu_2)^2
\end{aligned}
$$

Shape of the gaussian :

We can now expand Eq. (10) and examine in detail the unnormalized gaussian (i.e, it's mean, variance and nomalization constant.) In more detail:

$$\frac{1}{\sqrt{2\pi\sigma_1^2}}e^{-a(x-\mu_1)^2}\cdot\frac{1}{\sqrt{2\pi\sigma_2^2}}e^{-b(x-\mu_2)^2} = \frac{1}{2\pi\sigma_1\sigma_2}e^{-(a(x-\mu_1)^2+b(x-\mu_2)^2)} \qquad (13)$$

Using the result from Eq. (12) this becomes

$$\frac{1}{2\pi\sigma_1\sigma_2}e^{-\frac{ab}{a+b}(\mu_1-\mu_2)^2}e^{-(a+b)\left(x-\frac{a\mu_1+b\mu_2}{a+b}\right)^2} \qquad (14)$$

$$\frac{ab}{a+b} = \frac{1}{2(\sigma_1^2+\sigma_2^2)} \qquad (15)$$

$$a+b = \frac{1}{2}\frac{\sigma_1^2+\sigma_2^2}{(\sigma_1^2\sigma_2^2)} \qquad (16)$$

Thus,

$$\sigma_3 = \frac{(\sigma_1^2\sigma_2^2)}{\sigma_1^2+\sigma_2^2} = \left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right)^{-1} \tag{17}$$

$$\mu_3 = \frac{\mu_1\sigma_2^2+\mu_2\sigma_1^2}{\sigma_1^2+\sigma_2^2} = \sigma_3^2\left(\frac{\mu_1}{\sigma_1^2}+\frac{\mu_2}{\sigma_2^2}\right) \tag{18}$$

$$z_3 = \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}\exp\left(-\frac{1}{2}\frac{(\mu_1-\mu_2)^2}{(\sigma_1^2+\sigma_2^2)}\right) = \left(\frac{\sigma_3^2}{2\pi\sigma_1^2\sigma_2^2}\right)^{1/2}\exp\left(-\frac{1}{2}\frac{\sigma_3^2}{\sigma_1^2\sigma_2^2}(\mu_1-\mu_2)^2\right) \tag{19}$$

It is useful to note that if the two gaussian share the same mean and variance (denote them by μ and σ^2) the above results can be further simplified. i.e

$$N(\mu,\sigma^2)\cdot N(\mu,\sigma^2) = z_3 N(\mu_3,\sigma_3) \tag{20}$$

Where,

$$\sigma_3 = \frac{\sigma^2}{2} \tag{21}$$

$$\mu_3 = \mu \tag{22}$$

$$z_3 = \frac{1}{\sqrt{4\pi\sigma^2}} \tag{23}$$

3.2 Putting it all together

Recall in that in Eq. (9) we found

$$\begin{aligned}\int\left(\sum_{i=1}^{K_1} w_iG_i - \sum_{i=1}^{K_2} v_iF_i\right)^2 dx = \quad & \sum_{i=1}^{K_1} {w_i}^2\int {G_i}^2dx + 2\sum_{i=1}^{K_1}\sum_{j=i+1}^{K_1} w_iw_j\int G_iG_jdx \\ & -2\sum_{i=1}^{K_1}\sum_{j=1}^{K_2} w_iv_j\int G_iF_jdx + \sum_{i=1}^{K_2} {v_i}^2\int {F_i}^2dx + \\ & 2\sum_{i=1}^{K_2}\sum_{j=i+1}^{K_2} v_iv_j\int F_iF_jdx\end{aligned} \tag{24}$$

where $G_i \sim N(\mu_{G_i}, \sigma_{G_i})$ and $F_i \sim N(\mu_{F_i}, \sigma_{F_i})$

Applying the results from the previous section to each of the terms in (24) and using the fact that $\int zN(\mu, \sigma) = z$, we obtain

$$
\begin{aligned}
\int \left(\sum_{i=1}^{K_1} w_i G_i - \sum_{i=1}^{K_2} v_i F_i \right)^2 dx = \quad & \sum_{i=1}^{K_1} {w_i}^2 \frac{1}{\sqrt{4\pi\sigma_{G_i}^2}} \\
& + 2 \sum_{i=1}^{K_1} \sum_{j=i+1}^{K_1} w_i w_j \frac{1}{\sqrt{2\pi(\sigma_{G_i}^2 + \sigma_{G_j}^2)}} \exp\left(-\frac{1}{2} \frac{(\mu_{G_i} - \mu_{G_j})^2}{(\sigma_{G_i}^2 + \sigma_{G_j}^2)} \right) \\
& - 2 \sum_{i=1}^{K_1} \sum_{j=1}^{K_2} w_i v_j \frac{1}{\sqrt{2\pi(\sigma_{G_i}^2 + \sigma_{F_j}^2)}} \exp\left(-\frac{1}{2} \frac{(\mu_{G_i} - \mu_{F_j})^2}{(\sigma_{G_i}^2 + \sigma_{F_j}^2)} \right) \quad (25) \\
& + \sum_{i=1}^{K_2} {v_i}^2 \frac{1}{\sqrt{4\pi\sigma_{F_i}^2}} + \\
& 2 \sum_{i=1}^{K_2} \sum_{j=i+1}^{K_2} v_i v_j \frac{1}{\sqrt{2\pi(\sigma_{F_i}^2 + \sigma_{F_j}^2)}} \exp\left(-\frac{1}{2} \frac{(\mu_{F_i} - \mu_{F_j})^2}{(\sigma_{F_i}^2 + \sigma_{F_j}^2)} \right)
\end{aligned}
$$

4 Product of multiple gaussians

In section 3.1.3 the product of two gaussian density functions was computed. In this section we wish to generalize such derivation to the product of multiple gaussian densities. Using the product rule theorem, we can conclude that the product of n gaussian density functions is an unnormalized gaussian. Let's explore the shape of the parameters.

$$\prod_{k=1}^{K} N_X(\mu_k, \sigma_k^2) = \tilde{z} N_X(\tilde{\mu}, \tilde{\sigma^2}) \quad (26)$$

Where,

$$\tilde{\sigma^2} = \sum_{k=1}^{K} \frac{1}{\sigma_k^2} \quad (27)$$

$$\tilde{\mu} = \tilde{\sigma^2} \left(\sum_{k=1}^{K} \frac{\mu_k}{\sigma_k^2} \right) \quad (28)$$

$$\tilde{z} = \left(\frac{2\pi\tilde{\sigma}^2}{\prod_{k=1}^{K} 2\pi\sigma_k^2} \right)^{1/2} \prod_{i<j} \exp\left(-\frac{1}{2} \frac{\tilde{\sigma}^2}{\sigma_i \sigma_j} (\mu_i - \mu_j)^2 \right) \tag{29}$$