GMMs

Wednesday, 22 March 2023 8:59 PM

Gaussian distribution by itself may not be enough in modelling real datasets

Sometimes a linear superposition of two or more Gaussians may capture data better. Such superpositions can be formulated as mixture distributions.

For a superposition of K Gaussians we have

$$p(x) = \sum_{k=1}^{K} w_k N(x \mid \mu_k, \Sigma_k)$$

Each $N(x \mid \mu_k, \Sigma_k)$ is called a component of the mixture. The parameters w_k are called mixing coefficients

We have $\sum_{k=1}^{K} w_k = 1$

Also $p(x) \geq 0$, $N(x \mid \mu_k, \Sigma_k) \geq 0$

$\Rightarrow w_k \geq 0 \;\; \forall k$

$\therefore 0 \leq w_k \leq 1$ (Can be thought of as probabilities)

$$p(x) = \sum_{k=1}^{K} p(k)\, p(x \mid k) \qquad (w_k = p(k))$$

Think of $p(k)$ as prior of picking k^{th} component and $p(x \mid k) = N(x \mid \mu_k, \Sigma_k)$

i.e. probability of x conditioned on k

The posterior prob. $p(k \mid x)$ are given as

$$\gamma_k(x) = p(k \mid x)$$

$$= \frac{p(k)\, p(x \mid k)}{\sum_l p(l)\, p(x \mid l)}$$

$$= \frac{w_k N(x \mid \mu_k, \Sigma_k)}{\sum_l w_l N(x \mid \mu_l, \Sigma_l)}$$

Parameters of GMM are w, μ and Σ.

$$W = (w_1, \ldots, w_K)$$

$$\mu = (\mu_1, \ldots, \mu_K)$$

$$\Sigma = (\Sigma_1, \ldots, \Sigma_K)$$

How to get the values?

Log of likelihood is given by

$$\ln p(X \mid w, \mu, \Sigma) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} w_k N(x_n \mid \mu_k, \Sigma_k) \right\}$$

Another Interpretation :-

z is K-dim binary random variable where at a time some z_k is 1 and all other coordinates are 0.

$\therefore z_k \in \{0, 1\}$ and $\sum_k z_k = 1$

K possible states of the vector z

Try to define joint dist. $p(x, z)$ in terms of $p(z)$ and $p(x \mid z)$

Now $p(z_k = 1) = w_k$ $\left(0 \leq w_k \leq 1 \text{ \& } \sum_k w_k = 1 \right.$

$$p(z) = \prod_{k=1}^{K} w_k^{z_k}$$

Also $p(x \mid z_k = 1) = N(x \mid \mu_k, \Sigma_k)$

$$\therefore p(x \mid z) = \prod_{k=1}^{K} N(x \mid \mu_k, \Sigma_k)^{z_k}$$

Now

$$p(x) = \sum_z p(z)\, p(x \mid z)$$

$$= \sum_{k=1}^{K} w_k N(x \mid \mu_k, \Sigma_k)$$

$\therefore$ Marginal of x is GMM

If we have observations $x_1, \ldots, x_N$ and as $p(x) = \sum_z p(x, z)$ for every observed data point x_n there is corresponding latent variable z_n

What is the advantage?

Now $\gamma(z_k)$ (responsibility) i.e. $p(z_k = 1 \mid x)$

$$= \frac{p(z_k = 1)\, p(x \mid z_k = 1)}{\sum_l p(z_l = 1)\, p(x \mid z_l = 1)}$$

$$= \frac{w_k N(x \mid \mu_k, \Sigma_k)}{\sum_l w_l N(x \mid \mu_l, \Sigma_l)}$$

Maximizing log likelihood of GMM is more complex problem than the case of a single Gaussian. Because of summation over k inside log, log fn. does not act directly on Gaussian.

We don't get closed form soln. by equating derivative to zero.