Gaussian Mixture Regression

Alexander Sepúlveda

Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones, E3T
Universidad Industrial de Santander, Bucaramanga-Colombia

This writing shows how a GMM model can be used for regression (prediction). Let be X the inputs to the model, and Y the targets. Both are random variables. For simplicity let be y an output, but the analysis can be easily extended to more variables. Lets define $f_{X,Y}(x,y)$ as the joint probability density function. The output y can be estimated by using the concept of expected value, as follows,

$$\widehat{y} = m(x) = E\big[Y \mid X = x\big] = \int y f_y(y \mid X) dy \tag{1}$$

Let's define $Z = [Y, X]$; thus $f_Z(Z) = f_{Y,X}(y,x)$, where

$$f_Z(Z) = \sum_{j=1}^{K} \pi_j \cdot \mathcal{N}(Z; \mu_j, C_j) \tag{2}$$

and, $\mathcal{N}(Z; \mu_j, C_j)$ is a joint probability density function of dimension $d+1$ with mean μ and covariance matrix C where d is the dimension of X.

$$C_j = \begin{bmatrix} C_j^{YY} & C_j^{YX} \\ C_j^{XY} & C_j^{XX} \end{bmatrix}$$

In case we have only an output, C_j^{YY} is a scalar value; C_j^{XX} is the $d \times d$ input covariance matrix; C_j^{YX} is a $1 \times d$ raw vector; and, $C_j^{XY} = (C_j^{YX})^T$ is a $d \times 1$ column vector. In addition,

$$\mu_j = [\mu_j^Y, \mu_j^{x_1}, ..., \mu_j^{x_d}] = [\mu_j^Y, (\mu_j^X)^T]^T \tag{3}$$

$$E\big[Y \mid X = x\big] = \int y \cdot f_Y(Y \mid x) \cdot dy$$

Remember that the conditional probability is defined by,

$$f_Y(Y \mid X = x) = \frac{f_{YX}(y,x)}{f_X(x)}$$

where, the joint probability $f_{YX}(y,x)$ is approximated by mixture model in (2); and, the probability of the observation x ($f_X(x)$) in the denominator is obtained by marginalization of this joint probability, giving following result,

$$f_Y(y \mid x) = \frac{\sum_{j=1}^{K} \pi_j \cdot \mathcal{N}(\overbrace{y, x}^{z}; \mu_j, C_j)}{\underbrace{\int_y \underbrace{\sum_{j=1}^{K} \pi_j \cdot \mathcal{N}(y', x; \mu_j, C_j)}_{f(y,x)} dy'}_{\text{estamos hallando } f_X(x) \text{ por marginalización de } f(y,x) \text{ respecto a y}}}$$

$$f_Y(y \mid x) = \frac{\sum_{j=1}^{K} \pi_j \cdot \mathcal{N}(y, x; \mu_j, C_j)}{\sum_{j=1}^{K} \pi_j \cdot \mathcal{N}(x; \mu_j^X, C_j^X)}$$

Once again, the definition of conditional probability ($f_{YX}(y,x) = f_X(x) \cdot f_Y(Y \mid X = x)$) is used in order to decompose each of the normal joint probability $\mathcal{N}(y, x; \mu_j, C_j)$ in numerator, obtaining following expression,

$$f_Y(y \mid x) = \frac{\sum_{j=1}^{K} \pi_j \cdot \mathcal{N}(x; \mu_j, C_j) \cdot \mathcal{N}(y \mid x; \mu_j^{Y|X}, C_j^{Y|X})}{\sum_{j=1}^{K} \pi_j \cdot \mathcal{N}(y, x; \mu_j^X, C_j^X)}$$

$$E\big[Y \mid X = x\big] = \int y \cdot f_y(y/x) \cdot dy$$

$$\begin{aligned} E\big[Y \mid X = x\big] &= \int y \cdot f_Y(Y \mid x) \cdot dy \\ &= \sum_{j=1}^{K} \frac{\pi_j \cdot \mathcal{N}(X; \mu_i^X, C_i^X)}{\sum_{j=1}^{K} \pi_j \cdot \mathcal{N}(Y, X; \mu_j^X, C_j^X)} \cdot \overbrace{\int y \cdot \mathcal{N}(Y \mid X; \mu_i^{Y|X}, C_i^{Y|X}) dy}^{m_j(X)} \end{aligned}$$

According to the book *The Multivariate Normal Distribution* by Y. L. Tong, page 34,

$$m_j(X) = \mu_j^Y + C_j^{YX} \cdot \mathrm{inv}(C_j^X) \cdot (x - \mu_j^X) \tag{4}$$

Therefore,

$$E\big[Y \mid X = x\big] = \sum_{j=1}^{K} \beta_j(x) \cdot m_j(x) \tag{5}$$

where,

$$\beta_j = \frac{\pi_j \cdot \mathcal{N}(x; \mu_j^X, C_j^X)}{\sum_{j=1}^{K} \pi_j \cdot \mathcal{N}(x; \mu_j^X, C_j^X)} \tag{6}$$

The expression β_j ($\Pr(j \mid X = x)$) are also called responsabilities (eq. 2.192 in the book *Pattern Recognition and Machine Learning* by C. Bishop).