Mixture of Gaussians [1] as A Marginalization [2]

Qi Zhao

August 11, 2018

1. Introduction

The mixture of Gaussians model can be expressed as the marginalization of a joint probability distribution [4] between the observed data x and a discrete hidden variable h that takes values h $\in \{1...K\}$ (Figure 1). If it defines in Equation 1.

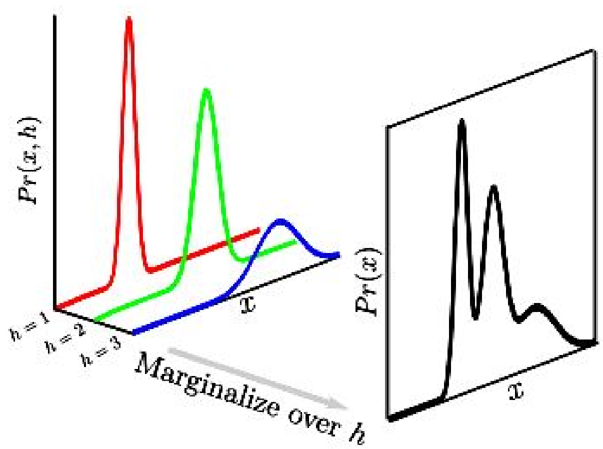

Figure 1. Mixture of Gaussians as a marginalization. The mixture of Gaussians can also be thought of in terms of a joint distribution Pr(x,h) between the observed variable x and a discrete hidden variable h. To create the mixture density it marginalizes over h. The hidden variable has a straightforward interpretation: it is the index of the constituent normal distribution [5].

$$\begin{aligned} Pr(x|h,\theta) &= Norm_x[\mu_h, \Sigma_h] \\ Pr(h,\theta) &= Cat_h[\lambda] \end{aligned} \tag{1}$$

where $\lambda = [\lambda_1 \ldots \lambda_K]$ are the parameters of the categorical distribution, then it can recover the original density using in Equation 2.

$$\begin{aligned} Pr(x|\theta) &= \sum_{k=1}^{K} Pr(x, h=k|\theta) \\ &= \sum_{k=1}^{K} Pr(x|h=k,\theta) Pr(h=k|\theta) \\ &= \sum_{k=1}^{K} \lambda_k Norm_x[\mu_k, \Sigma_k] \end{aligned} \tag{2}$$

2. Conclusions

Interpreting the model in this way also provides a method to draw samples from a mixture of Gaussians: it samples from the joint distribution Pr(x,h), and then discard the hidden variable h to leave just a data sample x. To sample from the joint distribution Pr(x,h) it first sample h from the categorical prior Pr(h) [3], then sample x from the normal distribution Pr(x|h) associated with the value of h. Notice that the hidden variable h has a clear interpretation in this procedure; it determines which of the constituent normal distributions is responsible for the observed data point x.

References

[1] S. Dasgupta. Learning mixture of Gaussians. In *IEEE Symposium on Foundations of Computer Science*, pages 634–644, 1999. 1

[2] E. G. Larsson and J. Jalden. Fixed-complexity soft MIMO detection via partial marginalization. *IEEE Transactions on Signal Processing*, 56(8):3397–3407, 2008. 1

[3] A. Piatti, M. Zaffalon, F. Trojani, and M. Hutter. Limits of learning about a categorical latent variable under prior near-ignorance. *International Journal of Approximate Reasoning*, 50(4):597–611, 2009. 1

[4] P. Stansell, J. Wolfram, and B. Linfoot. Improved joint probability distribution for ocean wave heights and periods. *Journal of Fluid Mechanics*, 503(503):273–297, 2004. 1

[5] C. M. Stein. Estimation of the mean of a multivariate normal distribution. *Annals of Statistics*, 9(6):1135–1151, 1981. 1