

1 Homework

1. Suppose that the p.d.f. of a random variable X has a 2-component mixture form:

$$p_\alpha(x) = \alpha p_1(x) + (1 - \alpha)p_2(x) \quad (1)$$

One component is the density model $p_1(x)$ and the other component is the density model $p_2(x)$. We know both $p_1(x)$ and $p_2(x)$. We do not know α . Given that $\{x_1, x_2, \dots, x_n\}$ are i.i.d. samples from the distribution of X , please give an EM algorithm for estimating α . (Describe the E-step and M-step clearly in your answer)

Hint: You may want to introduce the latent variable $Z = (z_1, z_2, \dots, z_N)$ to indicate which component “generated” each data item. $z_i \in \{1, 2\}$ for each i , and $z_i = k$ if the i^{th} sample was generated by the k^{th} mixture component. You can also introduce the similar latent variable z_i^j as that in my slides.