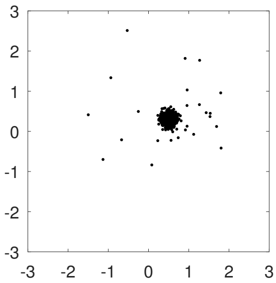
8 Machine Learning and Bayesian Inference (sbh11)

Evil Robot is shopping for Death Rays. He attends a presentation of two Death Rays, but is distracted and misses the demonstration, after which the target looks as follows:



He concludes that one is fast and accurate while the other is slower and inaccurate. To estimate the accuracy and speed of each, he models hits on the target as

$$p(\mathbf{x}|\boldsymbol{\theta}) = \pi\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \sigma_1\mathbf{I}) + (1-\pi)\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \sigma_2\mathbf{I})$$

where $\mathcal{N}$ is the normal probability density and $\boldsymbol{\theta}$ is a vector of all parameters.

(*a*) Expain why Evil Robot has chosen this model for the data. [3 marks]

(*b*) Let $\mathbf{X}$ denote the set of m hits, which are assumed independent and identically distributed. Write down an expression for the log-likelihood of $\mathbf{X}$, conditional on the parameters. [2 marks]

(*c*) Denote the two components of $p(\mathbf{x}|\boldsymbol{\theta})$ as the 'wide' and 'narrow' components. Let the ith hit have an associated indicator $\mathbf{z}_i = (z_i^{\text{narrow}}, z_i^{\text{wide}})$ where

$$z_i^x = \begin{cases} 1 & \text{if } \mathbf{x}_i \text{ in component } x \\ 0 & \text{otherwise.} \end{cases}$$

Let $\mathbf{Z}$ denote the collection of the m random variables $\mathbf{z}_i$. Show that

$$\log p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}) = \sum_{i=1}^{m} z_i^{\text{narrow}}[\log \mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}, \sigma_1\mathbf{I}) + \log \pi] + z_i^{\text{wide}}[\log \mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}, \sigma_2\mathbf{I}) + \log(1-\pi)].$$

[4 marks]

(*d*) The *EM algorithm* relies on the identity

$$\log p(\mathbf{X}|\boldsymbol{\theta}) = L(q, \boldsymbol{\theta}) + D_{\text{KL}}(q, p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}))$$

where q is an arbitrary distribution on $\mathbf{Z}$, D_{KL} is the Kullback-Leibler distance, and $L(q, \boldsymbol{\theta}) = \sum_{\mathbf{Z}} q(\mathbf{Z}) \log \frac{p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta})}{q(\mathbf{Z})}$. Explain how this gives rise to the two steps of the EM algorithm. [3 marks]

(*e*) Derive the EM update for the parameter π. [8 marks]