

Computational Statistics

Chapter 3: EM algorithm

1. Let w_1, \dots, w_n be an i.i.d. sample from a mixture of a normal distribution $\mathcal{N}(\mu, \sigma)$ and a uniform distribution $\mathcal{U}([-a, a])$, with pdf

$$g(w; \theta) = \pi \phi(w; \mu, \sigma) + (1 - \pi)c, \quad (1)$$

where $\phi(\cdot; \mu, \sigma)$ is the normal pdf, $c = (2a)^{-1}$, π is the proportion of the normal distribution in the mixture and $\theta = (\mu, \sigma, \pi)^T$ is the vector of parameters. Typically, the uniform distribution corresponds to outliers in the data. The proportion of outliers in the population is then $1 - \pi$. We want to estimate parameter θ using the EM algorithm

- (a) Using the function `rnorm` and `runif`, generate a sample of size $n = 100$. Draw a box plot of the data.
- (b) Write an EM algorithm for this problem, and apply it to the data, with different initializations.
- (c) Draw the estimated probabilities $1 - z_i^{(t)}$ of being an outlier, as a function of w_i . Does it make sense?
- (d) Compare the estimates with those computed using the `optim` function.