APMLA: assignement 1

Caterina De Bacco

Practicing EM and Gibbs sampling

Consider the problem of observing a set of heights and knowing that they represent two groups. You want to predict which are the members of the two groups based on this information. We will use a Bayesian mixture of 2 Gaussians for making this prediction.

Specifically, we assume the priors:

$$\mu_0, \mu_1 \sim \mathcal{N}(m, s^2) \quad , \tag{1}$$

$$\pi \sim Beta(a, b) \quad , \tag{2}$$

$$z_1, \dots, z_N \sim \text{Bern}(\pi) \quad . \tag{3}$$

Assume common fixed σ^2 for the variance of the mixture components.
We have that the height x_i of sample i, given all the parameters, is distributed as

$$x_i \sim \mathcal{N}(\mu_{z_i}, \sigma^2) \quad . \tag{4}$$

1 EM (MLE)

Consider an MLE approach, i.e. forget about the priors on the parameters μ and π.

(a) Write the expression for $\gamma_0(x_n)$, the posterior of $z_n = 0$ (i.e. the posterior of sample n beloniging to group 0).

(b) Write the EM updates for μ_0, μ_1 and π_0, π_1.

(c) Implement the EM on a Python Jupyter notebook.

(d) Print the final values of μ_0, μ_1 and $\pi = \pi_1/(\pi_0 + \pi_1)$.

(e) Plot the log-likelihood w.r.t. iteration time.

(f) Visualize the allocation of the cluster assignments $z_1, \dots, z_N$ in an effective way (you are free to choose how to show this result). The plot should convey the information about how the two clusters of points look like.

(g) How would the results change if you had considered the priors on the parameters μ and π?

2 Gibbs sampling

(a) Write the expression for $P(x_i|\mu, \pi)$ (i.e. without a dependence on z).

(b) Derive the conditionals $P(\pi|\mu, z, x)$, $P(\mu|\pi, z, x)$ and $P(z|\pi, \mu, x)$, needed for Gibbs sampling.

(c) Implement a Gibbs sampler in Python for this model.
Set $\sigma = 5cm$, $a = b = 1$ (for the Beta prior), $m = 175cm$, $s = 10cm$ (prior of the components' means).
Initialize $\pi = 0.5$, $\mu_0, \mu_1 = m$ and $z_1, \ldots, z_N$ randomly extracted i.i.d. from a Bernoulli(0.5).

(d) Plot the log-likelihood w.r.t. iteration time.

(e) Plot μ_0, μ_1 in the same plot as a function of iteration time.

(f) Plot π as a function of iteration time.

(g) Visualize the allocation of the cluster assignments $z_1, \ldots, z_N$ in an effective way (you are free to choose how to show this result). Similar for the analogous question for EM above.

3 General questions

(a) The true number of data in group 0 is $N_0 = 695$ and $N_1 = 562$. How many data points do the two approaches estimate in each of the groups?

(b) If you notice any difference between the ground truth and the estimated values, how could the difference decrease? (e.g. what would you change in the settings of the two algorithms?).

(c) Comment on the differences (if any) of the results between EM and Gibbs.