

Assignment Topic №3: EM algorithm

Dmytro Marushkevych

September 6, 2022

The third assignment topic is optimization via the EM algorithm. If you draw the topic “EM algorithm” at the oral exam, you will have to present a solution of one of the two assignments below.

Remember the five points:

- How can you test that your implementation is correct?
- Can you implement alternative solutions?
- Can the code be restructured e.g. by modularization, abstraction or object oriented programming to improve readability?
- How does the implementation perform (benchmarking)?
- Where are the bottlenecks (profiling), and what can you do about them?

As for the other assignments, performance tests are important, but remember to make correct comparisons. That is, if you benchmark two implementations against each other, use exactly the same convergence criterion. Alternatively, investigate how the algorithms converge as a function of *real* time (not iterations).

Test of convergence and experiments with convergence criteria are important to investigate. Robustness towards the initial choice of parameters is important, and testing of the implementation on several (simulated) data sets is a good idea.

Assignment 1: The EM-algorithm for the t -distribution

Let $Y = (X, W) \in \mathbb{R} \times (0, \infty)$ have joint density

$$f(x, w) = \frac{1}{\sqrt{\pi\nu\sigma^2}2^{(\nu+1)/2}\Gamma(\nu/2)} w^{\frac{\nu-1}{2}} e^{-\frac{w}{2}\left(1+\frac{(x-\mu)^2}{\nu\sigma^2}\right)}.$$

Then the marginal density of X is the t -distribution (why?) with location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma > 0$ and shape parameter $\nu > 0$. You can initially regard ν as fixed. Maximize the complete data log-likelihood

$$\sum_{i=1}^n \log f(y_i, w_i)$$

for i.i.d. observations over (μ, σ^2) and implement this as a function. Then implement the EM algorithm for estimating (μ, σ^2) , compare it to other optimization algorithms based on the marginal log-likelihood, and investigate how to compute the Fisher information. You can consider generalizing to estimating the shape parameter ν as well, but then the M-step cannot be computed analytically and you need the digamma function.

Remark: We can think of the joint density of Y as generating first W from a χ^2_ν -distribution, and then generating X conditionally on $W = w$ from a $\mathcal{N}(\mu, \nu\sigma^2/w)$ -distribution. For the E-step observe that the density of $W \mid X = x$ is $\propto f(x, w)$ with x fixed and recognize this as a Γ -distribution.

Assignment 2: Mixtures of t -distributions

The t -distribution with location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma > 0$ and shape parameter $\nu > 0$ has density

$$f(x \mid \mu, \sigma^2, \nu) = \frac{\Gamma((\nu + 1)/2)}{\sqrt{\pi\nu\sigma^2}\Gamma(\nu/2)} \left(1 + \frac{(x - \mu)^2}{\nu\sigma^2}\right)^{-(\nu+1)/2}.$$

In this assignment you will consider iid observations x_1, \dots, x_n from the two-component mixture of t -distributions, that is, the density is

$$pf(x \mid \mu_1, \sigma_1^2, \nu_1) + (1 - p)f(x \mid \mu_2, \sigma_2^2, \nu_2)$$

for the five parameters $p \in (0, 1)$, $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$. The shape parameters $\nu_1, \nu_2 > 0$ can be fixed.

Implement the Q -function and its gradient, and implement a generalized EM algorithm for estimating the parameters of the two-component mixture of t -distributions. Compare the EM algorithm with e.g. gradient descent or another optimization algorithm. Investigate also how to implement a computation of the Fisher information.

Use simulated data to test your implementations. It may be of interest to compare estimates of location and scale parameters with estimates from a two-component Gaussian mixture, in particular if you contaminate the sample with some outliers.