Mathematical Underpinnings Lab 9 08.05.2024

Task 1 (EM)

a) Generate $n = 500$ observations from the following mixture distribution:

$$0.3 * \mathcal{N}(1, 1) + 0.7 * \mathcal{N}(5, 2).$$

True parameter values: $\mu_0 = 1$, $\mu_1 = 6$, $\sigma_0^2 = 1$, $\sigma_1^2 = 2$, $\pi_0 = 0.3$, $\pi_1 = 0.7$. We will denote the set of al the parameters as θ - $\theta = (\mu_0, \mu_1, \sigma_0^2, \sigma_1^2, \pi_0, \pi_1)$ (there is a slight overparametrization as $\pi_1 = 1 - \pi_0$, but it doesn't matter). Draw a histogram of a sample and the true distribution.

b) Assume that the mixing variable G is known (meaning that $G = 0$ if the observation comes from the first normal distribution and $G = 1$ if it comes from the second). Write the formula of the loglikelihood: $\sum_{i=1}^{n} \log(p(x_i, G_i = g_i|\theta))$, $g = 0, 1$.

c) (E-step computations) Compute the expected values of not observed variables. (A hint: we don't observe G, so we need to compute $\gamma_i := \mathbb{E}G_i|X_i, \theta)$.

d) (M-step computations) Now we want to find the update equations for all the parameters θ:

- μ_0 (and analogously μ_1),
- σ_0^2 (and analogously σ_1^2),
- π_0 (and analogously π_1).

e) Implement EM algorithm:

- Draw initial values for parameters: $\hat{\mu}_0$, $\hat{\mu}_0$, $\hat{\sigma}_0$, $\hat{\sigma}_1$, fix $\hat{\pi}_0 = \hat{\pi}_1 = 1/2$.
- For each observation, calculate weights γ_i (E-step). In the formula obtained in step c), plug-in all current values of the parameters θ.
- Calculate new values of the parameters $\hat{\mu}_0$, $\hat{\mu}_0$, $\hat{\sigma}_0$, $\hat{\sigma}_1$, $\hat{\pi}_0$, $\hat{\pi}_1$ using the formula from d), using previously calculated weights γ_i instead of G_i.
- Repeat the last two steps (2-3) until the algorithm converges (you can set any stopping criterion based on loglikelihood, the fact that parameter estimators in successive steps will not differ much from each other, or simply set a maximum number of steps).

f) On one plot, draw the histogram of the generated sample, the true density, and the density estimated through the procedure described in steps 1-4.

g) Do the task for the mixture of three (or more) gaussian distributions.

Task 2* (MM)

a) Write down the formula for log-likelihood in logistic regression.

b) Compute Hessian of minus log-likelihood. How you can upper bound it? (In terms of matrices, we say that $H \geq M$ if $H - M$ is nonnegative semi-define).

c) (Majorization step) Write down the upper bound on minus log-likelihood.

d) (minimization step) Write down the formula for updates of the parameters (based on the upper bound).

e) What are the advantages and disadvantages of using MM compared to IWLS (iterative weighted least squares/Newton-Raphson)?

f) Implement MM algorithm for logistic regression and compare it with a sklearn function.