Pattern Recognition Homework 2

İpek Erdoğan

January 5, 2021

Expectation-Maximization Algorithm

Expectation-Maximization [1] is a clustering algorithm which has an iterative approach. It basically computes maximum likelihood estimates iteratively and updates the distribution parameters (π, Σ, μ) according to this likelihood information. This update can be implemented according to the log-likelihood and log-posterior [2].

In this homework, we implemented this solution to cluster a dataset which is a Gaussian Mixture model, includes 3 different Gaussian distribution.

We are trying to find best parameters theta (π, Σ, μ) to maximize the log-likelihood of each data to appropriate distribution (here we can think distribution as clusters) [3]. There is a paradox here. Since we don't know the distributions from the beginning, we can not calculate the likelihood. Since we can not calculate the likelihood, we can not maximize it and find the appropriate Gaussian distributions of dataset.

To break this loop and start calculation, we determine random inital parameters for 3 Gaussian distributions: $(\pi_0, \pi_1, \pi_2, \Sigma_0, \Sigma_1, \Sigma_2, \mu_0, \mu_1, \mu_2)$

Expectation

In this part, by using the parameters, we calculate the likelihoods of all the data points according to the different distributions. Than by using the formula down below, we calculate a posterior probability (sometimes people refer this responsibility) [4] for each of the data points and for each of the distributions. We will use this probabilities to update the parameters in maximization step.

$$\gamma(z_n k) = \frac{\pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n|\mu_j, \Sigma_j)}$$

Maximization

By using the probabilities come from the expectation step, we update the π, μ and Σ parameters in maximization step, to increase the each point's likelihood to appropriate Gaussian distribution. Which also means to devide data points into clusters in most correct way.

To update the π_k values, we sum the probabilities which came for the kth distribution, and divide it by the total data point count.

$$\pi_k = \frac{N_k}{N}$$

To update the μ_k values, we multiply the datapoints with their probabilities and sum. Then, we divide this sum to the sum of the probabilities.

$$\mu_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) x_n$$

To update the Σ values, we first calculate the differences between the data points and the mean, and then multiply this difference, the transpose of this difference and the probabilities of these data points. And we divide this value to the sum of the probabilities.

$$\Sigma_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk})(x_n - \mu_k)(x_n - \mu_k)^{\mathsf{T}}$$

We save the average likelihood of all points for all the distributions. As a convergence condition, we check the difference between this step's value and previous step's value. If the difference is smaller than "0.0000001", it means algorithm converged and code stops.

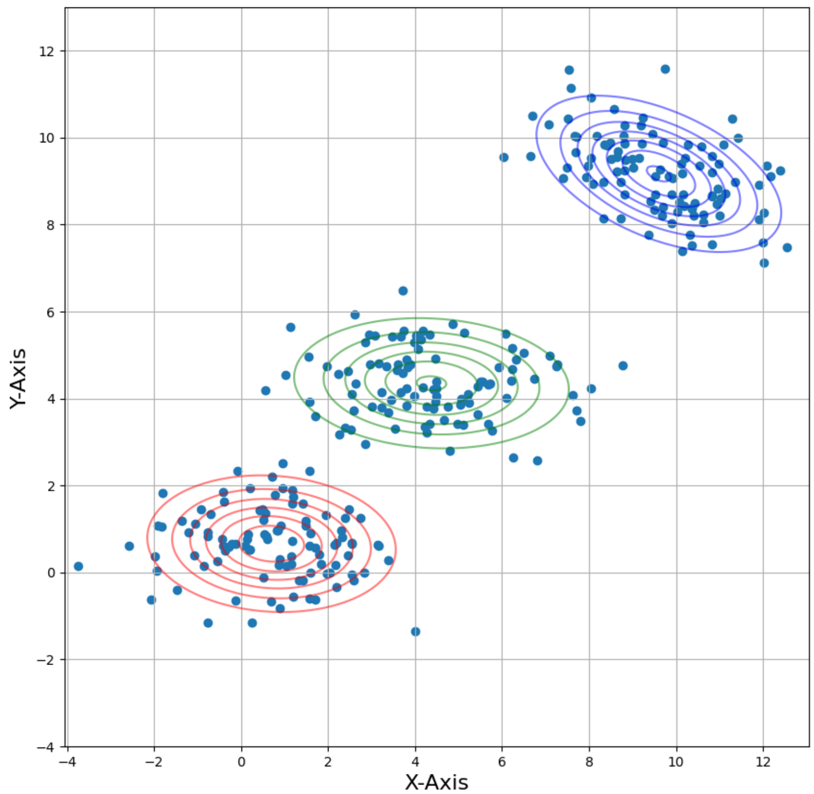

Figure 1: Result of EM algorithm

You can see the visualization of my solution in Figure 1. I found this solutions with 47 steps. The parameters I have found was:

$\pi_0 = 0.33350133224135536$

$\mu_0 = [4.37904703\ 4.35183928]$

$$\Sigma_0 = \begin{bmatrix} 2.74789528 & -0.1192322 \\ -0.1192322 & 0.61806456 \end{bmatrix}$$

$\pi_1 = 0.3331653516861432$

$\mu_1 = [0.70208295\ 0.6613623]$

$$\Sigma_1 = \begin{bmatrix} 2.11797378 & -0.09973243 \\ -0.09973243 & 0.64083765 \end{bmatrix}$$

$\pi_2 = 0.33333331607250144$

$\mu_2 = [9.60515914\ 9.16835945]$

$$\Sigma_2 = \begin{bmatrix} 2.01245136 & -0.64166751 \\ -0.64166751 & 0.82171149 \end{bmatrix}$$

Additional Notes

The formulas are taken from the Pattern Recognition Homework 2 explanation file.

In the script, I inspired by the blog here: https://towardsdatascience.com/implement-expectation-maximization-em-algorithm-in-python-from-scratch-f1278d1b9137

For the plotting part, I followed the same approach with this blog: https://medium.com/@prateek.shubham.94/expectation-maximization-algorithm-7a4d1b65ca55

References

[1] Dempster, A. P., N. M. Laird and D. B. Rubin, "Maximum Likelihood from Incomplete Data Via the EM Algorithm", *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 39, No. 1, pp. 1–22, 1977, `https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1977.tb01600.x`.

[2] Dellaert, F., "The Expectation Maximization Algorithm", , 07 2003.

[3] Moon, T. K., "The expectation-maximization algorithm", *IEEE Signal Processing Magazine*, Vol. 13, No. 6, pp. 47–60, Nov 1996.

[4] Gebru, I. D., X. Alameda-Pineda, F. Forbes and R. Horaud, "EM Algorithms for Weighted-Data Clustering with Application to Audio-Visual Scene Analysis", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 38, No. 12, pp. 2402–2415, Dec 2016.