

Model Selection on Mixture Models Using the Akaike Information Criterion.

Nabil Akbarazzima Fatih
Mat-Nr: 3271528

Abstract

The Expectation–Maximization (EM) algorithm is a popular tool in a wide variety of statistical settings, in particular in the maximum likelihood estimation of parameters when clustering using mixture models [2]. But how to select a model on mixture models using the Akaike Information Criterion and to improve the results of the vanilla EM algorithm for fitting mixture models to the data? We will take a look on model selection according to the Akaike Information Criterion and the strategies to improve the results of the vanilla EM algorithm. The experiment with Expectation Maximization (EM) Algorithm is performed in the following two cases: the responsibilities parameters are initialized randomly and with K-means. The result shows that the Expectation Maximization (EM) Algorithm initialized with K-means performs better than the one initialized randomly.

1 Introduction

While maximum likelihood estimation can find the “best fit” model for a set of data, it doesn’t work particularly well for incomplete data sets. The more complex EM algorithm can find model parameters even if you have missing data [1]. This paper presents how to perform model selection on mixture models using the Akaike Information Criterion, as well as the strategies to improve the results of the vanilla EM algorithm for fitting mixture models to the data. We implement the EM-algorithm for estimating the parameters of a mixture of k normal distributions $\mathcal{N}(x; \mu_j, \sigma_j)$ with means $\mu_j \in \mathbb{R}$ and standard deviations $\sigma_j \in \mathbb{R}$. We present the mathematical background of the Expectation Maximization and the Akaike Information Criterion. We underlay our results with plots.

2 Main analysis

2.1 Data

In this paper we use three data sets: iris, blood pressure, and generated synthetic data set.

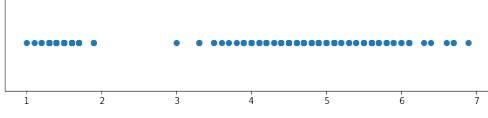


Figure 1: Petal lengths of iris data set

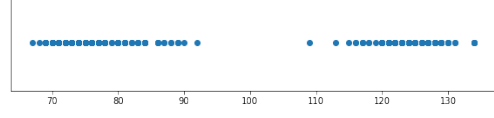


Figure 2: Blood pressure values of blood pressure data set

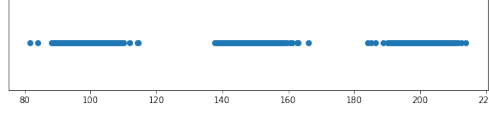


Figure 3: n data points randomly sampled from specified mixture of generated synthetic data set.

2.2 Background

First of all, we take a look at the needed mathematical background. In Expectation Maximization (EM) Algorithm for Gaussian Mixtures, we need to choose number of k components and initialize parameter θ of mixture $f(x; \theta)$, then we need to repeat the following steps until termination:

Step 1: Expectation (E): estimate responsibilities $\gamma(z_{ij})$ using current θ

$$\gamma(z_{ij}) = \frac{\pi_j N(x_i; \mu_j, \Sigma_j)}{\sum_{l=1}^k \pi_l N(x_i; \mu_l, \Sigma_l)} \quad (1)$$

$$n_j = \sum_i^n \gamma(z_{ij}) \quad (2)$$

Step 2: Maximization (M): re-estimate θ using current $\gamma(z_{ij})$

$$\mu_j = \frac{1}{n_j} \sum_{i=1}^n \gamma(z_{ij}) x_i \quad (3)$$

$$\Sigma_j = \frac{1}{n_j} \sum_{i=1}^n \gamma(z_{ij}) (x_i - \mu_j)(x_i - \mu_j)^T \quad (4)$$

$$\pi_j = \frac{n_j}{n} \quad (5)$$

We have the following equation for model selection according to Akaike Information Criterion (AIC):

$$2|\hat{\theta}| - 2\mathcal{L}_{\mathcal{D}}(\hat{\theta}) \quad (6)$$

Where $\mathcal{L}_{\mathcal{D}}(\hat{\theta})$ is the log-likelihood of estimate $\hat{\theta}$ and $|\hat{\theta}|$ is the number of free parameters to estimate. A mixture of two normal distributions has 6 parameters

$$\theta = (\pi_1, \pi_2, \mu_1, \mu_2, \gamma_1, \gamma_2)$$

but only 5 free parameters. The reason is that we can infer π_2 from π_1 via $\pi_2 = 1 - \pi_1$. Thus, we only need to estimate $k - 1$ instead of k mixing coefficients. The AIC uses this information when counting the number $|\hat{\theta}|$ of parameters to be estimated. For model selection, we choose a model with the smallest AIC value.

2.3 Experiments

We implemented the Expectation Maximization (EM) Algorithm for each of three data sets and for each $k \in \{1, \dots, 15\}$ fit a model with k components to the data. Then, Select the best model according to the Akaike Information Criterion. For default setting (Vanilla), the responsibilities (The method used to initialize the weights, the means and the precisions) are initialized randomly. The results of Vanilla Expectation Maximization (EM) Algorithm can be seen in the following plots:

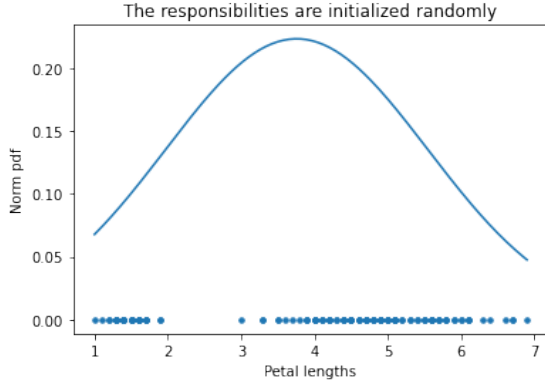


Figure 4: AIC value: 599.10 (Iris data set)

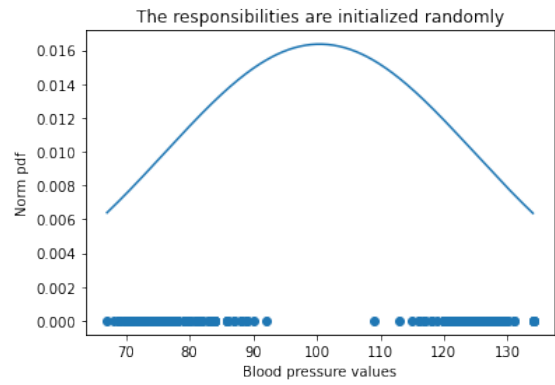


Figure 5: AIC value: 2790.41 (Blood pressure data set)

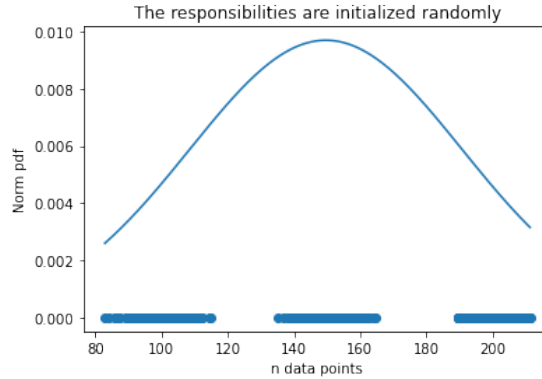


Figure 6: AIC value: 10276.95 (Generated synthetic data set)

Considering that in the results of Vanilla Expectation Maximization (EM) Algorithm if we initialized the responsibilities randomly, the performance of Expectation Maximization (EM) Algorithm looks poorly. So how can we improve the results of Expectation Maximization (EM) Algorithm? We changed the responsibilities parameter and initialized with K-means. We implement with the same steps like before for each of three data sets and for each $k \in \{1, \dots, 15\}$ fit a model with k components to the data, then select the best model according to the Akaike Information Criterion. The results of Expectation Maximization (EM) Algorithm after changing the responsibilities parameters can be seen in the following plots:

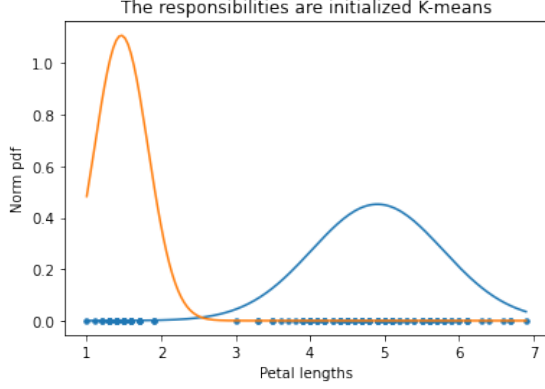


Figure 7: AIC value: 447.40 (Iris data set)

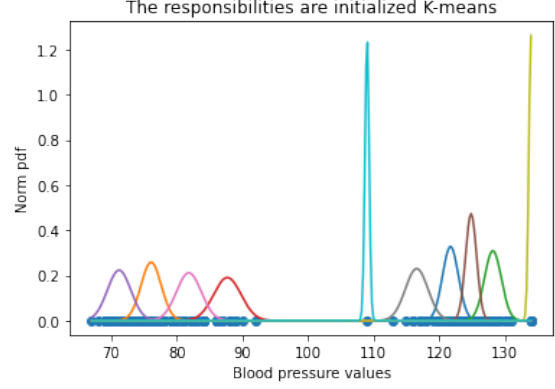


Figure 8: AIC value: 2173.55 (Blood pressure data set)

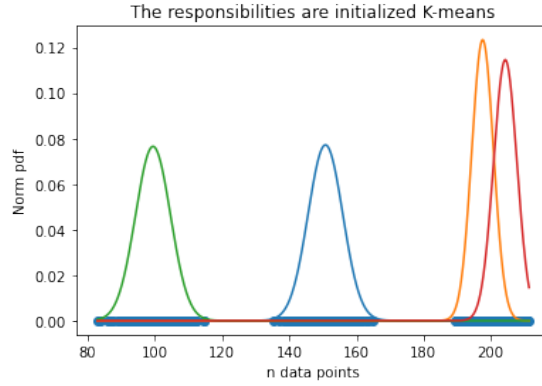


Figure 9: AIC value: 8301.20 (Generated synthetic data set)

From the plots, we can conclude that Expectation Maximization (EM) Algorithm performs well after changing the responsibilities parameters to K-means. We found that Expectation Maximization (EM) Algorithm initialized with K-means improve the Expectation Maximization (EM) Algorithm's performance.

3 Conclusion

In our experiments we implemented the Expectation Maximization (EM) Algorithm and did a model selection on mixture models using the Akaike Information Criterion. The experiment results show that the Expectation Maximization (EM) Algorithm is sensitive to the responsibilities parameters. The Expectation Maximization (EM) Algorithm performs better with K-means than randomly initializing the responsibilities parameters. So in the future, we can consider using K-means for the responsibilities parameters for the strategies to improve the results of Expectation Maximization (EM) Algorithm.

References

- [1] Stephanie Glen. “EM Algorithm (Expectation-maximization): Simple Definition”. In: *StatisticsHowTo.com* (2015). URL: <https://www.statisticshowto.com/em-algorithm-expectation-maximization/>.
- [2] Adrian O’Hagan, Thomas Brendan Murphy, and Isobel Claire Gormley. “Computational aspects of fitting mixture models via the expectation–maximization algorithm”. In: *Computational Statistics Data Analysis* 56.12 (2012), pp. 3843–3864. ISSN: 0167-9473. DOI: <https://doi.org/10.1016/j.csda.2012.05.011>. URL: <https://www.sciencedirect.com/science/article/pii/S0167947312002101>.