International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Expected Maximization Algorithm for Classification

Sripriya Muppidi[1], Bhatracharyulu N.Ch.[2]

Department of Statistics, University College of Science, Osmania University, Hyderabad-7, TS, India
muppidisripriya[at]gmail.com
dwarakbhat[at]osmania.ac.in

Abstract: ***In this paper an attempt is made to propose Expected Maximization algorithm for estimating the object belongs to the label class. The comparison made is with respect to their methods, merits, and demerits. The methods were implemented for a credit card bank data set and evaluated its accuracy.***

Keywords: Machine learning techniques, Credit Card, supervised learning

1. Introduction

Dempster, et al [1] established important fundamental properties of the algorithm. It is maximizing pastiche estimates of parameters based on the observed sample. It is an iterative process and is numerically stable, each iteration increasing the likelihood i.e. it is slow and, linearly converging. No evaluation of the likelihood nor its derivatives is involved and requires small storage space. Its simplified maximization step appears to be an advantage over other methods in some respects; this has proven to be a major failing because the algorithm does not seem to provide an estimate of the information matrix. Such derivative-based algorithms as quadratic optimization methods generally compute matrices that approximate the Hessian of the log-likelihood function, and thereby provide estimators of the information matrix as a by-product. McLachlan and Krishnan [2] used an iterative method for forming the (normal theory-based) linear discriminant function from partially classified training data.

The EM algorithm is typically easily implemented in two steps, the E-step of each iteration only involves taking expectations over complete data conditional distributions and the M-step of each iteration only requires complete data ML-estimation, which is often in simple closed form.

2. Expected Maximization Algorithm

Let $\underline{x} = (x_1, x_2, x_3, \ldots, x_n)$ be the observed sample drawn from a population with probability density function P(x, θ), where the value of the parameter is unknown. Let L($\underline{x}$,θ) be the likelihood function of observed sample and the log of the likelihood function be log L($\underline{x}$,θ). Start with some initial guess to evaluate the parameters and evaluate the Expected value of log Likelihood function. Evaluate the improved version of the parameter that maximizes the expected value of log of Likelihood function. Repeat the procedure until the values for the parameter are stabilizes.

The improved version of the parameter that maximizes the expected value of log of Likelihood function. Repeat the procedure until the values for the parameter are stabilizes. Maximizing the log probability of data is not tractable. There is no closed form solution to log L($\underline{\mathbf{x}}$). Using bayes theorem,

$$p(x) = \frac{p(x,z)}{p\left(\frac{z}{x}\right)} \Rightarrow \log p(x) = \log p(x,z) - \log p(z/x)$$

$$\Rightarrow \log p(x) = \log p(x,z) - \log p(z/x) - \log q(z) + \log q(z)$$

$$\Rightarrow \log p(x) = \log\left\{\frac{p(x,z)}{q(z)}\right\} + \left(\log\left\{\frac{p(z|x)}{q(z)}\right\}\right)$$

Multiply with q(z) on both sides

$$q(z).\log p(x) = q(z).\log\left\{\frac{p(x,z)}{q(z)}\right\} - \left(q(z)\log\frac{p(z/x)}{q(z)}\right)$$

$$\Rightarrow \int q(z)\log p(x)\, dz = \int q(z)\log\left\{\frac{p(x,z)}{q(z)}\right\}dz - \int\left(q(z)\log\frac{p(z/x)}{q(z)}\right)dz$$

$$\Rightarrow \log p(x) = \log\left\{\frac{p(x,z)}{q(z)}\right\}\int q(z)\log\left\{\frac{p(x,z)}{q(z)}\right\}dz - qzlogp(z|x)q(z)dz$$

$$\Rightarrow \log p(x) = F(q,\theta) + KL(p|q) \text{ where } F(q,\theta) = \int q(z)\frac{ln\, p(x,z)}{q(z)}dz$$

E-step: $q(z) \cong p(z|x)$ and **M-step:** $Max_\theta F(q,\theta)$

Algorithm:

Step 1: Consider a Statistical Model with a set of observed random sample $\underline{\mathbf{x}}$ that satisfies the model, that is let $\underline{\mathbf{x}} = (\mathbf{x_1, x_2, x_3, \ldots, x_n})$ be the observed sample drawn from a population with probability density function P(x, θ), where the value of the parameter θis unknown.

Step 2: Evaluate the Likelihood Function L($\underline{x}$,θ) of the observed sample and the log of Likelihood Function log L($\underline{x}$,θ).

Step 3: Choose some initial value for the parameter in the model and evaluate the Expected value of Log Likelihood function. Start with some initial guess to evaluate the parameters.

Step 4: Evaluate the improved version of the parameter that maximizes the expected value of Log of Likelihood function and evaluate the Expected value of log Likelihood function.

Step 5: Repeat the steps 3 & 4 until two successive iterates gives same value.

The algorithm is implemented for the credit card data and is illustrated in the following example.

Example: A real data set of 13,444 customers of a national bank with a feature vector $\underline{\mathbf{X}}$= (X_1, X_2, … X_{12}) of 12 attributes under study with a categorical response variable Y Sample Data Set of the status of credit-card is considered. For the classification, 80% (10,755) of the data is used for training the model and 20% (2,689) of the data set is used for testing the model and its accuracy. The variables / attributes under study are: Y: Credit card status; X_1: Age; X_2: Months living at current address; X_3 :1+No of dependents; X_4: No. of Major Derogatory reports; X_5: No. of Minor Derogatory reports ; X_6: Own-rent; X_7: Income; X_8: Self-employed; X_9: Income divided by number of dependents; X_{10}: Ratio of monthly credit card expenditure to yearly income; X_{11}:Average monthly credit card expenditure; X_{12}: Log of Spending.

Sample data set of Response (Y) & Feature vector (X) with data values												
Y	**X_1**	**X_2**	**X_3**	**X_4**	**X_5**	**X_6**	**X_7**	**X_8**	**X_9**	**X_{10}**	**X_{11}**	**X_{12}**
…	…	…	…	…	…	…	…	…	…	…	…	…
1	30.75	36	2	0	0	1	3166.667	0	12666.667	0.0252295	79.893336	4.3806924
1	36.75	14	1	0	0	0	1833.333	0	12500	0.0753976	138.228926	4.9289112
1	27.66667	16	1	0	0	0	1650	0	23900	0.0953847	157.384808	5.0586938
0	33.75	18	0	0	0	1	1833.333	1	31000	3.87E-04	17.5041667	3.7810954
1	25.91667	54	0	1	1	1	1918	0	23016	0.169264	324.64833	5.7827425
1	22.33333	2	0	0	0	0	2383.333	0	28600	0.0309584	73.7841625	4.3011441
1	45.16667	42	1	0	0	1	4000	0	26500	0.0561251	224.500373	5.4138774
1	38.58333	24	1	0	0	1	5000	0	30000	0.0202707	101.353331	4.6186127
0	42.41667	2	3	0	0	0	2916.667	0	8750	3.43E-04	69.3808347	4.6803215
1	43.25	118	4	0	1	1	3333.333	0	9000	2.67E-04	0.8888889	-0.117783
…	…	…	…	…	…	…	…	…	…	…	…	…

The results after implementation of EM algorithmusing R-program are given below:

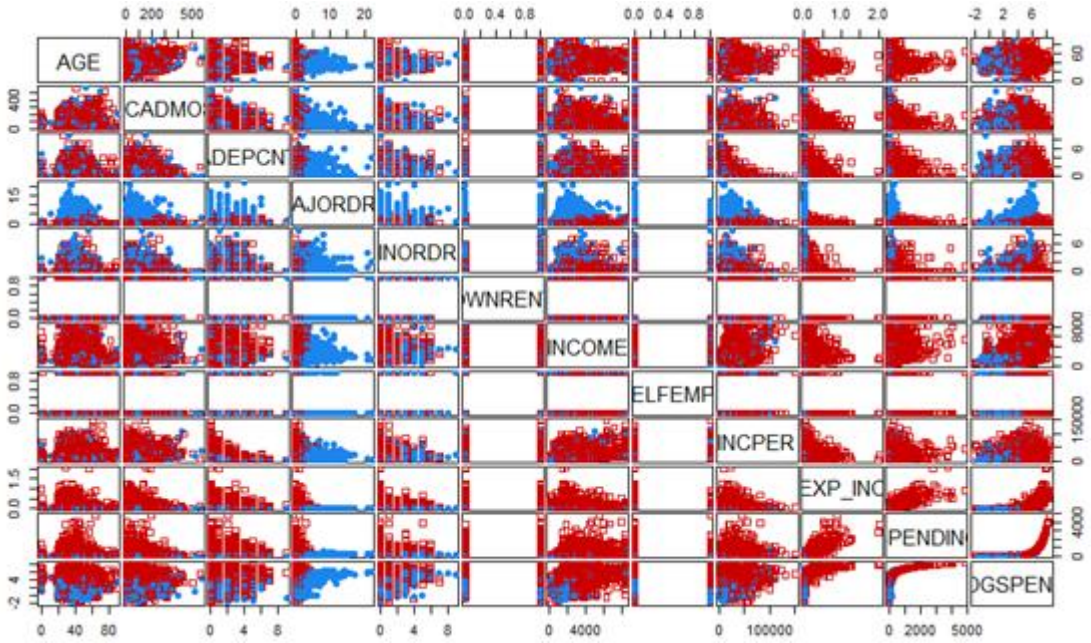

Figure 1: Correlation among variables

The Bayesian information criterion (BIC) is a criterion for model selection among a finite set of models; the model with the lowest BIC is preferred. It is based on the likelihood function.

Best BIC values			
	VEV, 2	EEV, 2	EEE, 1
BIC	-190844.5	-211314.44	-214052.9
BIC diff	0.0	-20469.97	-23208.4
Best ICL values			
	EEV, 5	VEV, 2	EEE, 1
ICL	-147982.6	-200371.8	-214052.87
ICL diff	0.0	-52389.2	-66070.25

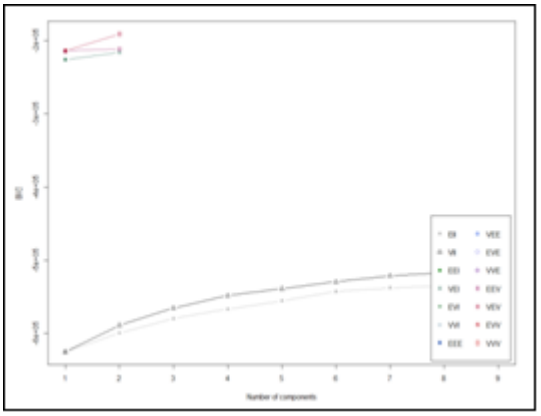

Figure 2: BIC values plot

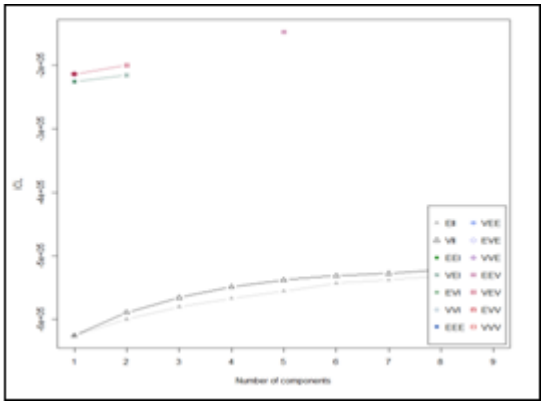

Figure 3: ICL values plot

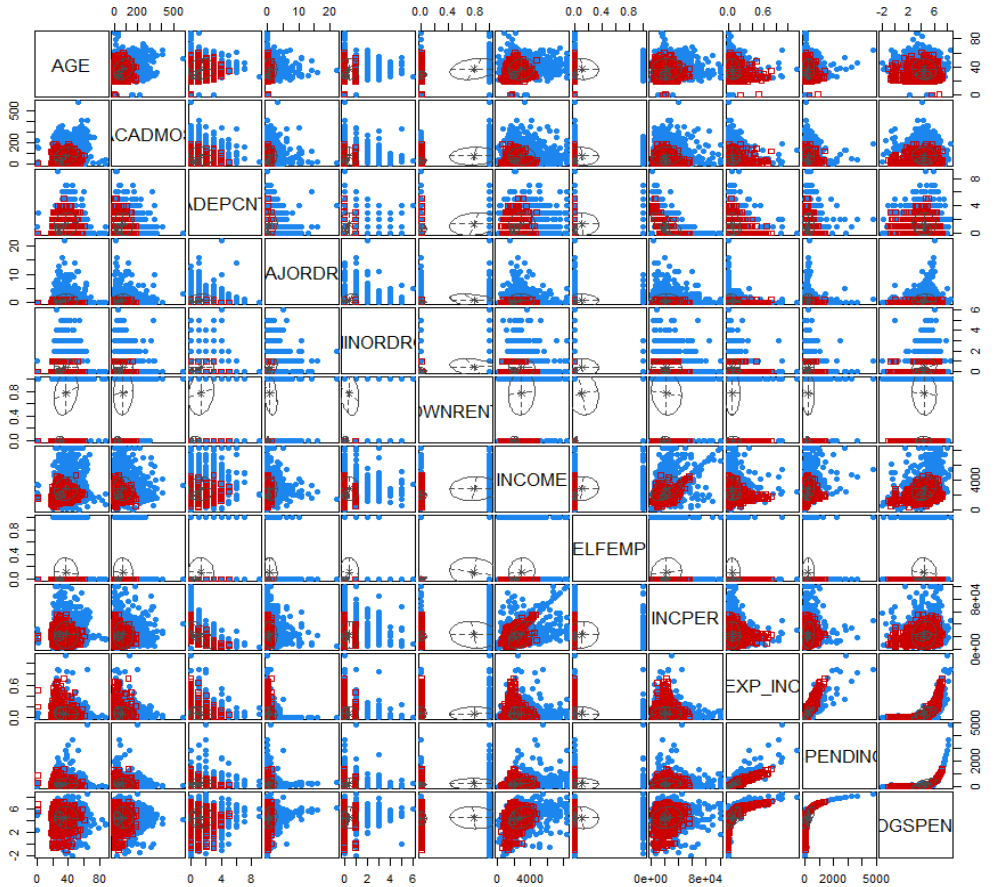

Figure 4: Classification by EM Algorithm

Remarks:

1) The EM-algorithm converge slowly and in problems where there is too much 'incomplete information'.
2) It is just like the Newton-type methods does not guarantee convergence to the global maximum when there are multiple maxima. Further, in this case, the estimate obtained upon the initial value.
3) In some problems, the E-step may be analytically intractable, although in such situations there is the possibility of effecting it via a Monte Carlo approach.
4) Unlike the Fisher's scoring method, it does not have an inbuilt procedure for producing an estimate of the covariance matrix of the parameter estimates

References

[1] Dempster A.P., Laird. N.M. and Rubin D.B. (1977): "Maximum Likelihood from Incomplete Data via the EM Algorithm, Journal of the Royal Statistical Society", Series B, Vol. 39, pp 1-38.
[2] McLachlan G.J. and Krishnan T. (2008): "The EM Algorithms and Extensions", Wiley Inter-science Publications, Second Edition.
[3] Park T. (1993): "A comparison of the generalized estimating equation approach with the maximum likelihood approach for repeated measurements", Statistics in Medicine, Vol. 12(18), pp 1723-1732.