DATA SCIENCE LAB

EM Algorithm for GMM

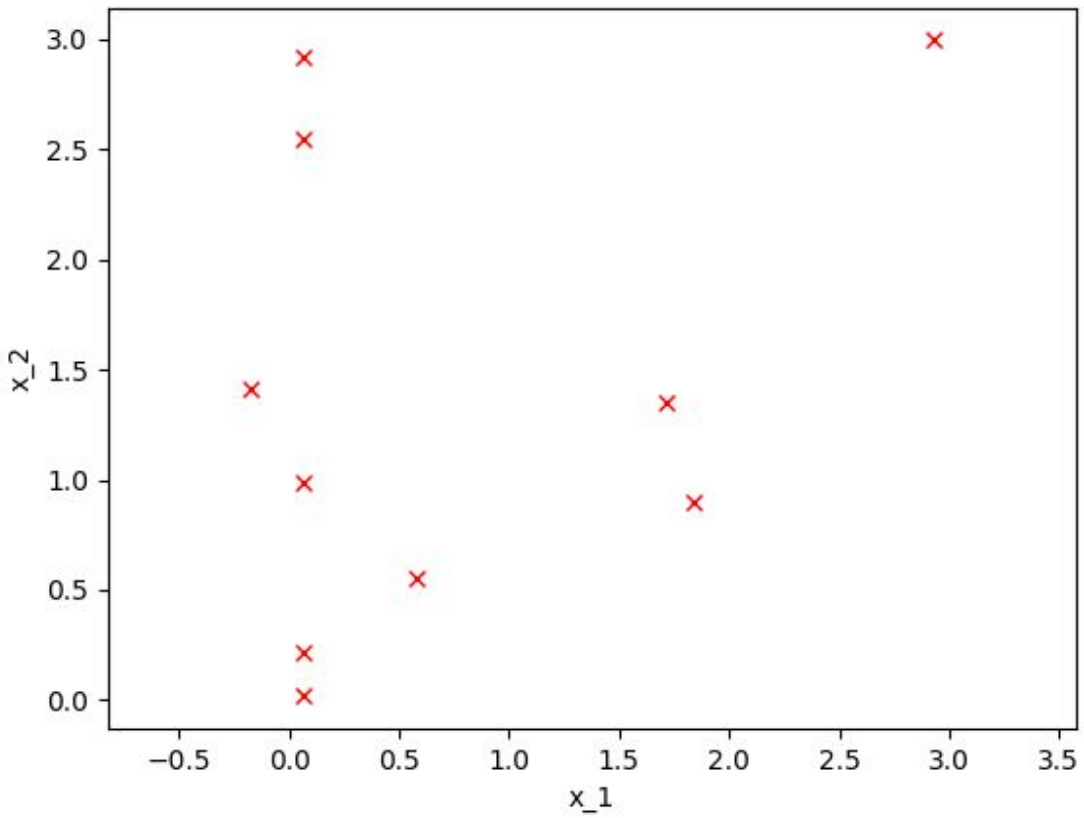
LIBIN N GEORGE
09/02/2018

Result

GMM can be used to predict the mean and variances for each component of GMM along with mixing coefficients. As part of the Lab I had implemented EM for GMM and results were for d=2 (where I can plot points and verify) Some samples are given below

Plots for d=2 (Input for EM algorithm)

1) **N = 10, d=2, k= 2**



Result of EM algorithm

Mean

[[0.10549484 1.2363555]

[2.15957339 1.74819128]]

Variance

[array([[0.04391549, -0.0558607],

[-0.0558607 , 1.09401573]]),

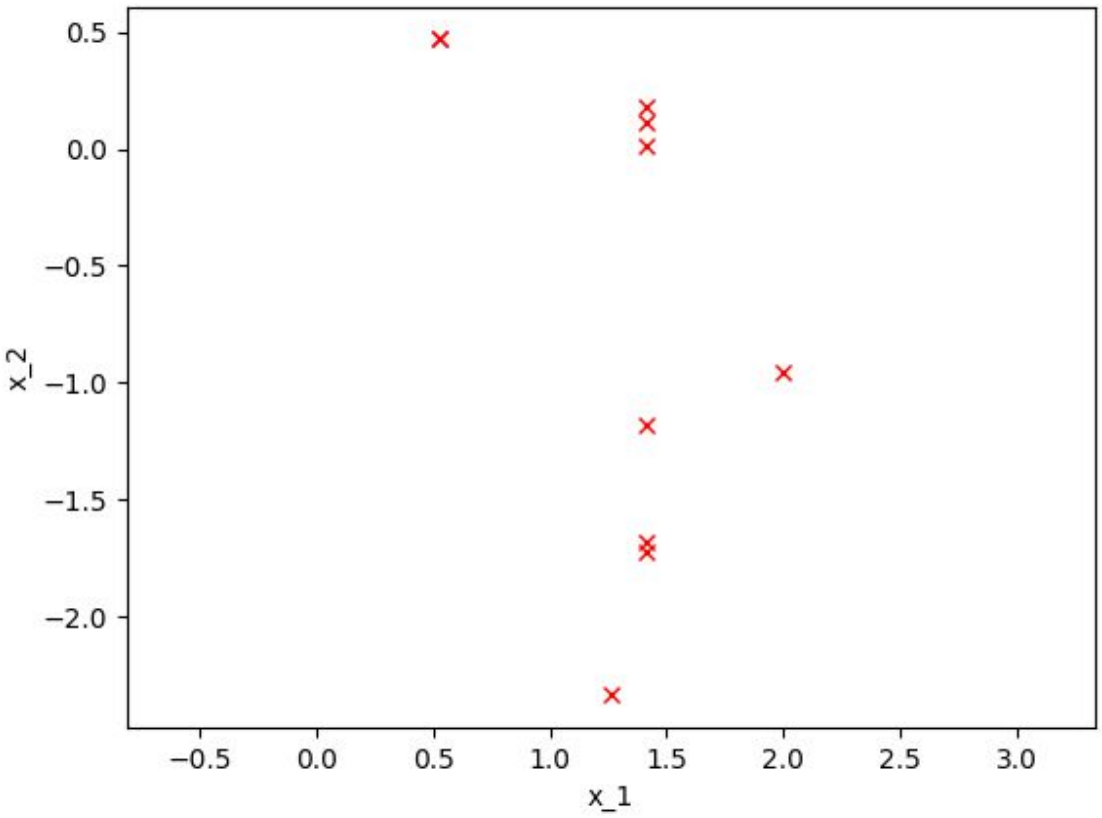
array([[0.29854633, 0.47079186],

[0.47079186, 0.81226983]])]

Mixing Coefficient

[0.6999999126322525, 0.3000000873677475]

2) d=2 n=10 k=5



[[1.41078732 -0.85570558]

[1.25887515 -2.33579823]

[1.25887515 -2.33579823]

[1.70456222 -0.47128219]

[0.52633033 0.46904129]]

```
[array([[0.       , 0.       ],
        [0.       , 0.71222723]]),
array([[2.09991207e-12, 1.12289071e-11],
        [1.12289071e-11, 7.74436004e-11]]),
 array([[1.59534739e-13, 6.73470461e-13],
        [6.73470461e-13, 2.87306701e-12]]),
array([[ 0.08630368, -0.14217086],
        [-0.14217086,  0.23420269]]),
 array([[ 1.68541082e-04, -5.62812792e-05],
        [-5.62812792e-05,  1.89182306e-05]])]
[0.49995692273681014, 0.004762471046142915, 0.09523752895494884,
0.19999998668313324, 0.20004309057896488]
```

Conclusion

MLE for GMM has no closed form solution hence Parameters are calculated using Expectation Maximization (EM) technique . As number of points increases the prediction of parameters becomes more easier. Ie EM becomes more accurate. A limit in the number of iterations are used to stop the EM algorithm in case of parameters are not converging easily.