EM based algorithm

Report:

First I initialized 3 sets of 50 random data points with mean values as 0, 5, 10 and a standard deviation value using numpy. Each time I modified the value of standard deviation (variance) using the same value for each Gaussian and observed the results. This data is given to GMM algorithm which returns calculated means and calculated variance. With each different variance, I executed the algorithm ‘n’ number of times. I then calculated the average values of mean and variance from these ‘n’ observations. These average values of means and variance are used to calculate the accuracies of these parameters. Accuracy of EM is calculated using the number of correct predictions.

Observations are noted below in the table:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Mean values | Variance | Calculated Mean values | Calculated Variance | Accuracy in calculating true means (difference with true means) | Accuracy in calculating true variance | Accuracy of EM (%) in assigning points to true mixture |
| 0, 5, 10 | 0.01 | 0.93786891, 5.22156679, 8.99589336 | 0.01859859 0.01859859 0.01859859 | 0.93786891, 0.22156679‬, 1.00410664‬ | 0.00859859 0.00859859 0.00859859 | 35.653 |
| 0, 5, 10 | 1 | 0.47964183, 4.68562297  , 8.97339355 | 0.07007064 0.07007064 0.07007064 | 0.47964183, 0.31437703, 1.02660645 | 0.92992936 0.92992936 0.92992936 | 30.167 |
| 0, 5, 10 | 4 | 0.34087022, 4.42254066, 9.7830623 | 0.49646801 0.49646801 0.49646801 | 0.34087022, 0.57745934, 0.2169377 | 3.50353199 3.50353199 3.50353199 | 30.8 |
| 0, 5, 10 | 25 | 3.88, 4.5861881, 12.61306953 | 0.17659936 0.17659936 0.17659936 | 3.88 0.63819855 2.9065537 | 24.82340064 24.82340064 24.82340064 | 34.053 |
| 0,5,10 | 9 | 1.70024141, 3.64277421, 10.42509654 | 0.17003185 0.17003185 0.17003185 | 1.70024141 0.04491756 0.90008044 | 8.82996815 8.82996815 8.82996815 | 32.906 |