**QUESTION 1:**



Figure 1: 4 Component GMM



Figure 2: BIC Score vs Number of Components for 10, 100, 1000, 10000 Samples

COMMAND OUTPUT:

Dataset 1

Order 1, BIC Score 73.0890

Order 2, BIC Score 77.7689

Order 3, BIC Score 75.6872

Order 4, BIC Score 78.8282

Order 5, BIC Score 81.9692

Order 6, BIC Score 88.8769

Best order of GMM 1

Dataset 2

Order 1, BIC Score 643.6861

Order 2, BIC Score 601.8075

Order 3, BIC Score 591.5106

Order 4, BIC Score 600.8061

Order 5, BIC Score 612.9097

Order 6, BIC Score 625.2938

Best order of GMM 3

Dataset 3

Order 1, BIC Score 6256.6670

Order 2, BIC Score 6003.6085

Order 3, BIC Score 5800.5308

Order 4, BIC Score 5818.8500

Order 5, BIC Score 5839.0156

Order 6, BIC Score 5859.3737

Best order of GMM 3

Dataset 4

Order 1, BIC Score 62546.1144

Order 2, BIC Score 59658.3922

Order 3, BIC Score 57489.1025

Order 4, BIC Score 57277.3967

Order 5, BIC Score 57304.9610

Order 6, BIC Score 57332.2195

Best order of GMM 4

OUTPUT TABLE:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Dataset 1 | Dataset 2 | Dataset 3 | Dataset 4 |
| Order 1 | 73.089 | 643.6861 | 6256.667 | 62546.1144 |
| Order 2 | 77.7689 | 601.8075 | 6003.6085 | 59658.3922 |
| Order 3 | 75.6872 | 591.5106 | 5800.5308 | 57489.1025 |
| Order 4 | 78.8282 | 600.8061 | 5818.85 | 57277.3967 |
| Order 5 | 81.9692 | 612.9097 | 5839.0156 | 57304.961 |
| Order 6 | 88.8769 | 625.2938 | 5859.3737 | 57332.2195 |

Report:

1. First generate 4 component GMM as true data (Figure 1)
2. Generate 10, 100, 1000, 10000 sample datasets
3. For each dataset, perform K fold cross validation to generate BIC scores after fitting samples using EM algorithm with varying number of hyperparameters.
4. Minimum BIC scores determine the best hyperparameter fits.
5. Output table summarizes results with the highlighted values being the best BIC score
6. Figure 2 shows that with few samples, there is not enough to tell if many hyperparameters are the best fit, so the BIC score assumes the 10 samples are from a GMM with only one model. With 100 samples there is a U shaped BIC score vs number of components which indicated that the optimum is at order 3 since with this number of samples is just between not enough sample and too much samples where the cross fold validation penalizes for not enough and too many hyper-parameters and the lowest BIC score is the “sweet spot”. With more samples, the BIC scores effectively determine the best model. Additionally, the BIC score is based off the two-term equation :

*bic = -2log(logL) + numParam\*log(numObs)*

where L is the maximized likelihood for the estimated model. With fewer samples, the second term the penalty for model parameters has a higher effect. On the other hand, with more samples, the first term has a higher effect and dominated the second term.

**QUESTION 2:**

****

****

****

****

****

**RESULTS:**

The following table summarizes the results where although each classifier performs well on the training set, the MAP classifier performed the best with the dataset.

|  |  |
| --- | --- |
|  | P(error) |
| MAP | .0400 |
| LDA | .0430 |
| LOGISTIC REGRESSION | .0044 |