Name: Andrew Gan
ECE 20875 – HW10
18-Nov-2019
Prof. Milind Kulkarni

Fitted mixture model formulas for k = 2, 3, 4, 5, 6 using tol = 1

All values rounded to 3 or 4 decimal places.
Directly printed to console by running gmm_em.py as main module

k = 2
0.471 N(x | 3.189, 1.699) + 0.529 N(x | 11.002, 9.563)
ll = -1091.8566

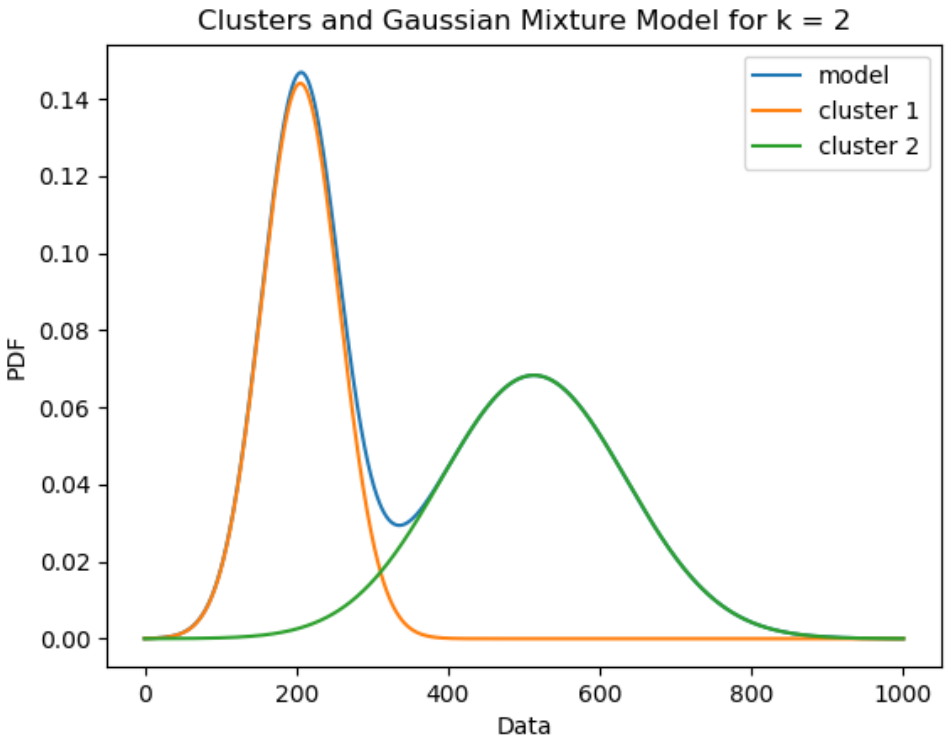
k = 3
0.249 N(x | 2.028, 0.344) + 0.245 N(x | 4.468, 0.08) + 0.506 N(x | 11.316, 7.742)
ll = -999.7635

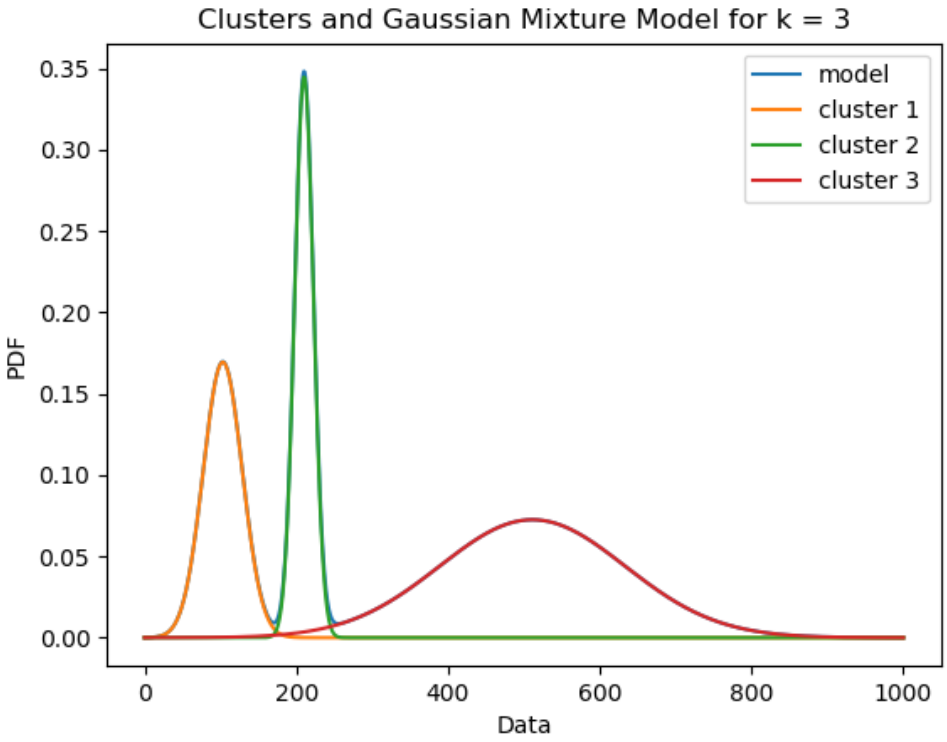
k = 4
0.25 N(x | 2.03, 0.345) + 0.25 N(x | 4.47, 0.081) + 0.25 N(x | 8.893, 0.353) + 0.25 N(x | 13.903, 1.519)
ll = -910.9247

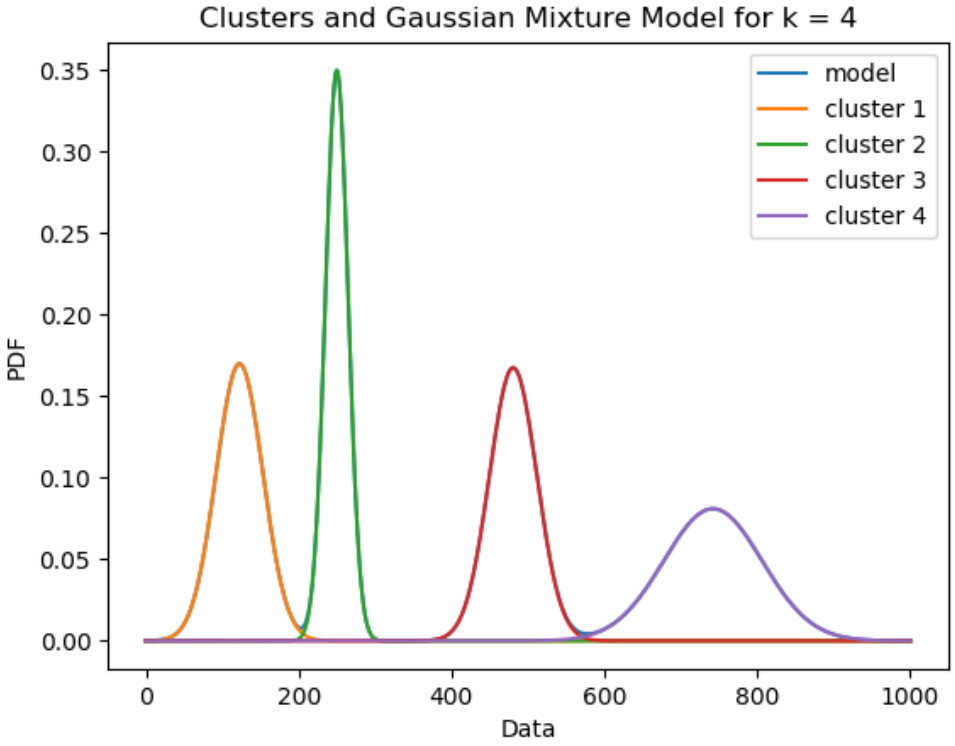
k = 5
0.25 N(x | 2.03, 0.345) + 0.25 N(x | 4.47, 0.081) + 0.057 N(x | 8.727, 0.302) + 0.192 N(x | 8.942, 0.358) + 0.25 N(x | 13.903, 1.519)
ll = -910.8393

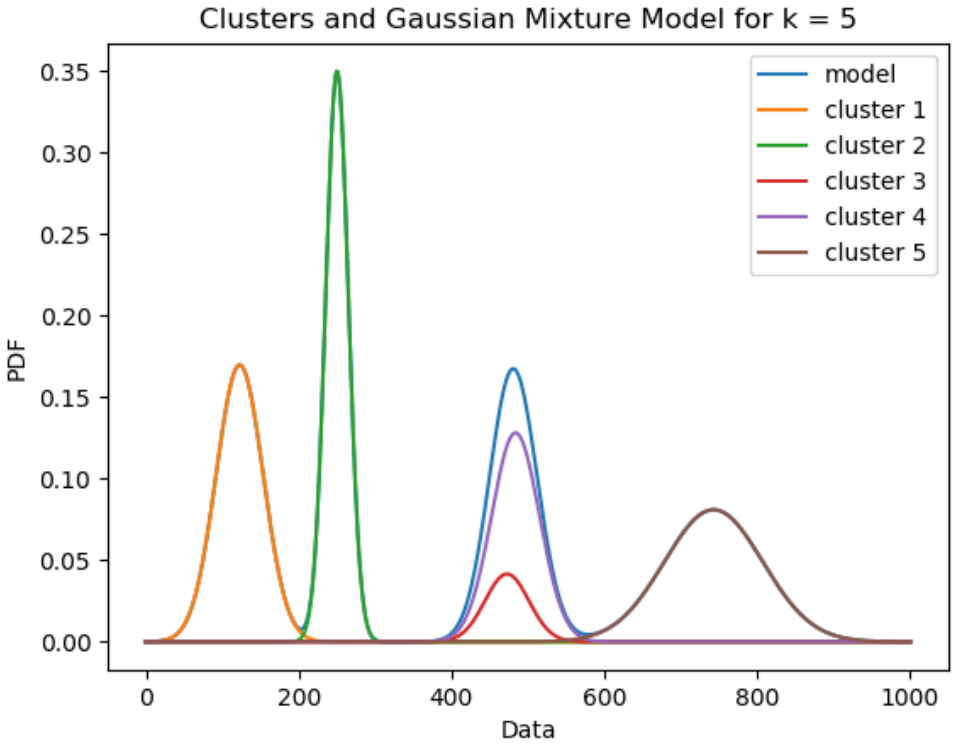
k = 6
0.224 N(x | 1.994, 0.333) + 0.031 N(x | 2.609, 0.776) + 0.245 N(x | 4.473, 0.08) + 0.249 N(x | 8.892, 0.353) + 0.013 N(x | 13.291, 1.489) + 0.238 N(x | 13.933, 1.51)
ll = -912.4968

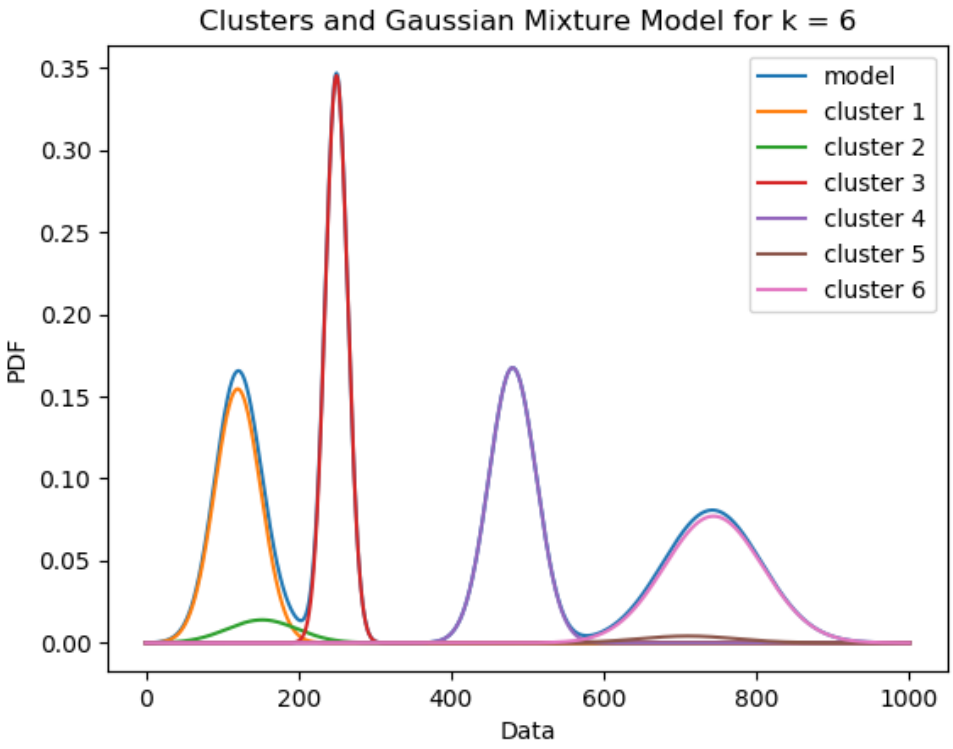
It is observed that the log-likelihood increased from -1091 for k = 2 to -910 for k = 5, and then decreased again to -912 for k = 6


Clusters and Gaussian Mixture Model for k = 2
model
cluster 1
cluster 2
PDF
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
200
400
600
800
1000
Data


Clusters and Gaussian Mixture Model for k = 3
model
cluster 1
cluster 2
cluster 3
PDF
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
200
400
600
800
1000
Data


Clusters and Gaussian Mixture Model for k = 4
model
cluster 1
cluster 2
cluster 3
cluster 4
PDF
Data
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
200
400
600
800
1000


Clusters and Gaussian Mixture Model for k = 5
model
cluster 1
cluster 2
cluster 3
cluster 4
cluster 5
PDF
Data
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
200
400
600
800
1000



K = 4 seems to be the number of clusters that best represents the nature of the data, due to it having the second highest log-likelihood, and being the highest k value without significant overlap in the plot of the clusters.