Ryan Boyd
ECE 20875
HW 10

3.

$$p_X(x) = \sum_{i=1}^{k} \pi_i \mathcal{N}(x|\mu_i, \sigma_i^2)$$

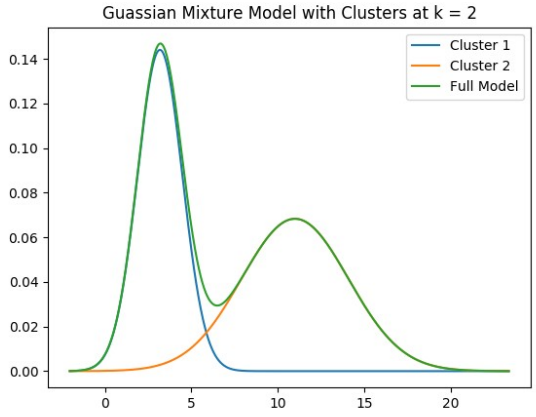
K = 2
Pi = [0.47057958 0.52942042]
Mu = [3.18871898 11.00216753]
Varis = [1.69933919 9.56272622]
K = 3
Pi = [0.24912167 0.24480779 0.50607054]
Mu = [2.02765463 4.46752085 11.31561928]
Varis = [0.34397885 0.08024393 7.74156218]
K = 4
Pi = [0.25015607 0.24984393 0.24962165 0.25037835]
Mu = [2.02957993 4.47020299 8.89291217 13.90251977]
Varis = [0.34477083 0.08109582 0.35335439 1.51933857]
K = 5
Pi = [0.25018635 0.24981365 0.05729871 0.1923332 0.25036809]
Mu = [2.0298245   4.47025391 8.72704385 8.94237416 13.90268862]
Varis = [0.34522648 0.08108099 0.30182166 0.35809268 1.51870127]
K = 6
Pi = [0.22350959 0.03103757 0.24545284 0.24948078 0.01287102 0.2376482]
Mu = [1.99364547 2.60915902 4.47329922 8.89247316 13.29140356 13.93310912]
Varis = [0.33284439 0.77645284 0.08044074 0.35312495 1.48867194 1.50964533]

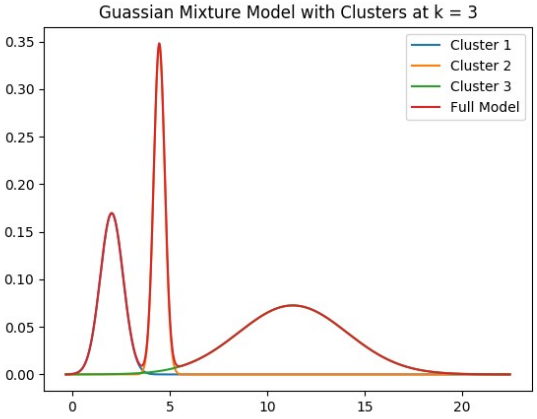
Log Likelihoods:

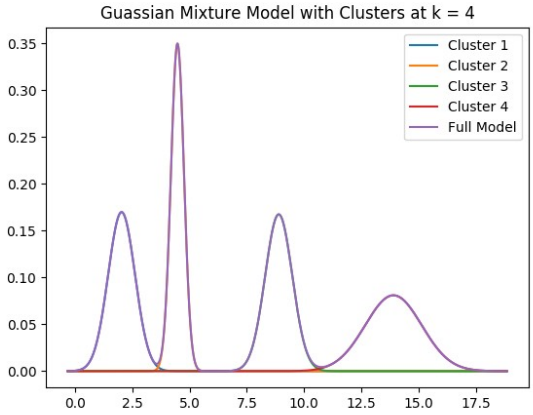
K = 2
Log Likelihood = -1091.8565606736677
K = 3
Log Likelihood = -999.7635118345282
K = 4
Log Likelihood = -910.9246936882649
K = 5
Log Likelihood = -910.8393241677736
K = 6
Log Likelihood = -912.4968389442097

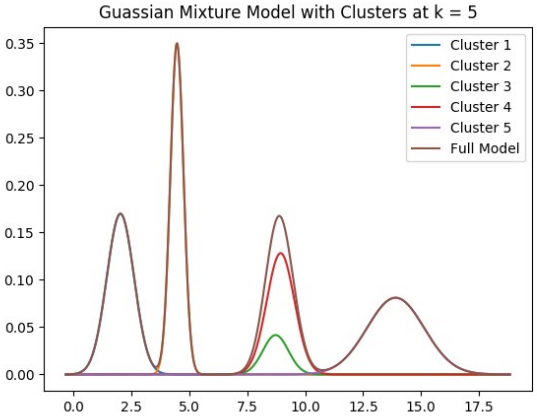
The log likelihoods are largest at K = 4 and K = 5(K = 5 likelihood is only ~.085 larger). Starting at k = 2, the log likelihoods increase until K = 4, 5, and then decrease at k = 6.

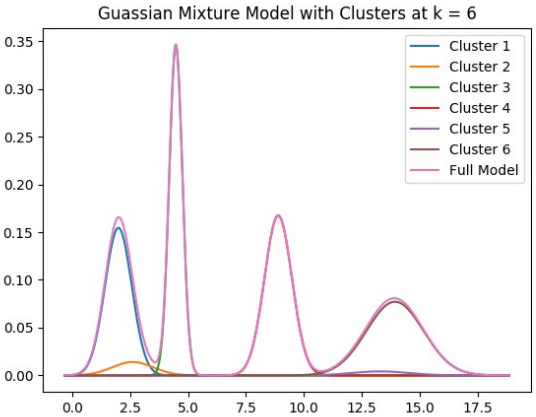
4.











This data set has 4 clusters. This is because the log likelihood for k=4 and k = 5 are the largest. Looking at the k = 5 graph, there is one extra cluster in the mixture model, causing another of the clusters to be smaller than it should be. In the k = 4 graph, the 4 clusters almost perfectly fit the mixture model.