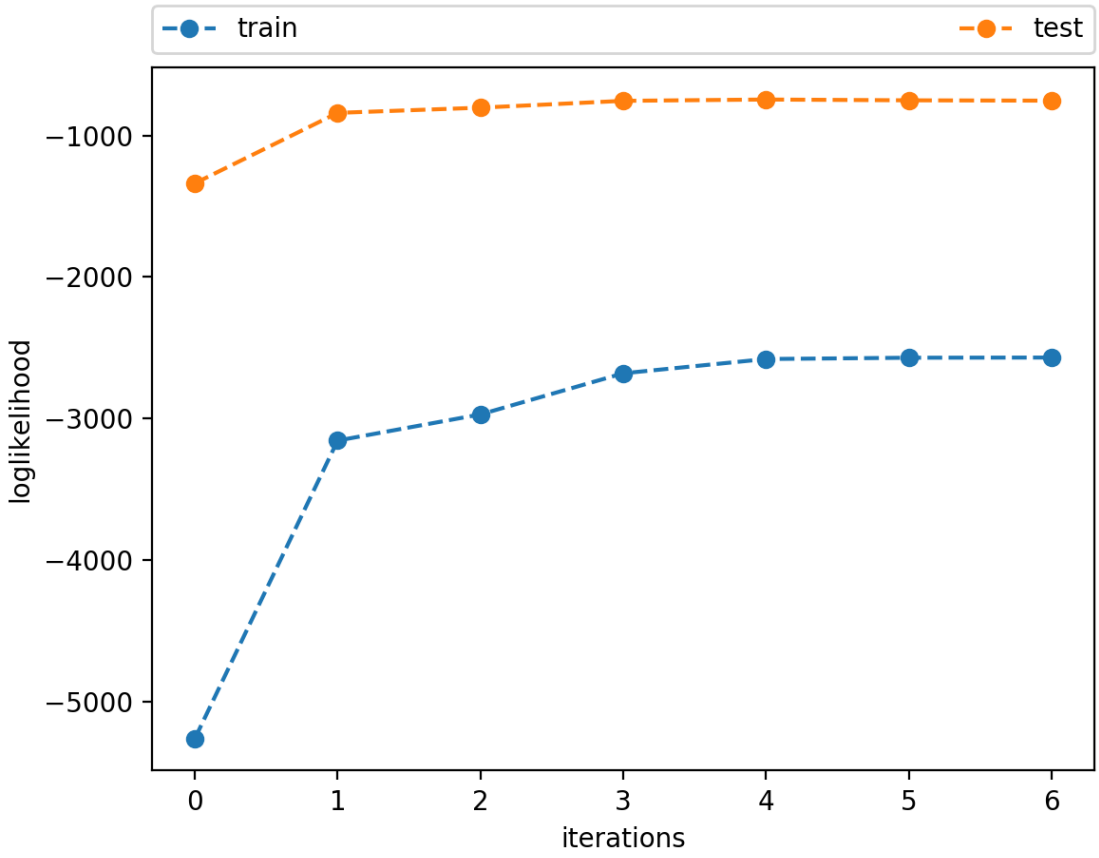
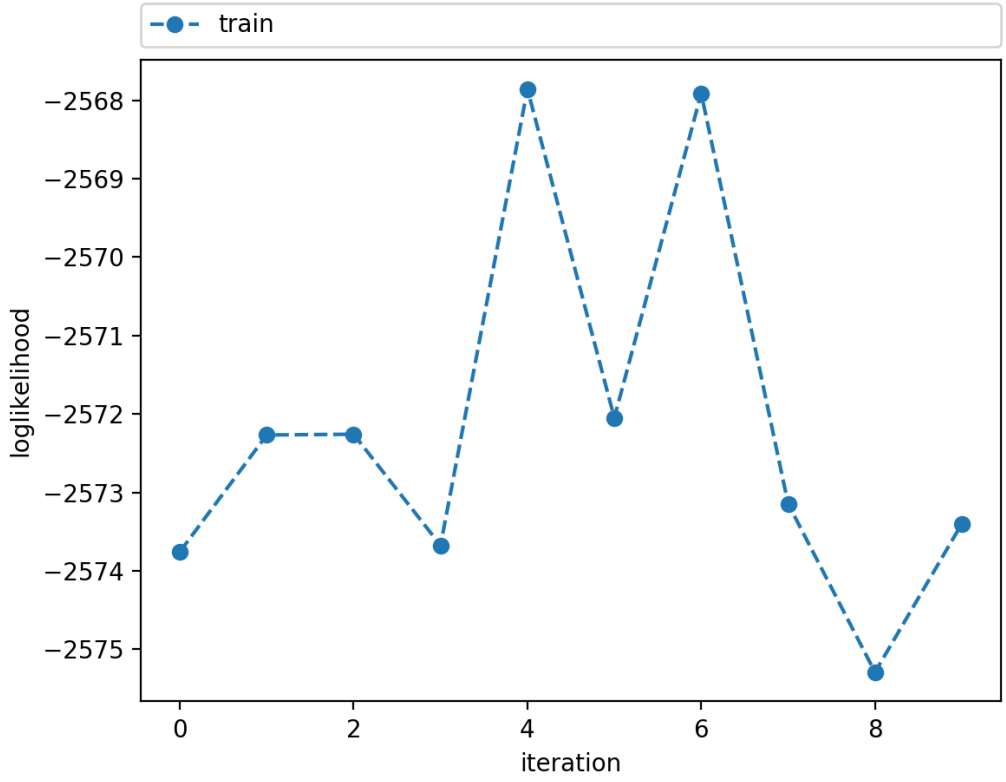
(a) 2 points Plot train and test set likelihood vs. iteration. How many iterations does EM take to converge?

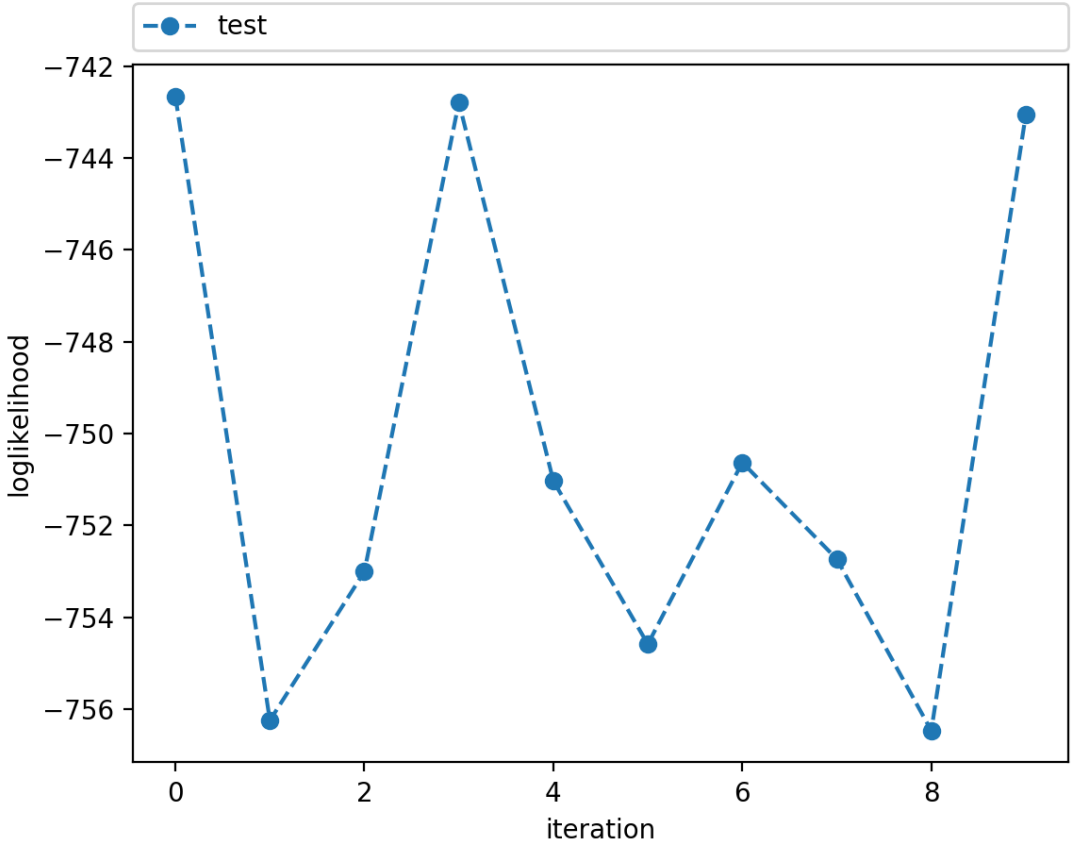


This varies between 5 – 14, usually between 7 – 9

(b) 2 points Run the algorithm 10 times with different random seeds. How much does the log likelihood change from run to run?

It does not change very much. < 2% change.


train
−2568
−2569
−2570
−2571
−2572
−2573
−2574
−2575
loglikelihood
0
2
4
6
8
iteration

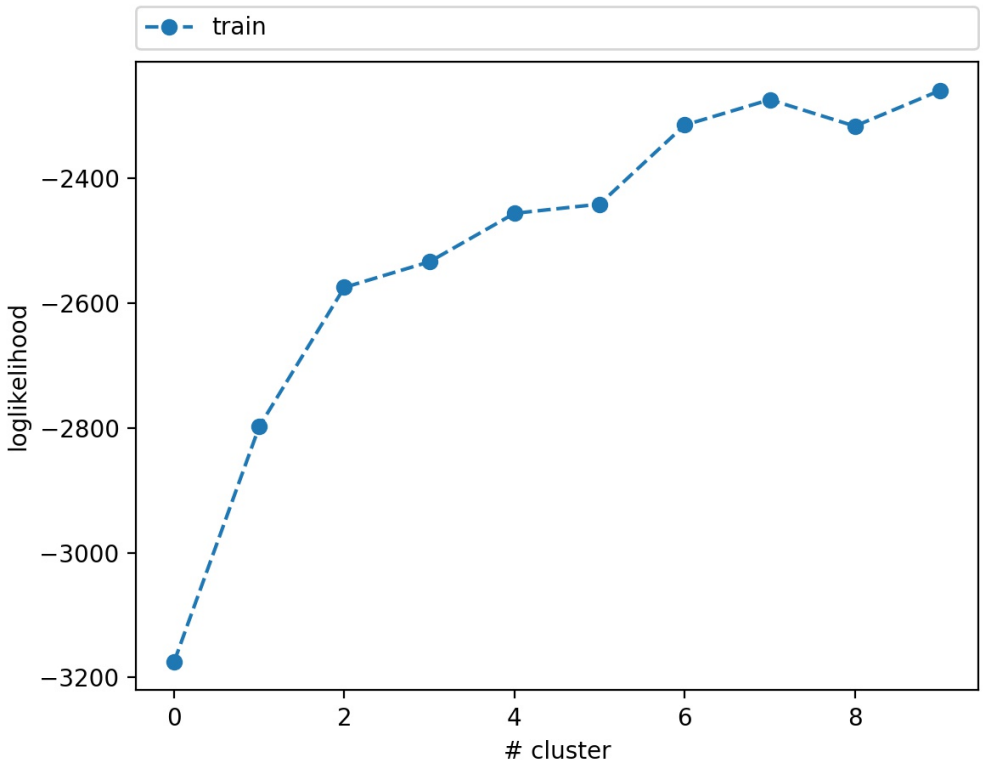

test
−742
−744
−746
−748
−750
−752
−754
−756
loglikelihood
0
2
4
6
8
iteration

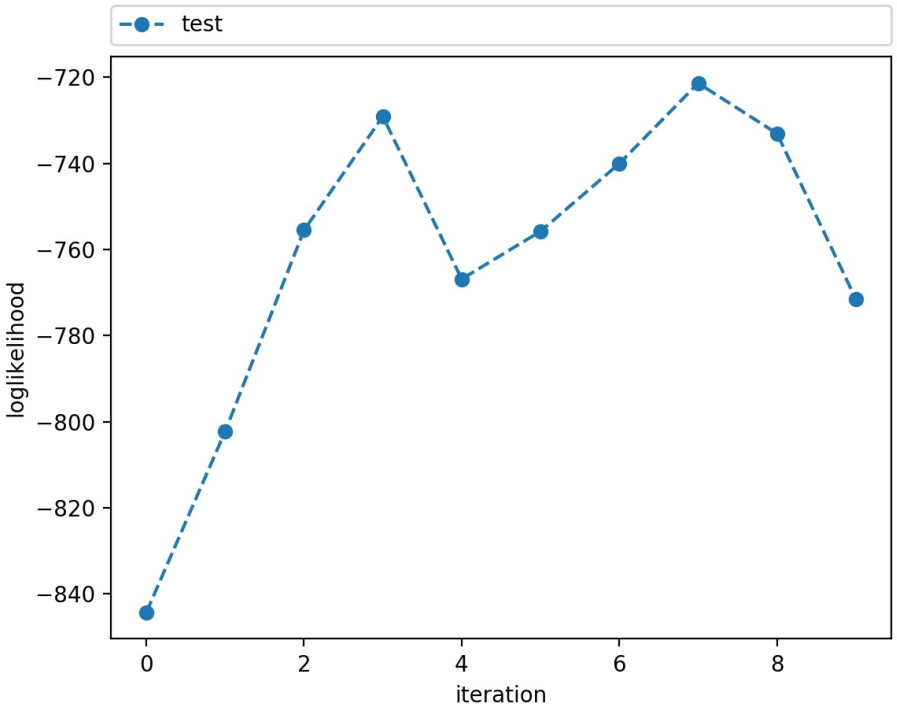
(c) 2 points Infer the most likely cluster for each point in the training data. How does the true clustering (see wine-true.data) compare to yours?

95.07% correct

(d) 3 points Graph the training and test set log likelihoods, varying the number of clusters from 1 to 10. Discuss how the training set log likelihood varies and why? Discuss how the test set log likelihood varies, how it compares to the training set log likelihood, and why. Finally, comment on how train and test set performance with the true number of clusters (3) compares to more and fewer clusters and why.

The training set log likelihood generally increases as the clusters increase, this is because it allows the data sets in group themselves such that they are more similar to each other. The testing set log likelihood will increase for a time until the clusters is greater than the true value and then it will begin to decrease, this is because the training data is being classified into groups that don't really exist and so the testing data will also be grouped that way and it's wrong. The testing likelihood is much higher than the training likelihood because it has a smallest size. The true number for the training is higher than the lower cluster and lower than the higher clusters because of what I have previously said. The true number for the testing is higher than the lower cluster and higher than the higher clusters because too few of clusters will cause groups to merge and not be accurate and too many clusters will be overtraining of the train data.




test
−720
−740
−760
−780
−800
−820
−840
loglikelihood
0
2
4
6
8
iteration