output

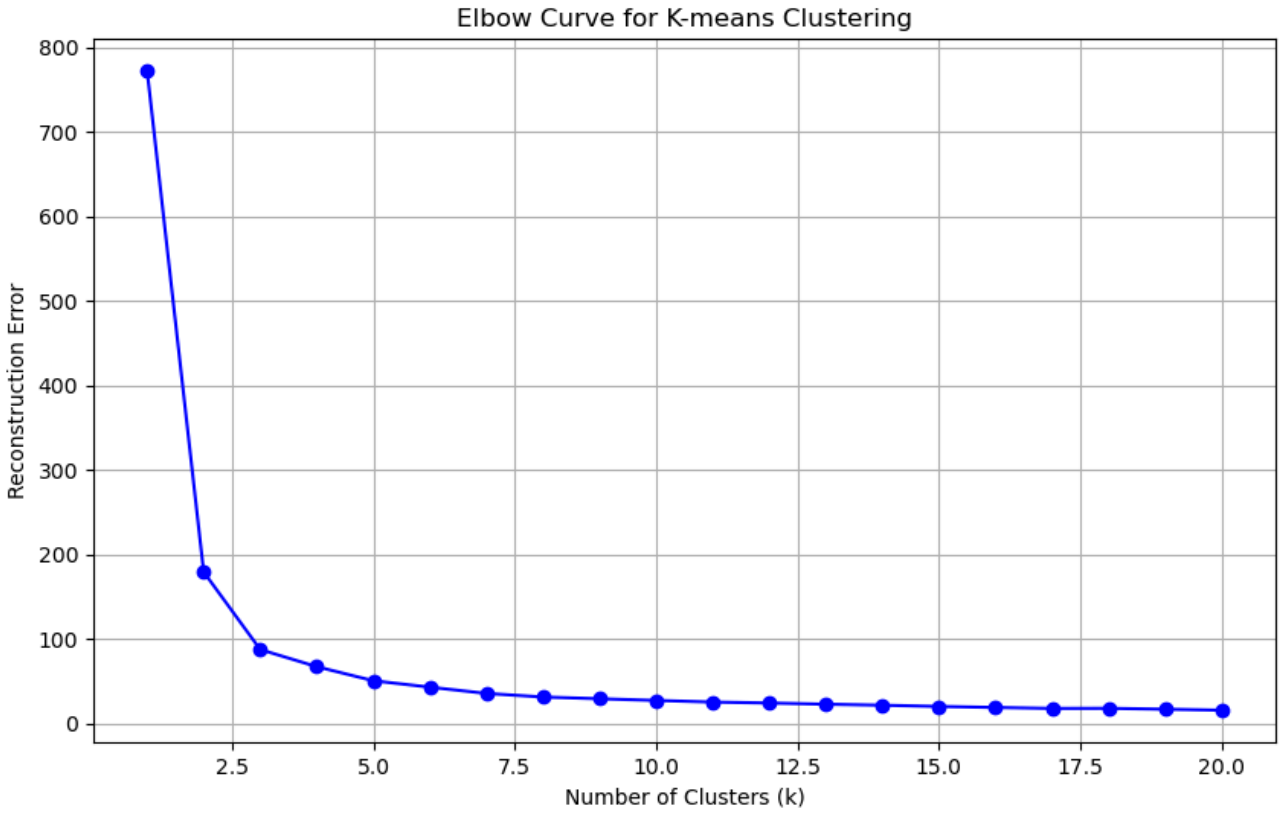
Part 1: K-means Clustering

Plotting Elbow Curve

k = 3:

Confusion Matrix:
[[0 49 0]
[2 0 48]
[36 0 14]]

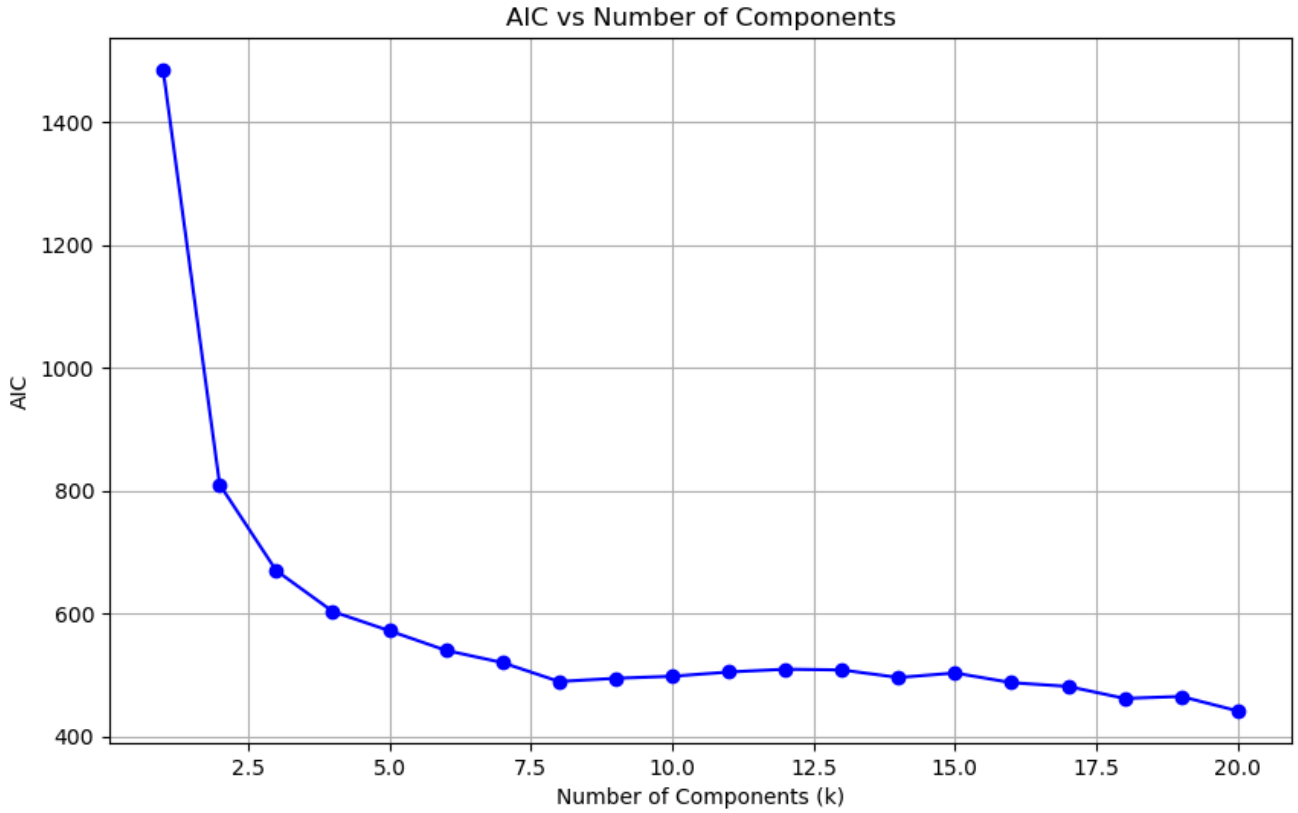
Accuracy Score: 0.8926



Question 1: Yes because k=3 is the elbow and there are 3 classes

GMM Results using AIC elbow k = 3

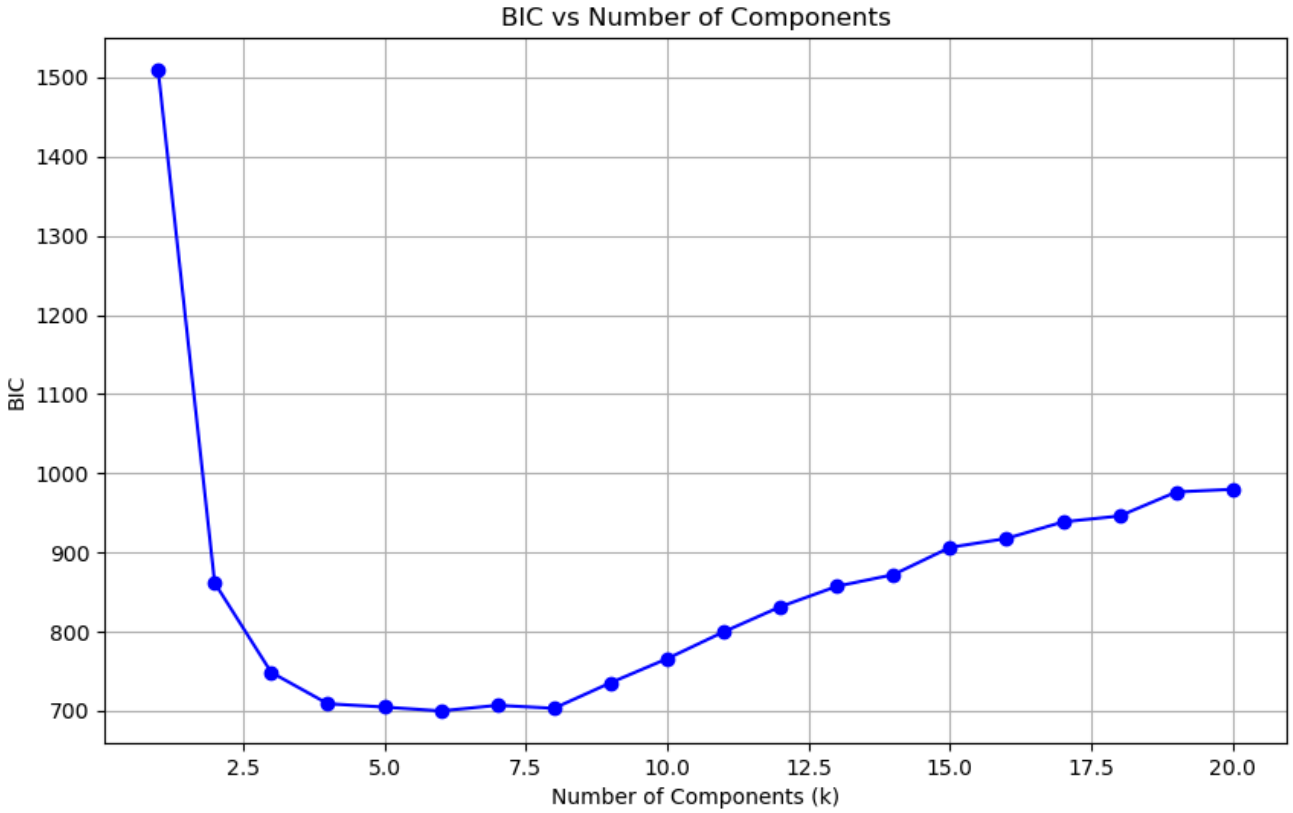
Confusion Matrix:
[[49 0 0]
[0 49 1]
[0 14 36]]



Accuracy Score: 0.8993

GMM Results using BIC elbow k = 3

Confusion Matrix:
[[49 0 0]
[0 49 1]
[0 14 36]]

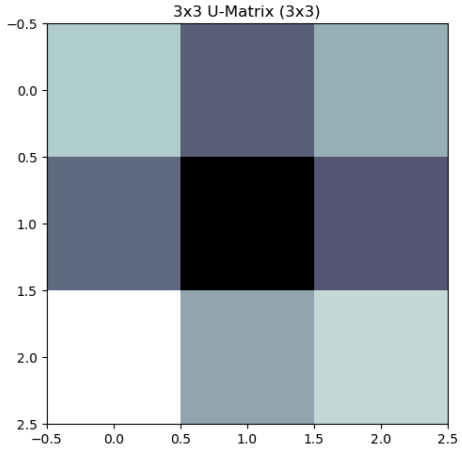


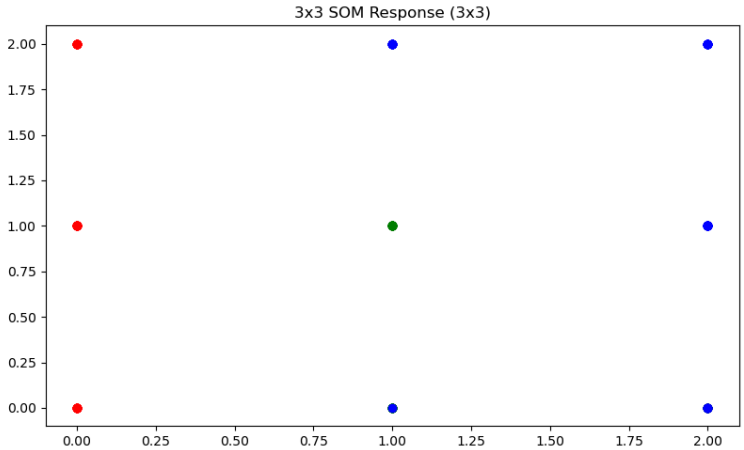
Accuracy Score: 0.8993

Question 2a: Yes because again k=3 is the elbow and thats how many classes there are

Question 2b: Yes same reason

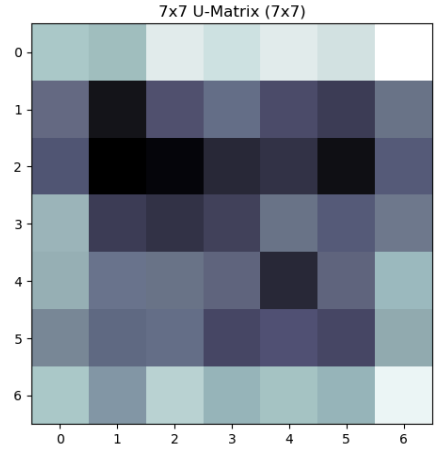
Training 3x3

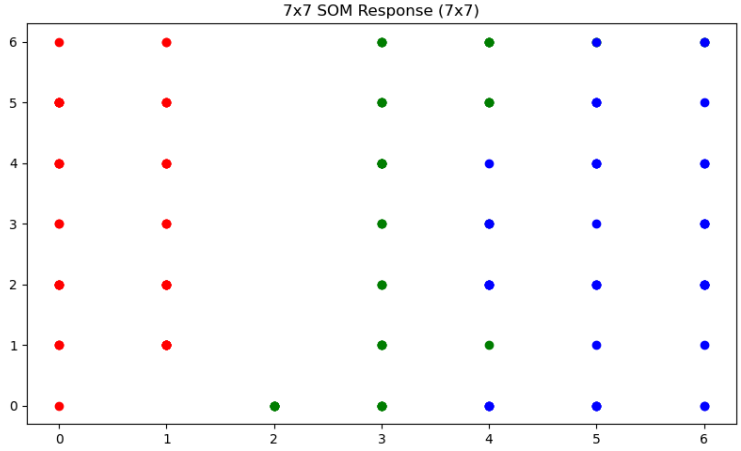




Quantization Error for 3x3 grid: 0.1293

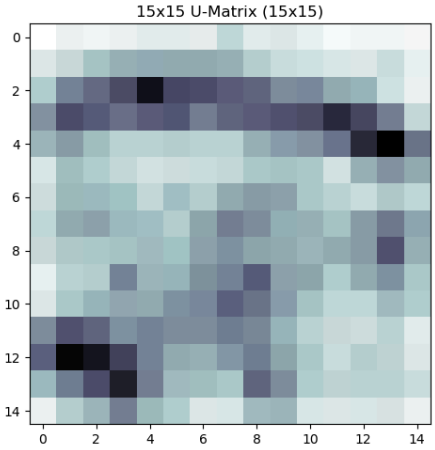
Training 7x7

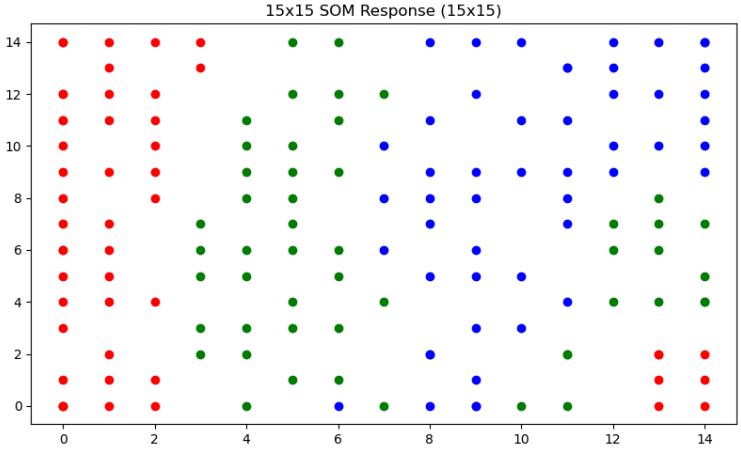




Quantization Error for 7x7 grid: 0.0589

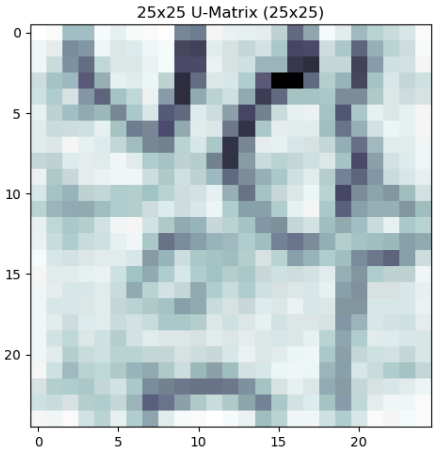
Training 15x15

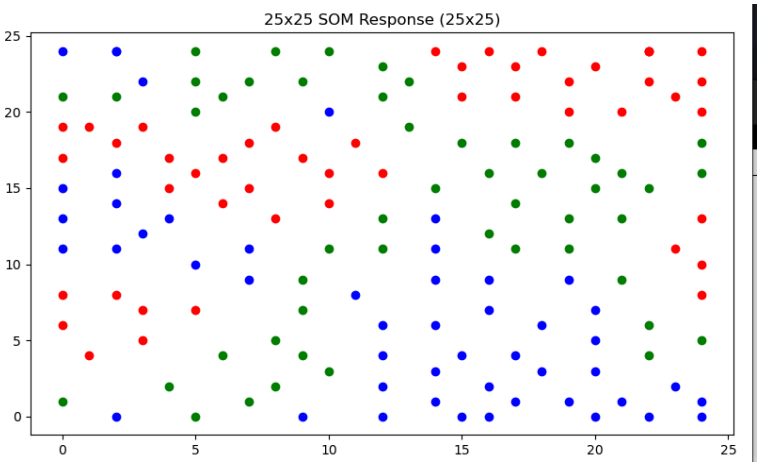




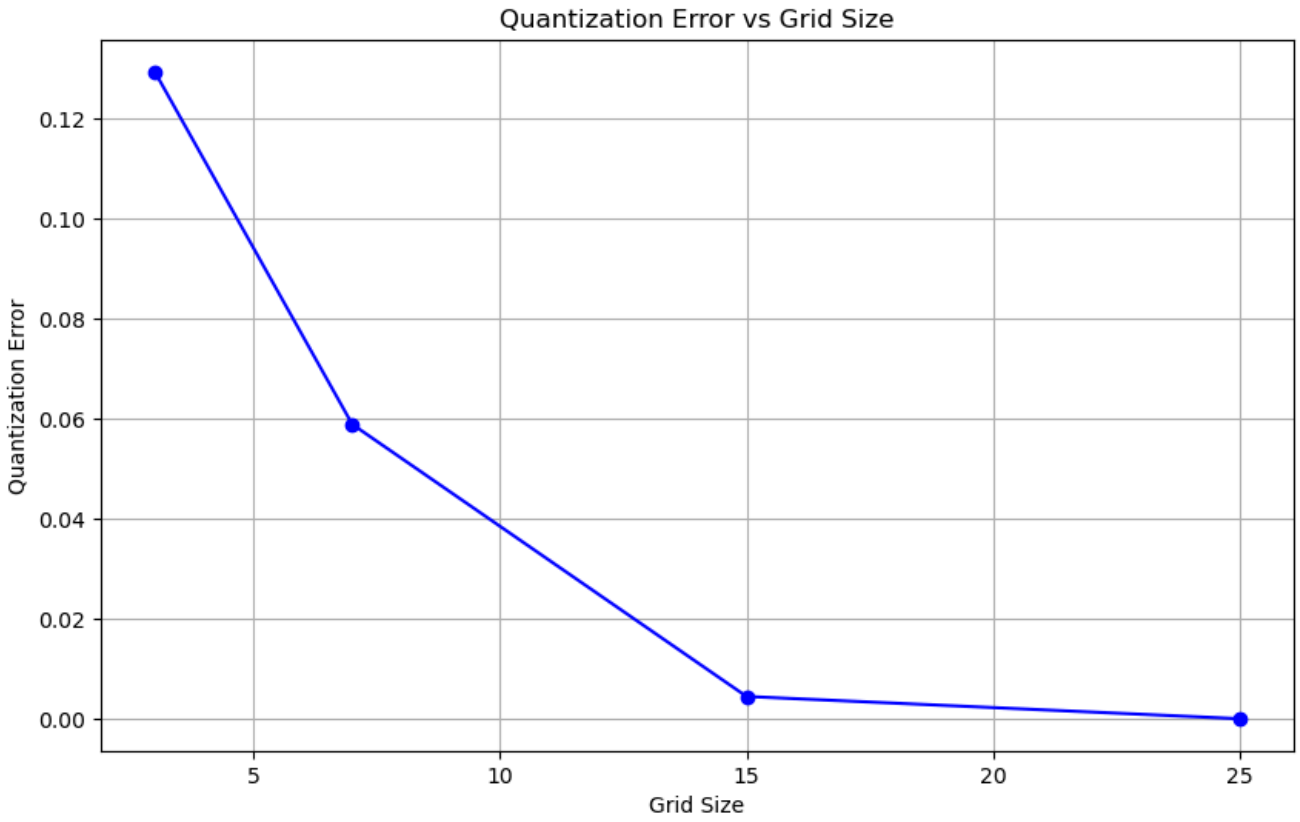
Quantization Error for 15x15 grid: 0.0045

Training 25x25





Quantization Error for 25x25 grid: 0.0000



Question 3a: The elbow is between 7x7 and 15x15 so I will say 7x7

Question 3b: The better the grid size, the better the lower the error but it reaches a point of overfitting. It also takes longer to train

Question 3c: 7x7 because the dataset only has 150 samples, 25x25 would be too much and wouldn't actually be a good classifier