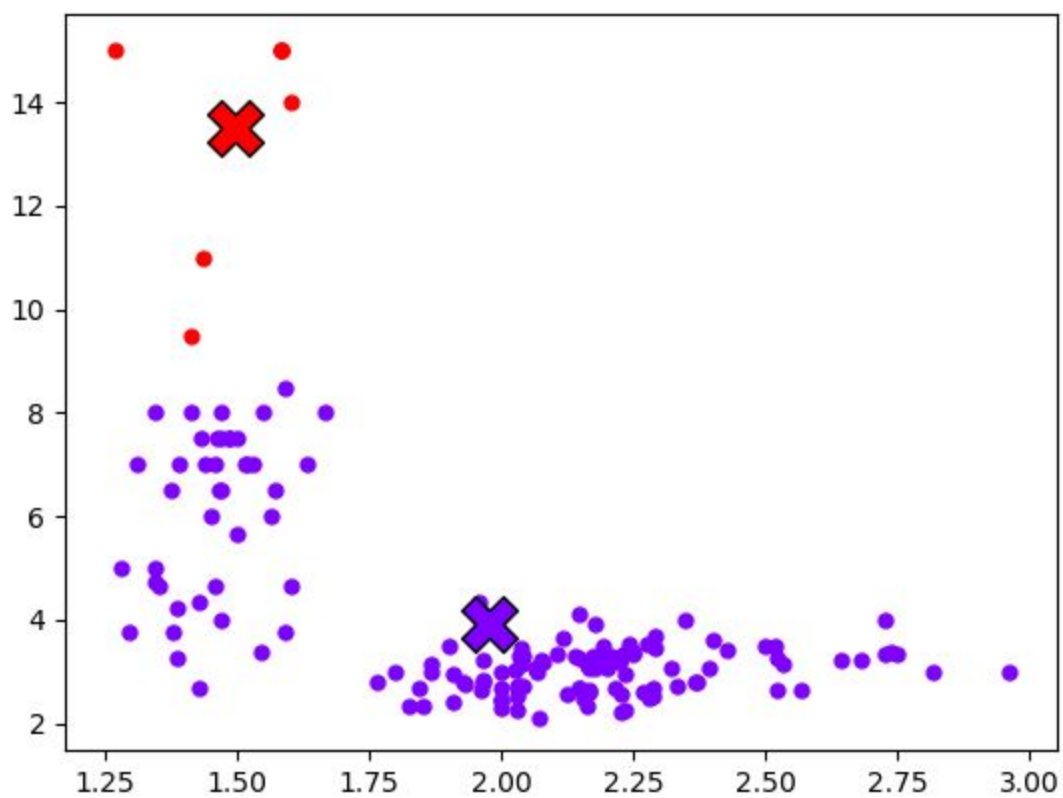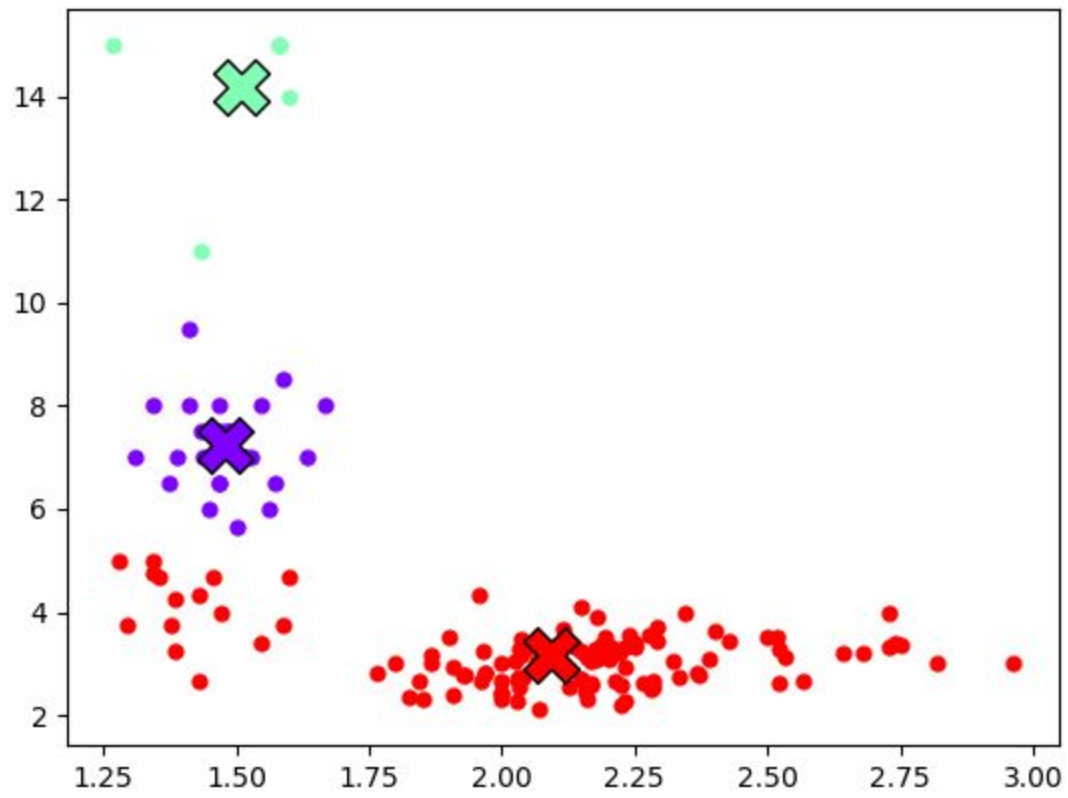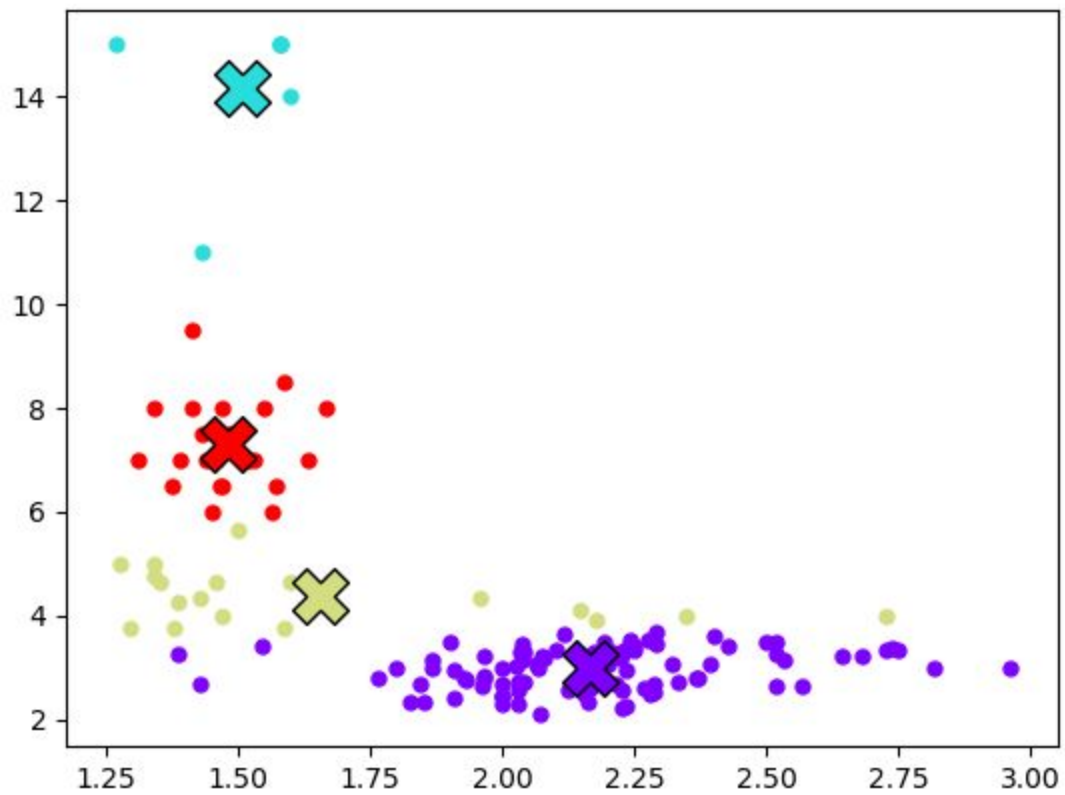Kam Li

1.   Looking through the plots, I have determined that 3 cluster is the best clustering number.
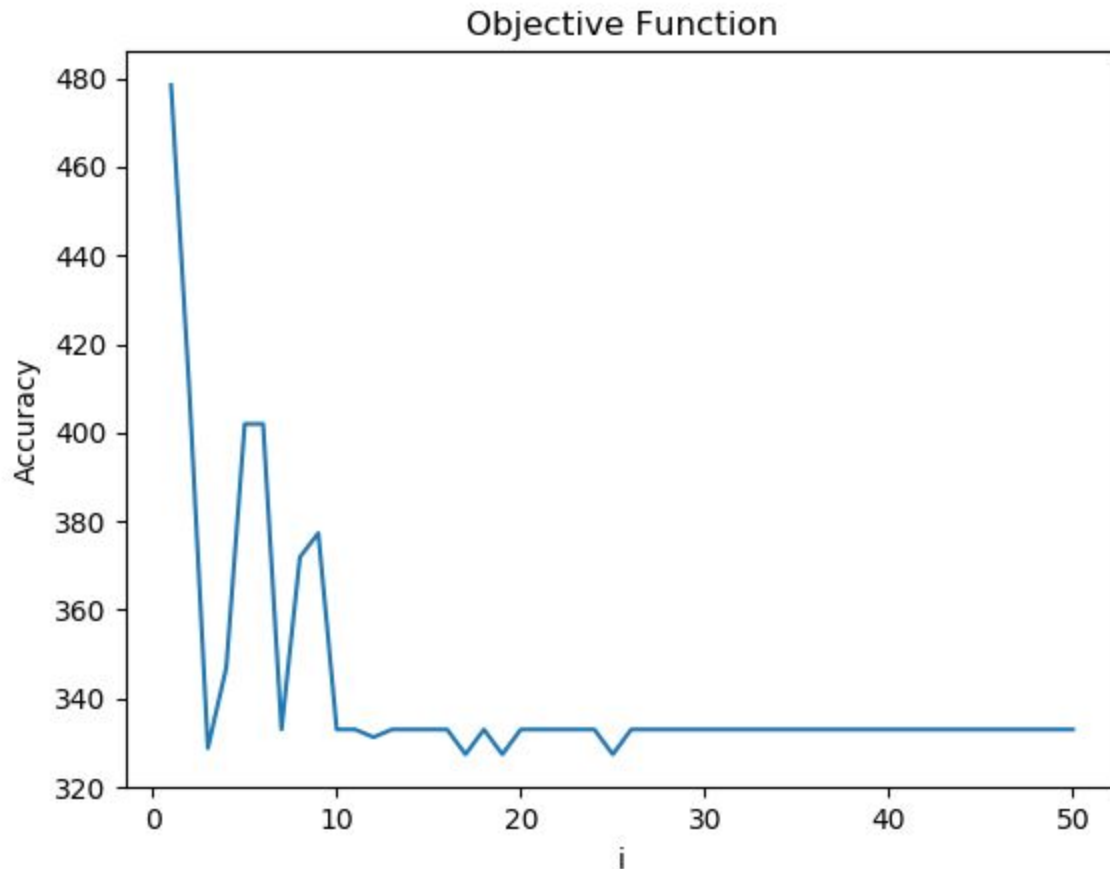

-     2 Clusters

- 3 Clusters

- 4 Clusters

- Out of all three of these plots, two is okay, three is even better, four has overlaps in our data. So three is the best choice.

Objective Function

- Accuracy over 50 iterations

2.
- **Centering or mean removal**
    - Centering or mean removal is the process of centering our features to zero. To apply centering on a data matrix, we need to find the features mean and subtract the mean from its corresponding feature.

- **Scaling of a feature to a range[a,b]**
    - Scaling is the process of reducing our data to an interval, where a is the smallest possible value and b is the biggest possible value. To apply scaling on a data matrix, we first have to find the smallest and biggest value of each feature. Then we would subtract by the smallest value and divide by the biggest value on its corresponding feature.
- **Standardization**
    - Standardization is the process of scaling our data so that there is no means and one variance. To apply standardization on a data matrix we need to find the mean and

standard deviation of each feature. Then we subtract the mean and divide by the standard deviation off of its corresponding feature.
- **Normalization**
    - Normalization is the process of fitting out features value into a interval. Such interval is usually from 0 to 1. We do this to ensure our data is in the same scale as each other. We do not want one feature with a extremely high value skew our PCA. To apply normalization on a data matrix , we get use the norm of the data. Meaning we just find the length of the vector and divide every entry by that length.
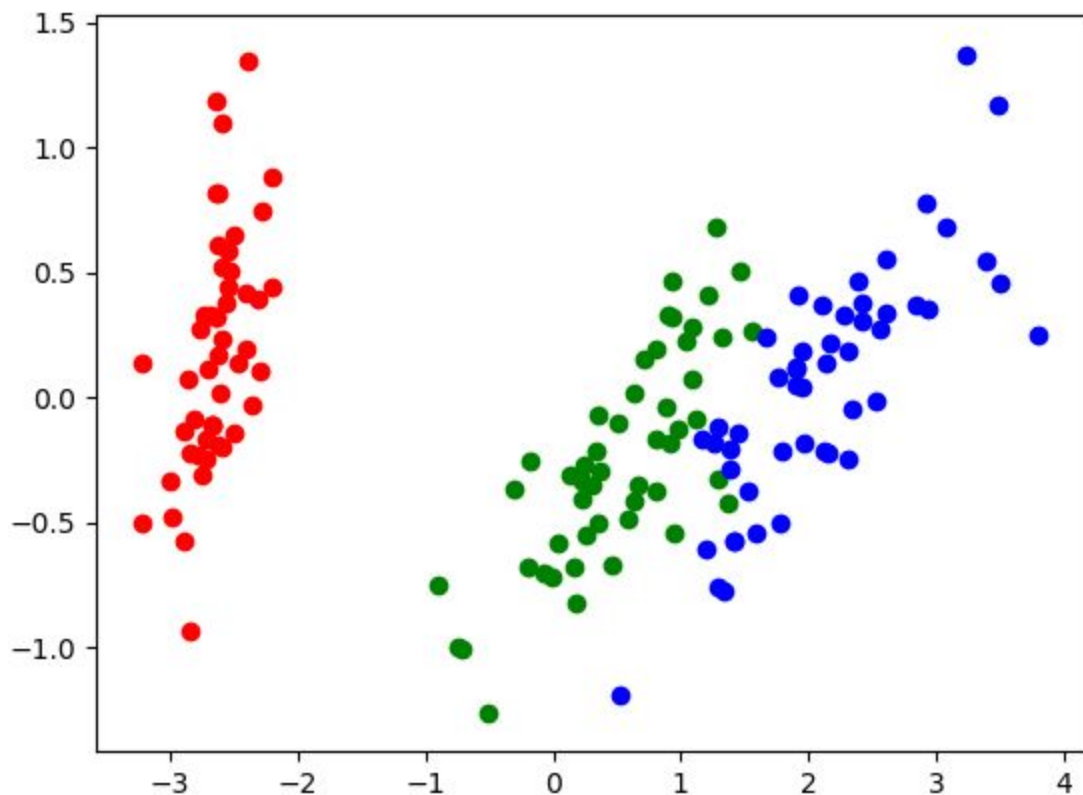
3. PCA

1.
Variance: [0.92461621]
Variance: [0.92461621 0.05301557]
Variance: [0.92461621 0.05301557 0.01718514]
Variance: [0.92461621 0.05301557 0.01718514 0.00518309]

- Original cover 92.46 over the first principal component

3.