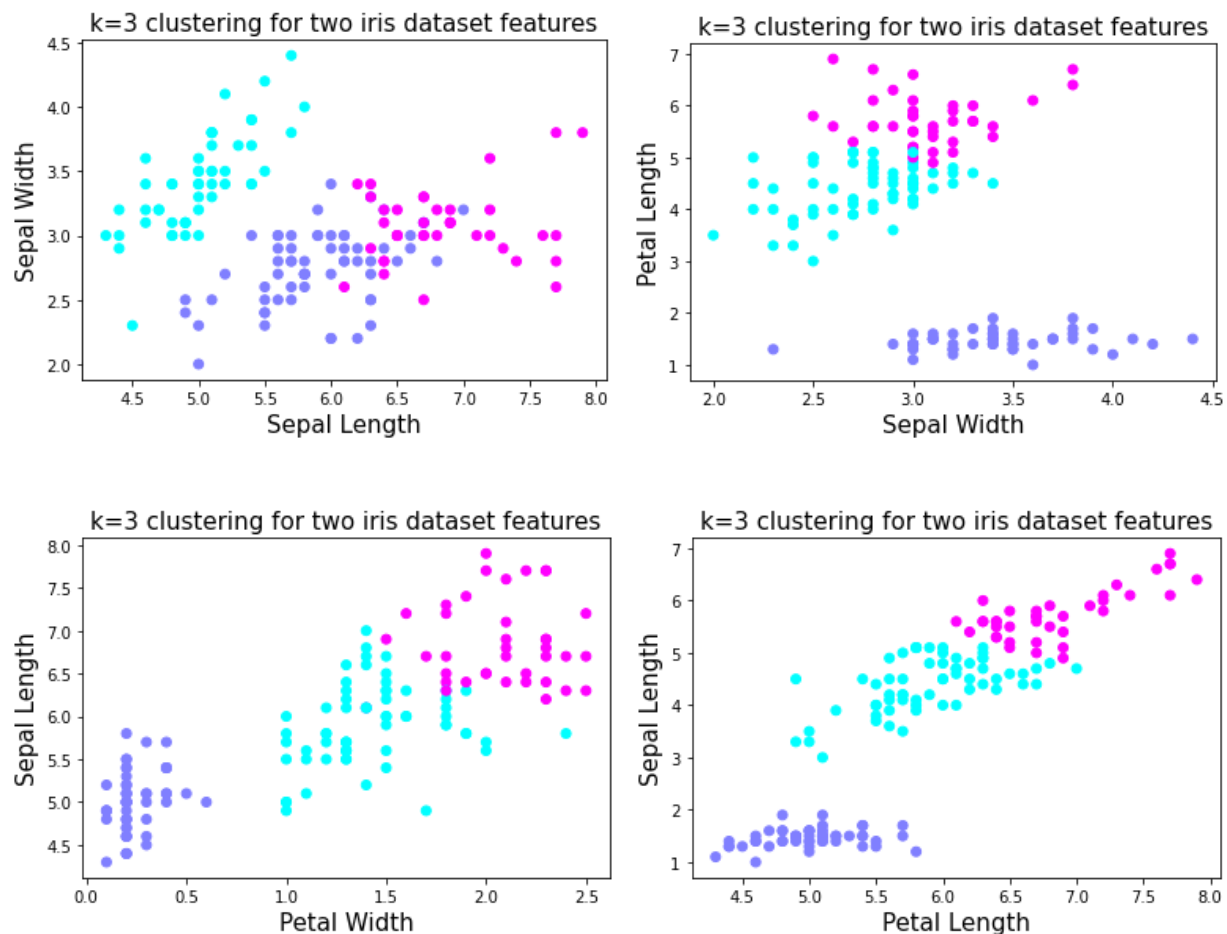


1. According to your results (i.e., elbow_k), are there 3 species of iris represented in the iris data set?

According to the k-value corresponding to the elbow of the reconstruction_error vs. k_clusters plot, the optimal number of species/classes appears to indeed be 3 for the iris dataset. I include four examples below of two dataset features each, with the samples color-coded by cluster integer. As illustrated, the three clusters that the model found do well at parsing the data if we only consider this 2D display, aside from the occasional partial entangling of classes.



2a. According to your AIC results (i.e., aic_elbow_k), are there 3 species of iris represented in the iris data set?

I ran GMM with AIC results a few times, with each iteration yielding an elbow value of either 4 or 5 (i.e., never 3). Despite the inherent variation in clustering models, then, AIC does not generate a plot with a k=3 elbow, meaning that the results do not quite support there being 3 species of iris in this dataset.

2b. According to your BIC results (i.e., `bic_elbow_k`), are there 3 species of iris represented in the iris data set?

Also despite the aforementioned variation, BIC consistently generated a $k=3$ elbow. These results do indeed support the claim that the iris dataset contains three classes/species.

#####

Output Screenshots

```
In [563]: %run ~/Desktop/KimConger_Assignment6/RunKMeansClustering.py
```

```
~~~~~Part One~~~~~
```

```
KMeans run with elbow_k = 3 (found algorithmically)
Confusion Matrix
[[50.  0.  0.]
 [ 0. 48.  2.]
 [ 0. 14. 36.]]
accuracy score: 0.8933333333333333
```

```
KMeans run with k = 3
Confusion Matrix
[[50.  0.  0.]
 [ 0. 48.  2.]
 [ 0. 14. 36.]]
accuracy score: 0.8933333333333333
```

```
In [774]: %run ~/Desktop/KimConger_Assignment6/RunGMM.py
```

```
~~~~~Part Two~~~~~
```

```
GMM run with AIC elbow_k = 4 (found algorithmically)
Confusion Matrix
[[50.  0.  0.  0.]
 [ 0. 27.  0. 23.]
 [ 0. 15. 35.  0.]
 [ 0.  0.  0.  0.]]
# clusters > # classes, so conventional accuracy_score not available.
```

```
GMM run with BIC elbow_k = 3 (found algorithmically)
Confusion Matrix
[[50.  0.  0.]
 [ 0. 49.  1.]
 [ 0. 14. 36.]]
accuracy score: 0.9
```

```
GMM run with k = 3
Confusion Matrix
[[50.  0.  0.]
 [ 0. 50.  0.]
 [ 0. 14. 36.]]
accuracy score: 0.9066666666666666
```

