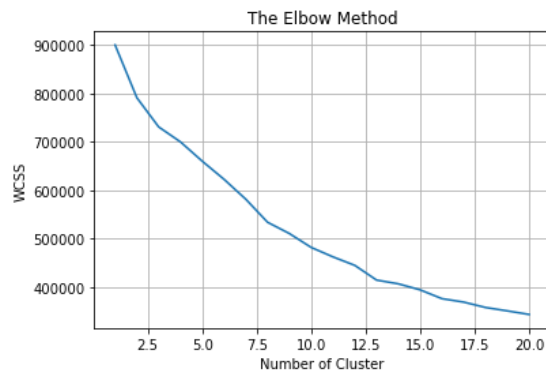


Experimentation on Hotel reservation dataset

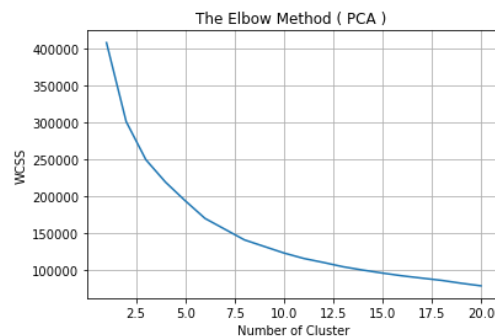
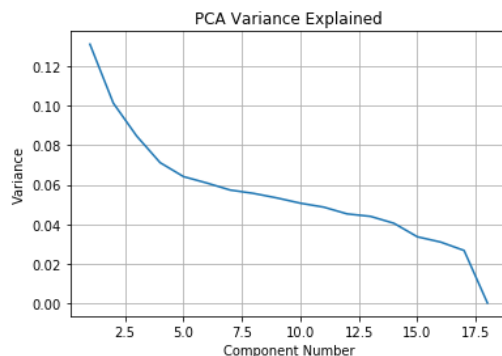
K-Means

From the chart below we can see that the optimum number of clusters for the dataset would be 8. Since there is a slight elbow formed. Hence without optimizing the dataset we get an optimum of 8 clusters.



PCA

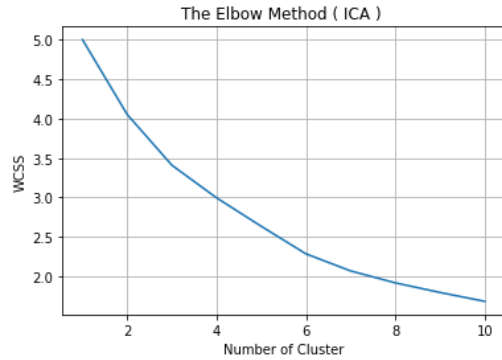
Now I've implemented PCA as part of the Dimensionality Reduction concept, PCA finds the maximum variance and the next orthogonal and finds the component which explains the second-best variance. From our SGEMM dataset, we could see that the variance of the dataset explained Hence I'm using first 6 components to pass it to the K – Means algorithm.



From the above graph it is seen that at 8 clusters the error has decreased by more than 5 times. Hence after implementing the PCA on K means the error has been well limited by the algorithm.

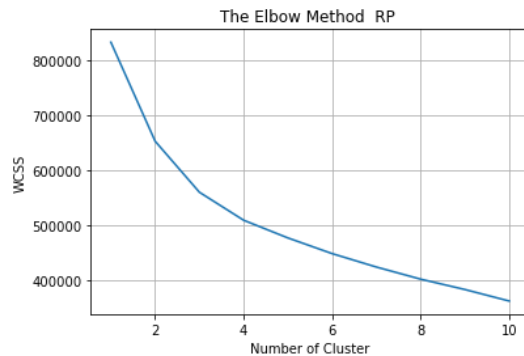
ICA

For this type of dimensionality reduction, we have set the number of components to be 6 and have developed the ICA to extract the features/signals which gives the best mutually independence. With this we will plot the K-Means. For 6 clusters we get a low error value of 2.25, which is better to the before methods.



Randomized Projections

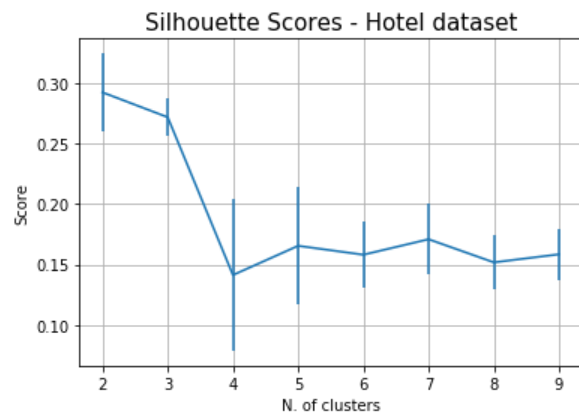
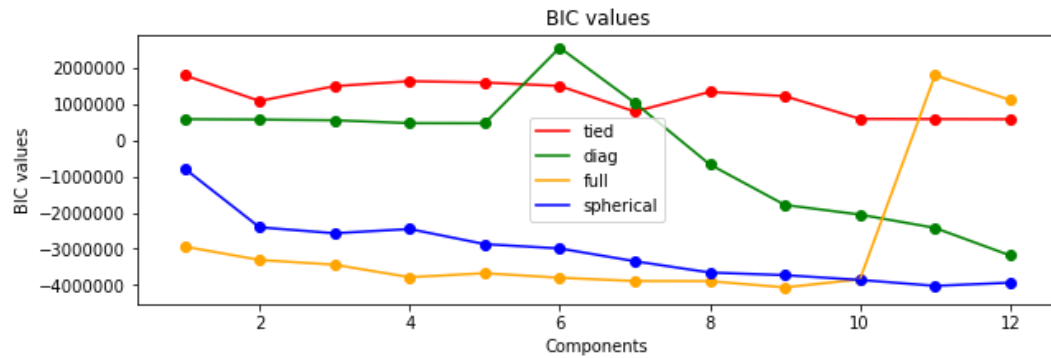
This module implements a simple and computationally efficient way to reduce the dimensionality of the data by trading a controlled amount of accuracy (as additional variance) for faster processing times and smaller model sizes. This module implements two types of unstructured random matrix: Gaussian random matrix and sparse random matrix.



Implementing Randomized projections to the dataset and applying K-means yields a better result than what was done without any algorithms. The error at 6 clusters is half what we found without any optimization.

Expectation Maximization

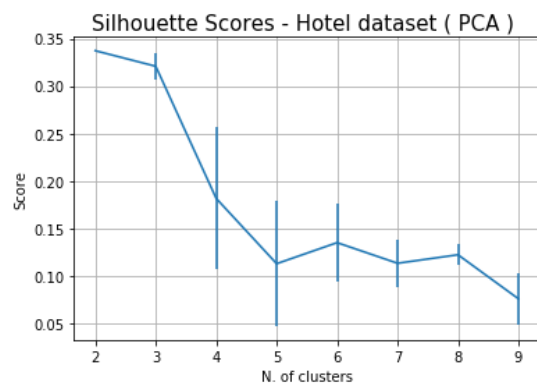
Initially I did the experimentation by identifying the covariance matrix which yields the best resultant model. Hence, I experimented with different covariance types and obtained the best type using the AIC and BIC values where the values peaked for the spherical model. The I performed K-means algorithm on the spherical model with Euclidian distance and obtained the following.



From the silhouette scores, the coefficient varies between -1 and 1. A value close to 1 implies that the instance is close to its cluster is a part of the right cluster. Whereas, a value close to -1 means that the value is assigned to the wrong cluster. we can see that 6 number of clusters has the optimum number but if we can reduce more to 3 clusters we get the best hence we can proceed between either of the models.

PCA

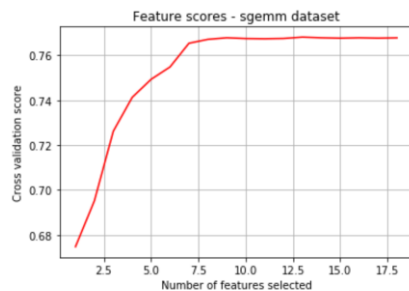
After Implementing PCA we see that the 6 number of clusters is better chosen for this dataset is also an optimum number of clusters that can reproduce the data as much as possible.



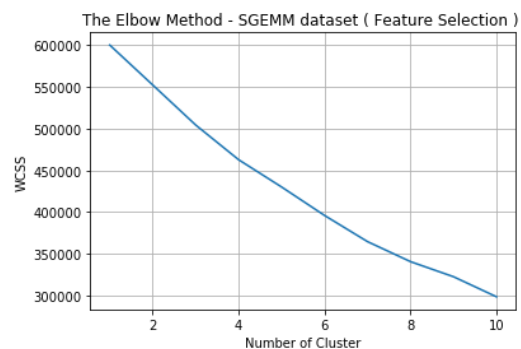
Feature Selection

Feature selection module can be used for feature selection/dimensionality reduction on sample sets, either to improve estimators' accuracy scores or to boost their performance on very high-dimensional datasets. In this module I've done feature selection on random forest algorithm and obtained the following results.

Optimal number of features : 13
[5 2 1 1 1 1 1 1 4 6 3 1 1 1 1 1 1]

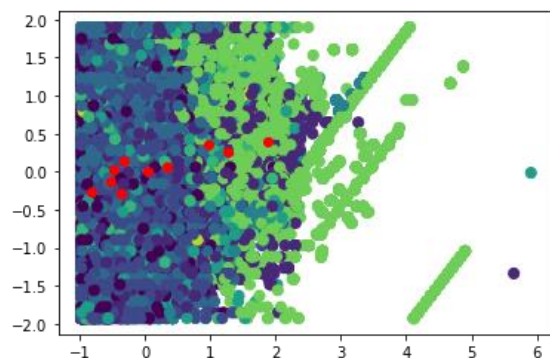


From the graph we can see that the minimum optimum number of features is 10 but the optimum number itself would be 13. But it gives a very high CV score around 0.78.



Now after Implementing feature selection the resultant graph shows us how the errors have reduced. But for just the K-means model the best results were given by PCA.

Best Clustering Model- FS



Applying ANN on the above models

Feature Scaling

When applying feature scaling on the neural network we get a reduced feature which are more significant than in the other models. I've uploaded the confusion matrix and accuracy values for this

```
The accuracy is
0.73392

The confusion matrix is:
[[6829 1006]
 [2320 2345]]
```

model.

Randomized Projection

When applying randomized projection on the neural network we get a reduced feature which are more significant than in the other models. I've uploaded the confusion matrix and accuracy values for this model.

```
The accuracy is
0.71112

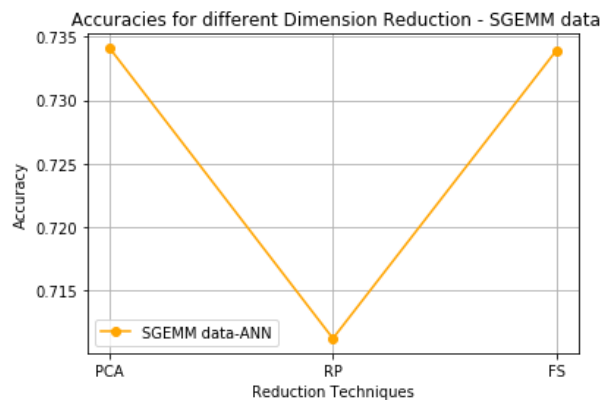
The confusion matrix is:
[[6590 1245]
 [2366 2299]]
```

PCA

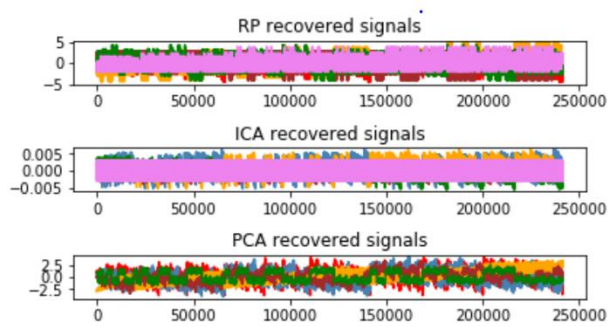
When applying PCA on the neural network we get a reduced feature which are more significant than in the other models. I've uploaded the confusion matrix and accuracy values for this model.

```
The accuracy is
0.734

The confusion matrix is:
[[6838 997]
 [2328 2337]]
```



Now when I plot the values on the graph, we can see that the feature scaling gives the highest accuracy followed by PCA. RP gives a poor accuracy hence we can consider the other two methods



The energy graph shows that RP has majorly three clusters are dominant energies. In PCA there are multiple signals which are dominant. The ICA is similar to the RP signal.