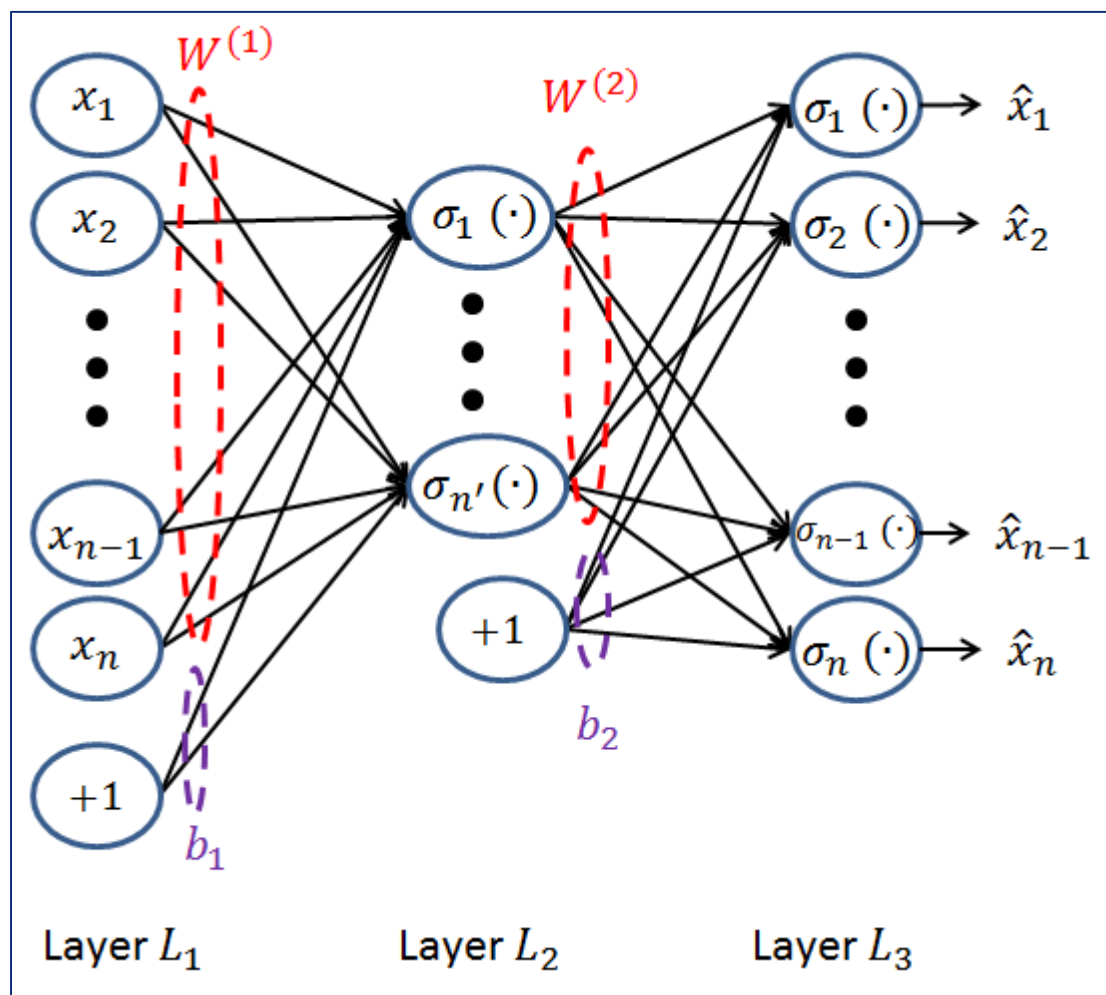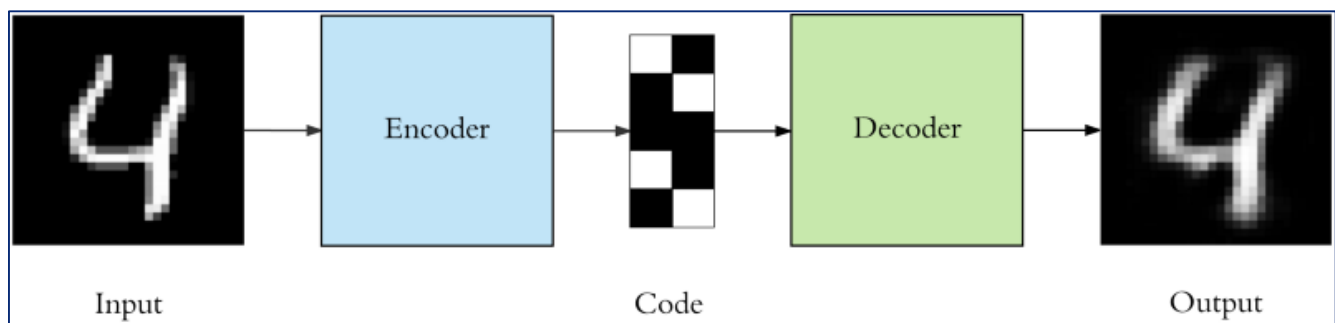# SMAI ASSIGNMENT- 9 REPORT

_____

MARCH 30

_____

AYUSH KUMAR DWIVEDI
2018802002
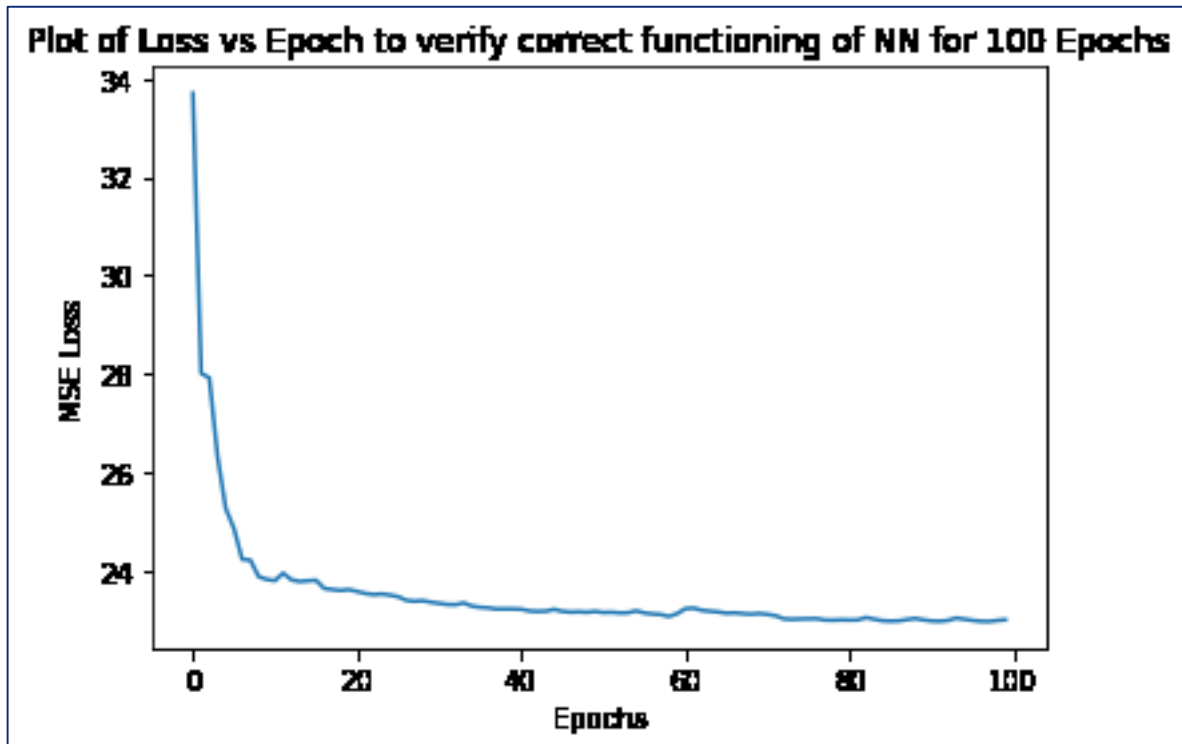
Autoencoder for Dimensionality Reduction

Structure of Autoencoder:

The following is the structure of autoencoder which I have implemented:

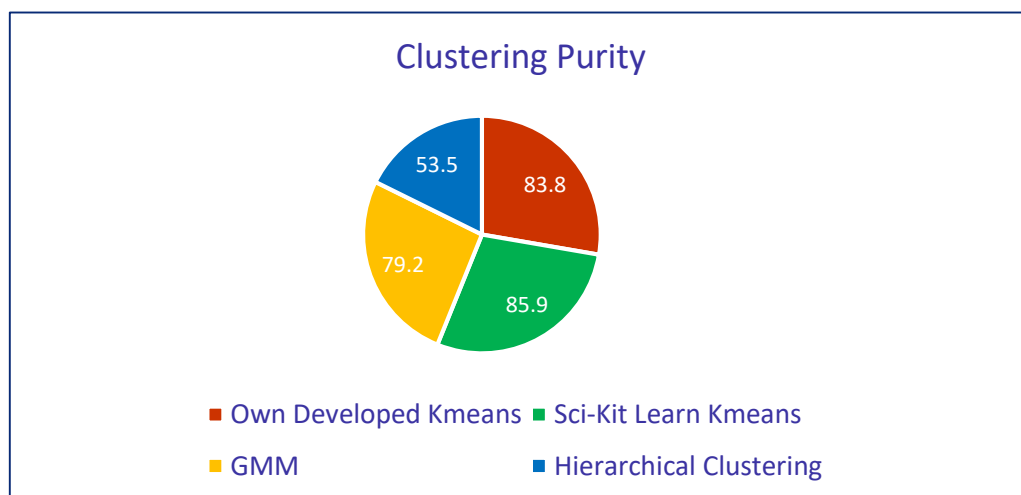# Checking if the Neural Network is Working properly or not ?

The following is the graph of Cost Function Vs No. of Epochs

I implemented 3 layer Autoencoder with the following type of activation functions:
- Linear, Sigmoid, ReLu, Tanh

The following are the results of purity in clustering thereafter:

```
Hyperparameters set to:

    Nuerons in Hidden Layer: 14
    Learning Rate: 0.001
    Number of Epochs: 10
    Activation Function: sigmoid


=================================================================

Autoencoder Running !

Data reduced to new dimention: 14


=================================================================

Doing Clustering by selecting reduced number of dimentions in PCA as
per threshold of 10%

Clustering Started !!!

Initial Centroids Set to:
[[ 24  36  35  98 167 140  46  87 232 102  73 182  92 150]
 [126 228 216 191 143 219 200 112 237 154  48  34  46 200]
 [ 38 224  51  64 186 142 128  38  93 209  77 111  32  21]
 [  3 132  92 144  16 133  78 213 188  92  39 136 137  94]
 [ 26 137 217 122 190  42  74 171  38 204 175 125  17  24]]

Distance moved by Centroids in next interation
[4.50651750e+02 8.03514748e+09 4.54804354e+02 4.41976244e+02
 4.93080115e+02]

Distance moved by Centroids in next interation
[ 8.24621125  0.         15.62049935  4.12310563 51.17616633]

Distance moved by Centroids in next interation
[22.60530911  0.         18.94729532  1.41421356 60.81118318]

Distance moved by Centroids in next interation
[21.14237451  0.          5.56776436  3.16227766  5.83095189]

Distance moved by Centroids in next interation
[12.28820573  0.          1.          2.64575131 33.3166625 ]

Distance moved by Centroids in next interation
[ 7.14142843  0.          0.          1.41421356 89.70507232]

Distance moved by Centroids in next interation
[ 2.23606798  0.          0.          1.         96.7367562 ]
```

```
Distance moved by Centroids in next interation
[  1.41421356   0.           1.           1.          117.69876805]

Distance moved by Centroids in next interation
[0. 0. 0. 0. 0.]

Clustering Completed !!!

================================================================
=======

Purity while using own developed KMeans:0.837947035762861

Purity while using KMeans from Sci-Kit Learn:0.8585886870949676

================================================================

GMM Started !!!

GMM Completed !!!

Purity while using GMM:  0.7917433394671574

================================================================

Hierarchical Clustering Started !!!
Hierarchical Clustering Completed !!!

Purity while using Hierarchical Clustering:  0.5346427714217138

================================================================
```

Kernel Density Estimation

I applied PCA to reduce data to three different dimensionalities as follows: 15, 25, 30
Later I applied Grid Search cross validation to obtained the best Bandwidth and then performed KDE on all three different dimensionalities and obtained the following results:

```
Digit Data Loaded !
The loaded data has the shape(1797, 64)


================================================================

Performing Grid Search Cross-Validation to optimize the bandwidth
Best Bandwidth using 15 features: 3.79269019073225


================================================================

Performing Grid Search Cross-Validation to optimize the bandwidth
Best Bandwidth using 25 features: 2.9763514416313175


================================================================

Performing Grid Search Cross-Validation to optimize the bandwidth
Best Bandwidth using 30 features: 2.9763514416313175


================================================================
```
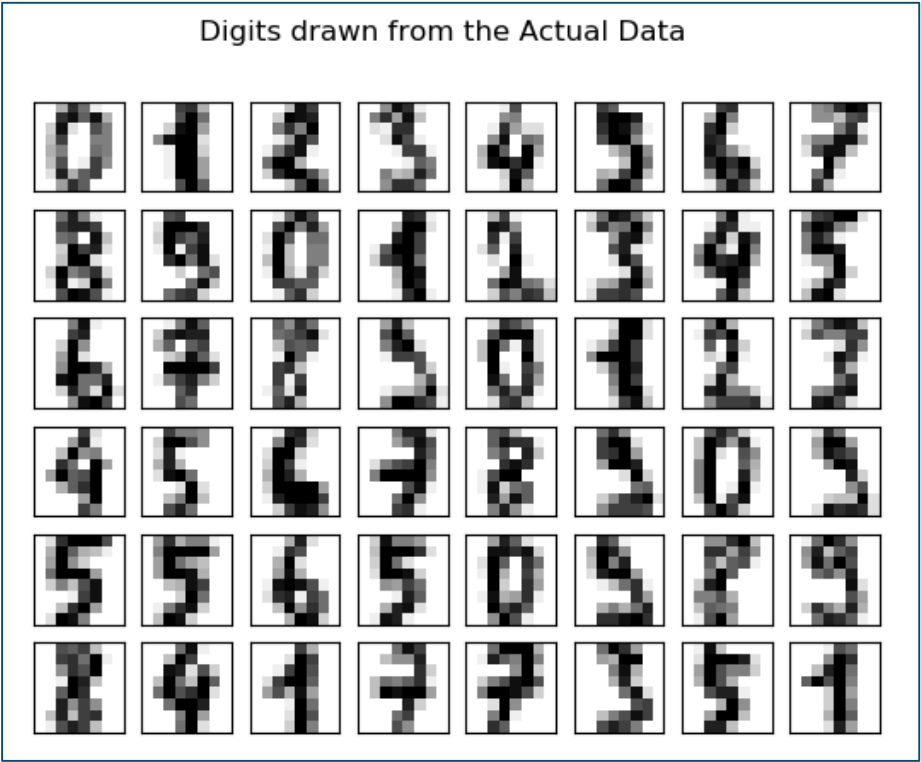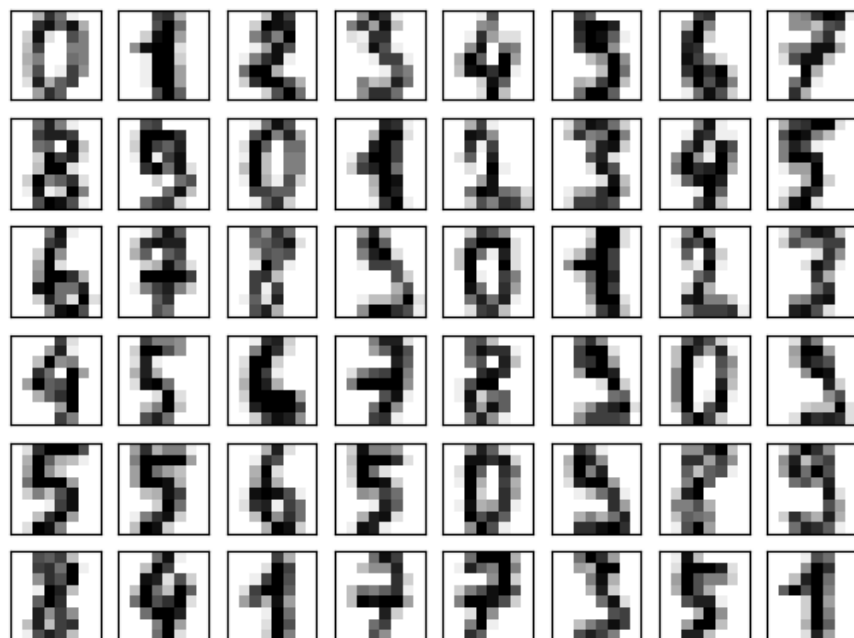
No. of PCA Reduced Dimension: 15

Digits drawn from the Actual Data

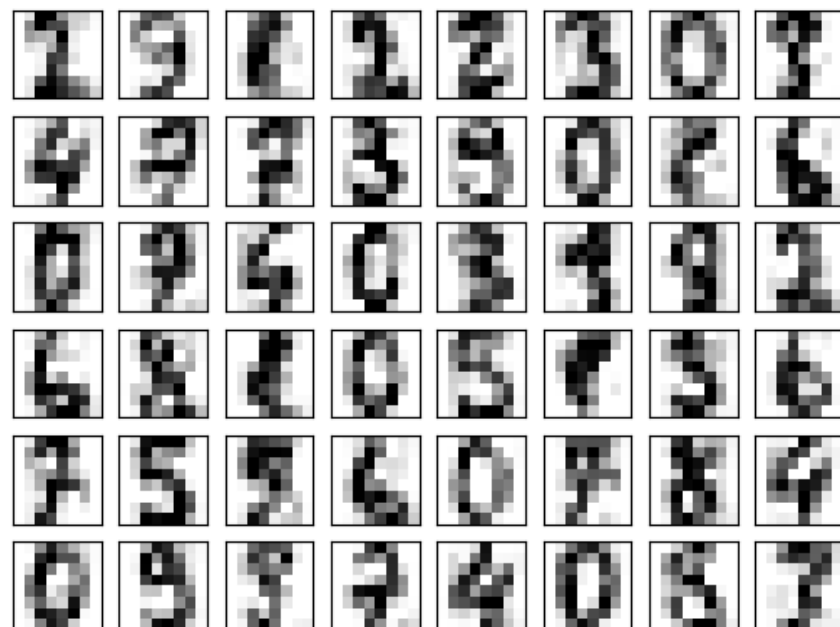Digits drawn from the Kernel Density Model
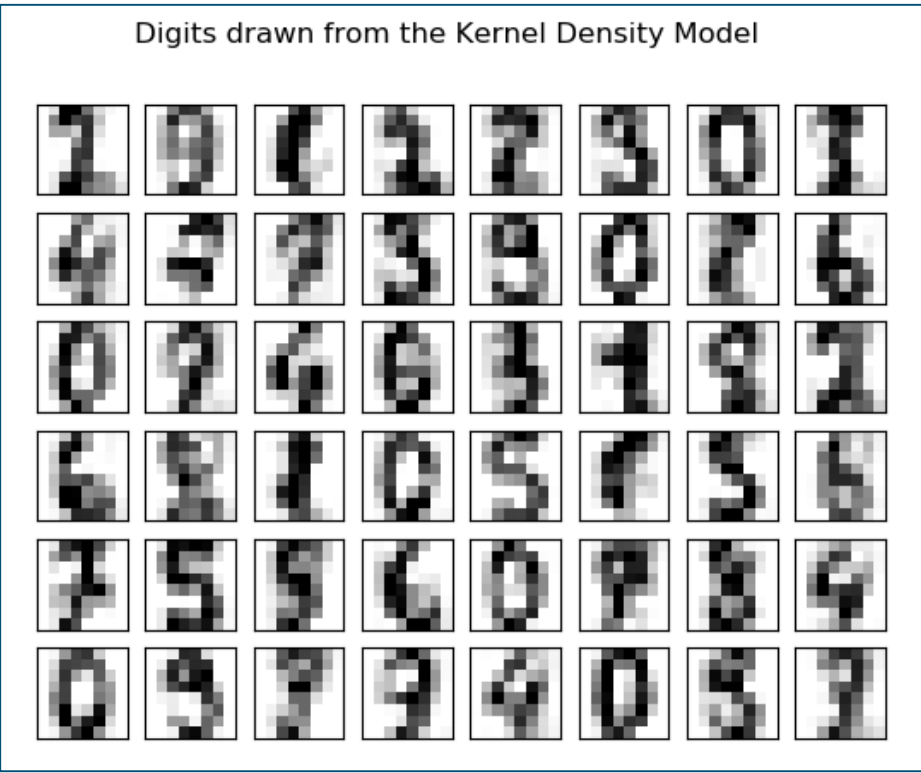
No. of PCA Reduced Dimension: 25
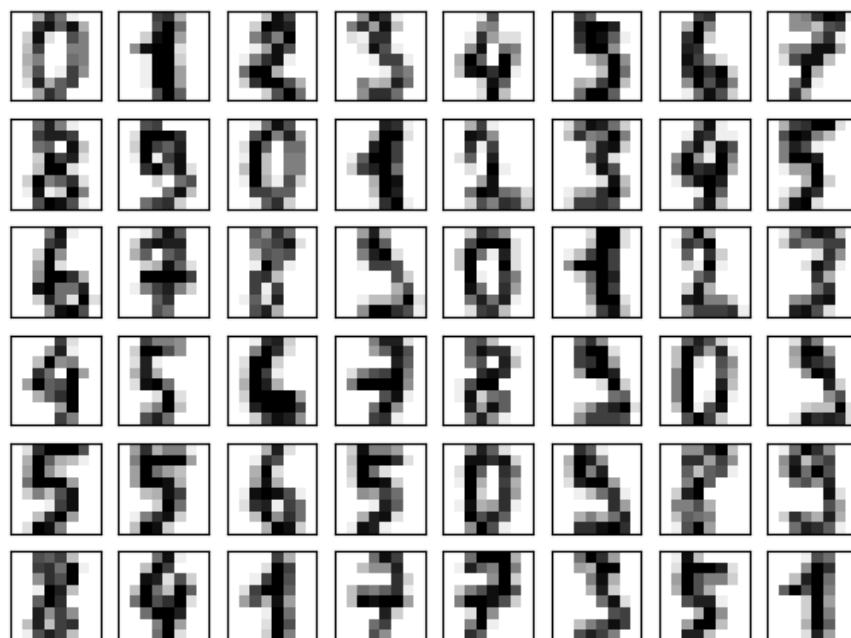
Digits drawn from the Actual Data

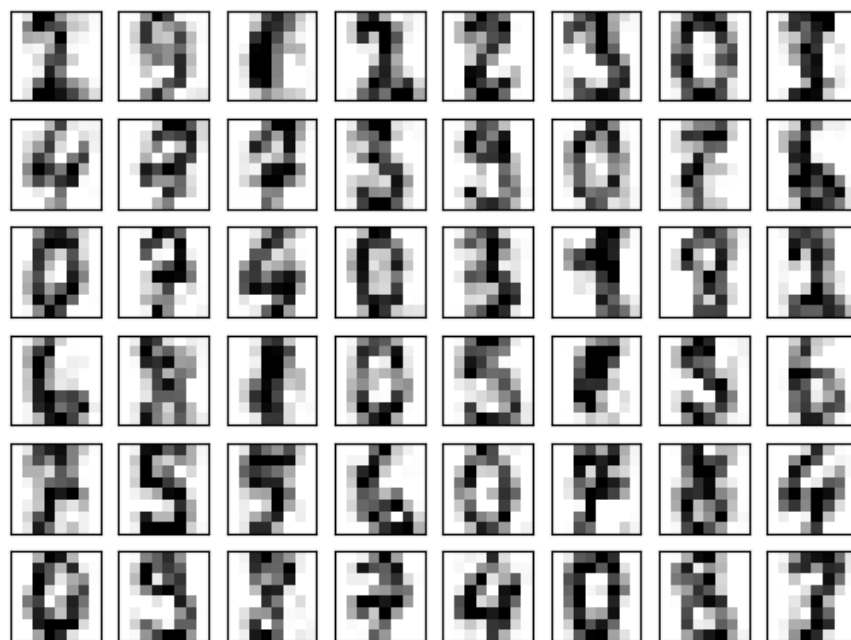Digits drawn from the Kernel Density Model

No. of PCA Reduced Dimension: 30

Digits drawn from the Actual Data

Digits drawn from the Kernel Density Model

GMM Density Estimation

I applied PCA to reduce data to three different dimensionalities as follows: 15, 25, 30 Later I applied Bayesian Information Criteria(BIC) to select the number of component and then performed GMM Density Estimation on all three different dimensionalities. The following are the obtained results:

```
Digit Data Loaded !
The loaded data has the shape(1797, 64)

================================================================

Selecting number of components for GMM using Bayesian Information
Criteria(BIC)
No. of components selected using 15 features: 190
GMM Converged: True

================================================================

Selecting number of components for GMM using Bayesian Information
Criteria(BIC)
No. of components selected using 25 features: 130
GMM Converged: True

================================================================

Selecting number of components for GMM using Bayesian Information
Criteria(BIC)
No. of components selected using 30 features: 110
GMM Converged: True
================================================================
```
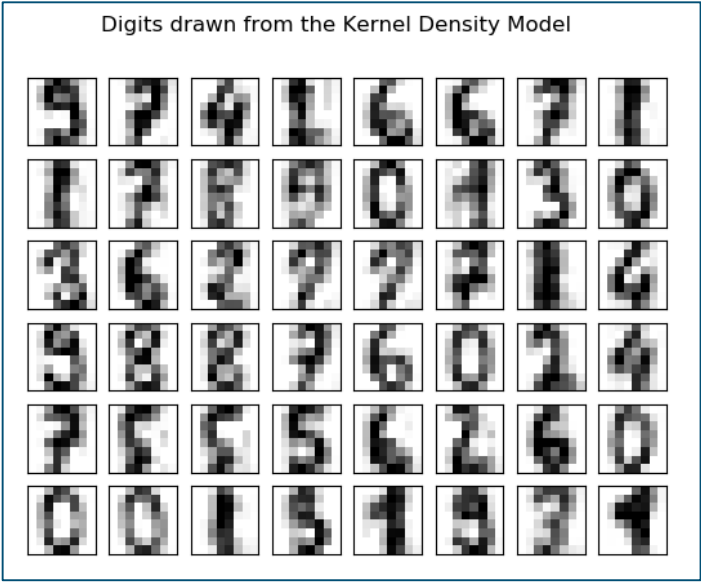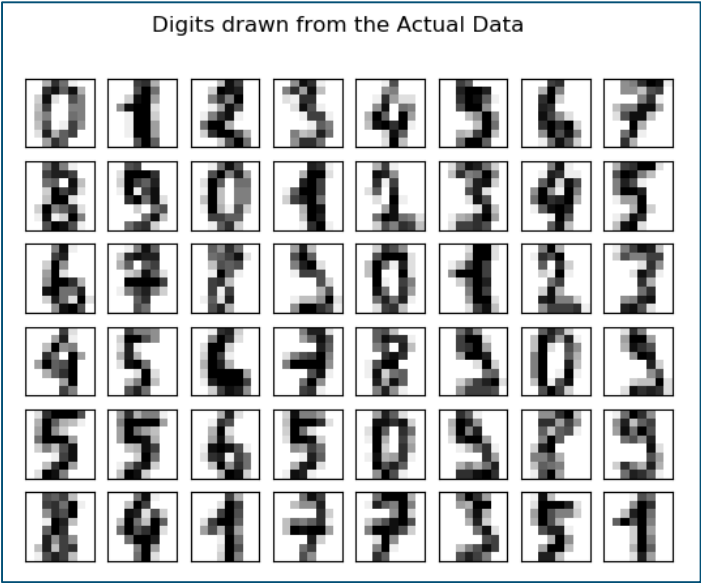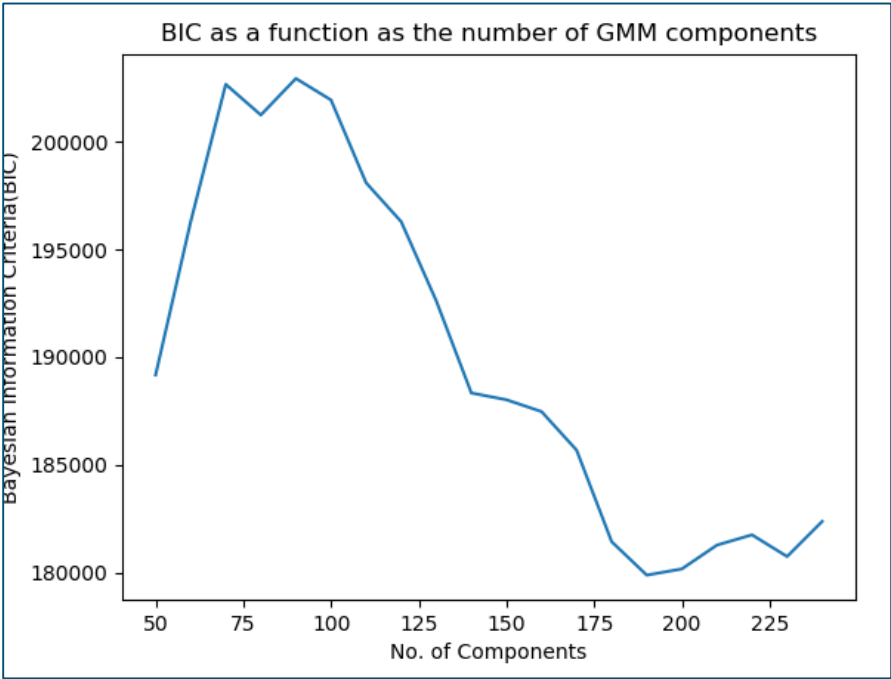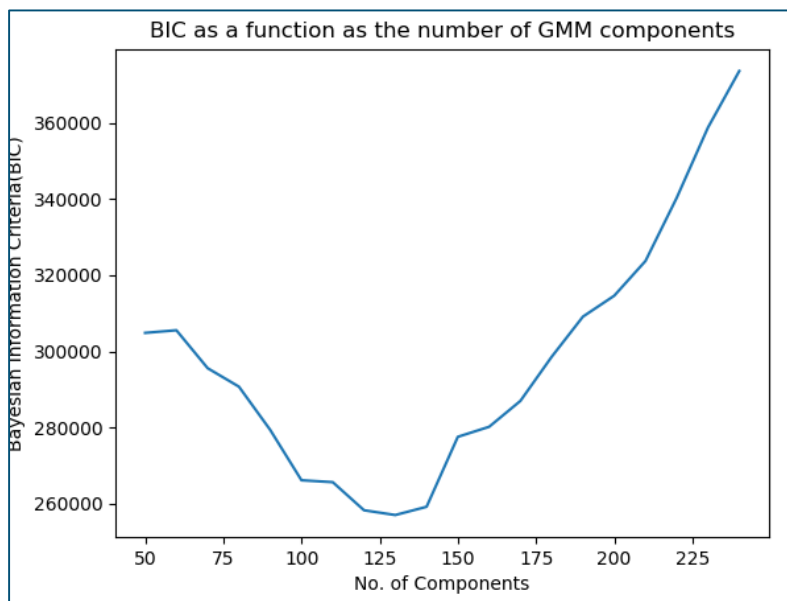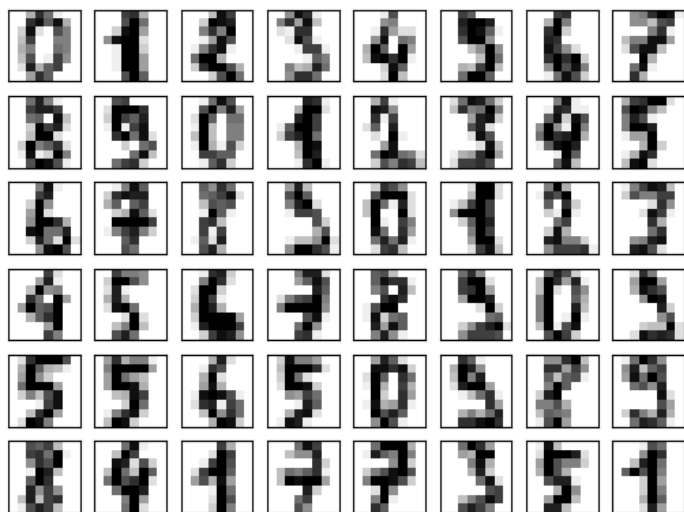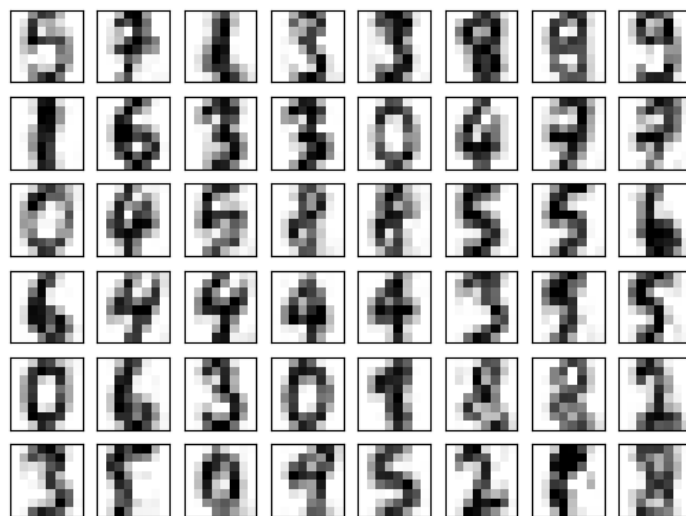
No. of PCA Reduced Dimension: 15

BIC as a function as the number of GMM components

Digits drawn from the Actual Data

Digits drawn from the Kernel Density Model

No. of PCA Reduced Dimension: 25

BIC as a function as the number of GMM components

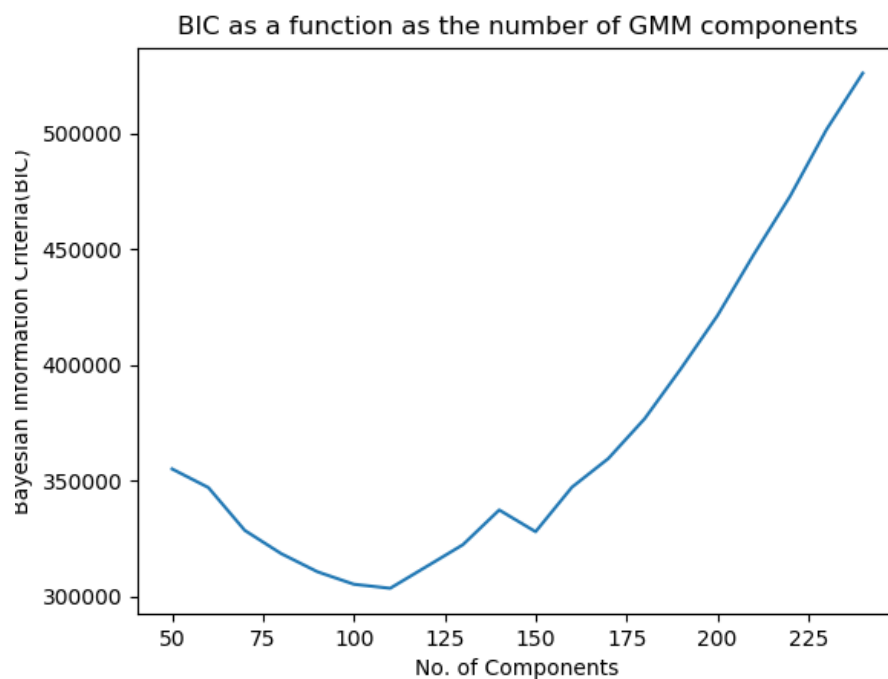Digits drawn from the Actual Data

Digits drawn from the Kernel Density Model

No. of PCA Reduced Dimension: 30

BIC as a function as the number of GMM components

Digits drawn from the Actual Data

Digits drawn from the Kernel Density Model