(a.) Clustering is form of Unsupervised Learning, where labels of dataset are not given.

Agglomerative Hierarchal Clustering, is a greedy approach of merging points in to one cluster. Its steps are as below:

```
1) Given set of Data Points $\{x_1, x_2, x_3, \dots, x_n\}$
2) Select a measure function such as Euclidean Distance.
3) For i=1 to n:
        a. Make each data point as cluster.
4) Cluster set, C= $\{c_{1,}c_2, c_3, \dots , c_n\}$
5) While Cluster Size > 1 do
        - Calculate if $(c_{min1}, c_{min2})$=minimum Distance between $c_i$ & $c_j$ for all $c_i. c_j\ in\ C$
              ○ Add $\{c_{min1}, c_{min2}\}$ to Cluster set C
```

Strengths: No prior information of dataset required.

Weakness: Very High time complexity, so if number of data points are large it will take huge amount of time.

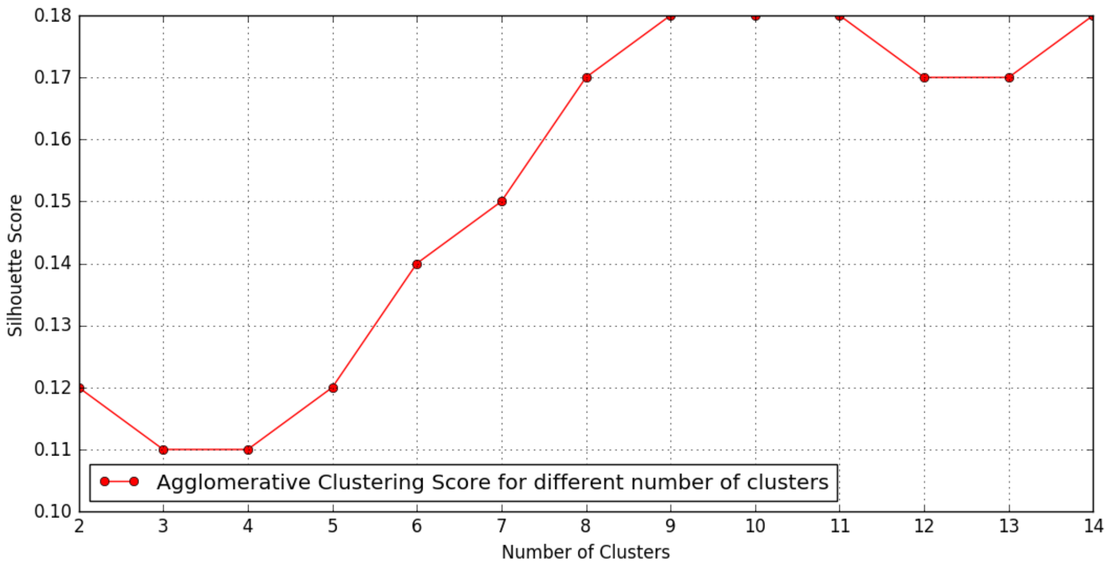
b. **Silhouette Score** can be defined as measure of similarity of a data point to its Cluster, as compared to other Clusters. It can be defined as below:

$s = \frac{b-a}{\max(a,b)}$ where b, average dissimilarity of I with all other data within same cluster.

a, lowest average dissimilarity with neighbor cluster.

If s value is close to 1 that means data is appropriately clustered.

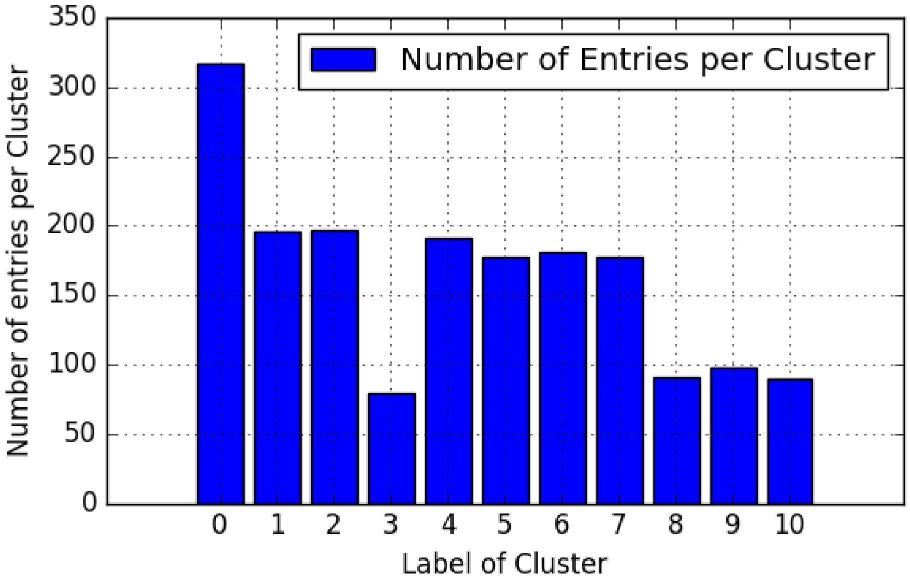
c.



d. In certain situations, there is no effect of adding the number of clusters. Elbow method is one way of determining the optimal number of clusters by plotting sum of squared distances vs Number of clusters, which determines at what point optima of number of clusters has been achieved.

In this case instead of Sum of Squared Euclidean distance, Silhouette Score had been considered. If we look at plotting, there's a clear stability can be seen from 9-11. So, Basically if number of Clusters are either 9,10 or 11. It will give same level of accuracy.

e.



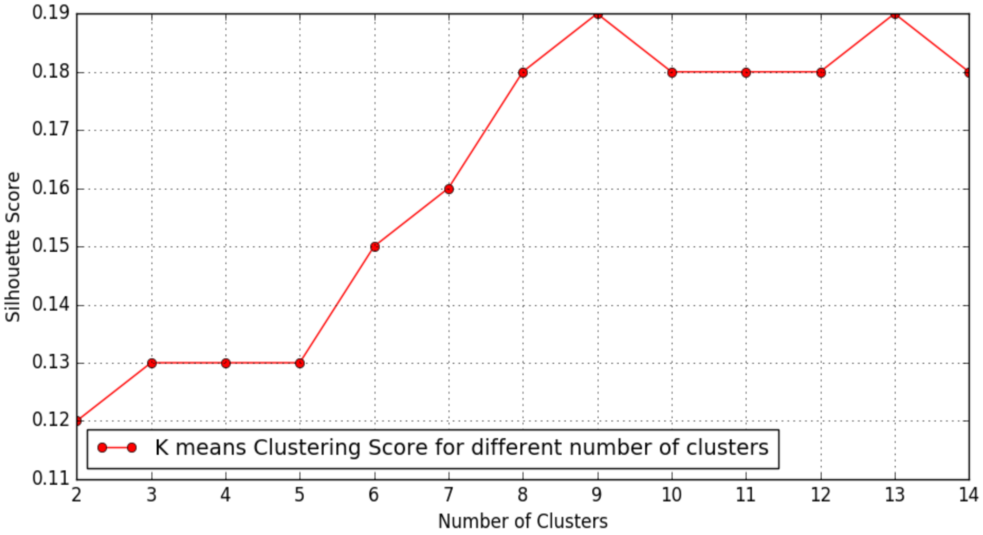
f. Optimal Value Achieved is 9, Its because certain digits have almost similar shape and size for example 9 looks like 4 in many cases which might led algorithm to believe that there are only 9 possible classes instead of 10.

g. **K-means:** K means Clustering works on objective of minimizing the Euclidean distance of every point to its cluster centroid. It follows given steps:

1) Place K points randomly in space as Cluster centroid points.
2) Calculate distance of each point to the Cluster Centroid point and assign them to cluster which is closest to it.
3) Recompute Centroid Cluster points by taking mean of all Cluster points.
4) Stop. When Centroid Points stop changing.

Distance Function: Euclidean(x,y)= $\sum(y-x)^2$

Centroid Point= $\frac{\sum_{i \in cluster\ points} x_i}{number\ of\ cluster\ points}$



Using K means we can see the minimal stability can be reached at 10,11 or 12. Although its minimum value is exactly equal to number of exact labels of dataset. But other points are outside range, thus Hierarchal Agglomerative Clustering is a better method.

2.

a. **Principal Component Analysis** is one of efficient Dimensionality Reduction Techniques, It transforms data points to new set of points(principal components), which can be defined as linear combination of original set. First Component is chosen which has large variance, Second Component is chosen and orthogonal always to first component.

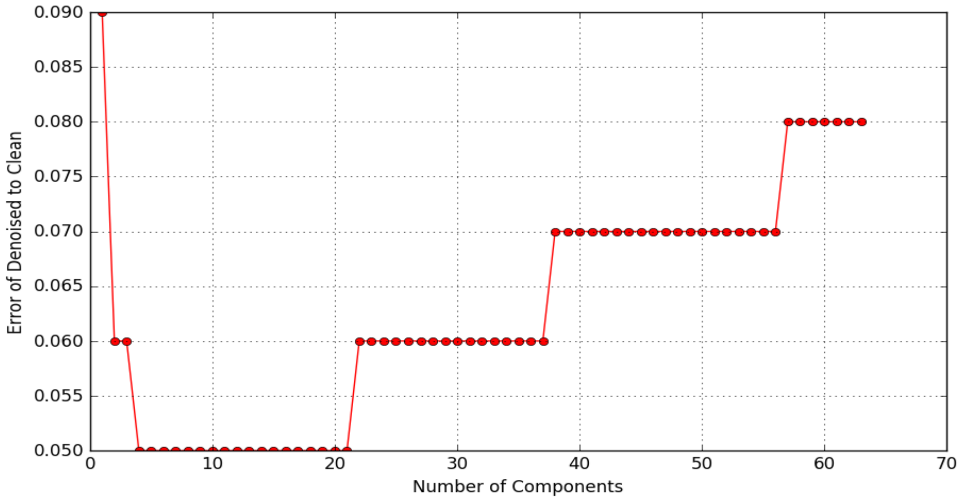
Algorithm of PCA is as follows:

1) Compute Covariance Matrix: $\Sigma = \frac{1}{m}\sum_{i=1}^{n}(x(i))(\,x(i))^T$
2) Compute Eigen Vectors of Covariance Matrix, Using Singular Value Decomposition.[n X n matrix]
   [U,S,V]=SVD(Σ)
   Reduced U=U[:][1:k]
   Z=Inverse(Reduced U)* x

Strengths: Non Parametric Technique. It works nicely if there is redundancy in data. It considers data as whole without taking account of labels.

Weakness: It has very High Computational Complexity of $O(D^2N + D^3)$, thus it is not suitable for large scale data.

b.



c. Principal component Analysis have several attributes but the important one is number of components to be kept, i.e; to what number of dimensions would you like to visualize data.

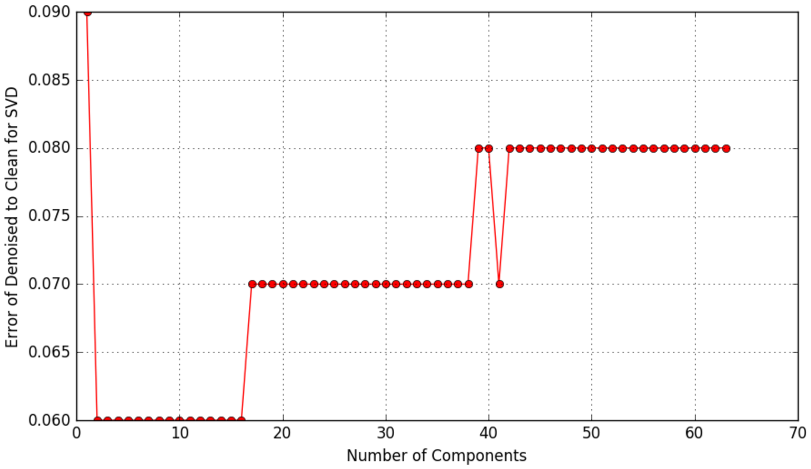
Over here **n_components** values has been varies from **0 to 64**, and obtained results show minimal error for range of components from **5 to 21**.

d. **Truncated SVD**: Its one of Linear Dimensionality Reduction Methods, by means of Singular Value Decomposition. It is almost similar to PCA but It doesn't centralized data before calculating matrix of singular value decomposition.

For Truncated SVD as well, number of components can be considered as hyperparameter.

Results obtained after varying Hyperparameter values is as follows:

Here Minimum Error is 0.6 which is greater than 0.5 of Principal component analysis. Thus PCA is best method for this dataset.


Error of Denoised to Clean for SVD
Number of Components