Annex

Appendix A DTW-based K-Means clustering

Given a univariate time series dataset, X is divided as $D = X_1, X_2, ..., X_n$ using a sliding window approach with a window size w. Then, these training samples are grouped into clusters $c = c_1, c_2, .., c_k$ using K-Means clustering with DTW distance for computing similarity between two time series. Initially, the process begins with 2 clusters and the corresponding Silhouette score [1] is computed. The clustering process is repeated for a number of clusters starting from a *min* value to some *max* value along with computation of the Silhouette score for every chosen number of clusters. Following the clustering process and score computation, the clusters and their centroids are plotted and carefully evaluated manually for different patterns and different range of mean values present in the existing data. A manual check for patterns reflected in the produced clusters ensures that prominent patterns, such as a rapid increase or decrease in trend, are adequately captured by the clustering process. Clustering process is terminated when the results obtained fulfill manually checked criteria and retrieve the collected clusters. Algorithm 1 describes the general procedure for DTW-based K-Means clustering.

Figure 1 visualizes the clusters obtained during the DTW clustering step. In the example shown in Figure 1, there are 7 clusters considered for the Reliance stock price data. We started by selecting 3 clusters that had a better Silhouette score. It was observed that the pattern represented by Cluster 5 in the figure was not captured. Choosing between 5 or 7 clusters did not vary much in terms of Silhouette score, but 7 clusters resulted in a compact range of values (Y-axis) that was desirable when assigning a new test point to a cluster and finding its nearest neighbors.

Algorithm 1: Algorithm for DTW based K-Means clustering

```
Result: Clusters
X = GetTrainingSet();
cluster, centroid, Silhouette_score = GetDTWBasedClusters(X, 2);
Repeat;
for ClusterNumber in range(min, max, step) do
    cluster, centroid, Score = GetDTWBasedClusters(X,
     ClusterNumber, Silhouette_score);
    plot(clusters, centroids);
    if ManualCheckOfClusters(clusters, centroids) then
        break;
    else if Silhouette_score != Score then
        Silhouette_score != Score;
        ClusterNumber -= step
end
return Clusters;
```

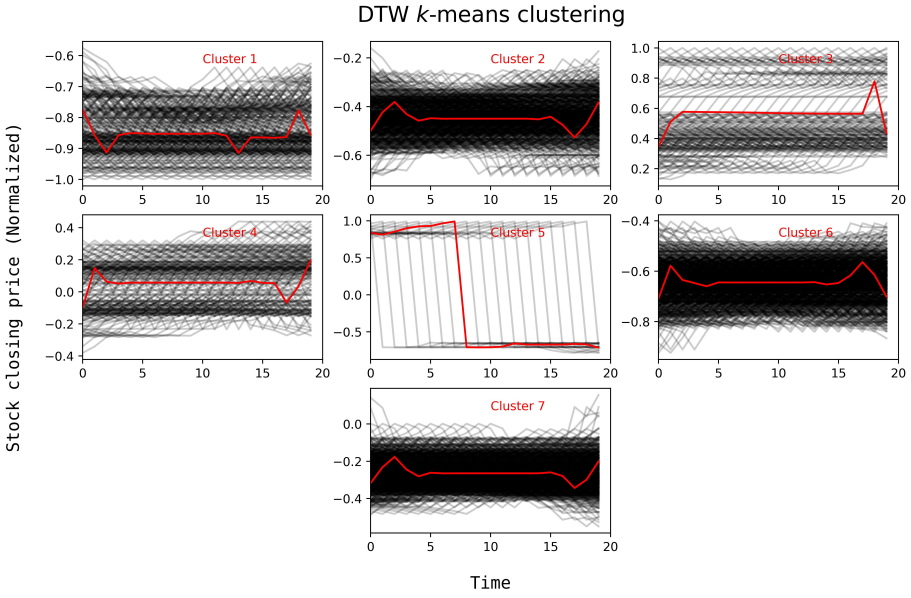

Figure 1: DTW based K-Means clustering for Reliance stock price data.

References

[1] Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, 1987.