CS 674 : HW1

Bhavika Tekwani (btekwani@gmu.edu)

1. Classification Accuracy

Dataset	1 NN with euclidean distance	1NN with Dynamic Time Warping	KNN and Dynamic Time Warping with 20% warping window	KNN with LB_Keogh Dynamic Time Warping
1	85.22	100	31.22	85.22
2	78.28			
3	51.65			
4	78.88	98.88	8.64	
5	99.54			

2. Analysis

As can be expected, 1 Nearest Neighbour gives the best performance across Euclidean distance and different versions of Dynamic Time Warping.

In the datasets given to us, the length of time series for each dataset is equal for the train and test sets. The complexity of Dynamic Time Warping is given by O(mn) where m and n are the lengths of the two time series. Since m = n in our case, the complexity of dynamic time warping will be O(Lm^2) where L is the number of time series instances in the dataset.

The computation proved to be very expensive for dynamic time warping - both with and without a warping window. Infact, on datasets 2, 3 and 5, DTW is still running.

We use the distance matrix from DTW to find its nearest neighbours and consequently the class labels. The complexity of k-Nearest Neighbours is O(nk+nd) where k is the no of neighbours, O(nd) is the time required to compute distances to all data points. O(nk) is the time to find k closest neighbours.

Our implementation of kNN and 1NN is relatively fast and DTW proves to be the only bottleneck. We run kNN multiple times with different values of k to observe the change in classification accuracy along with k. The value of k may vary from 1 to $\sqrt{n}$ where n is the number of samples in the test set. Therefore we vary the range for k depending on the dataset we're currently analyzing.

The results are presented below:

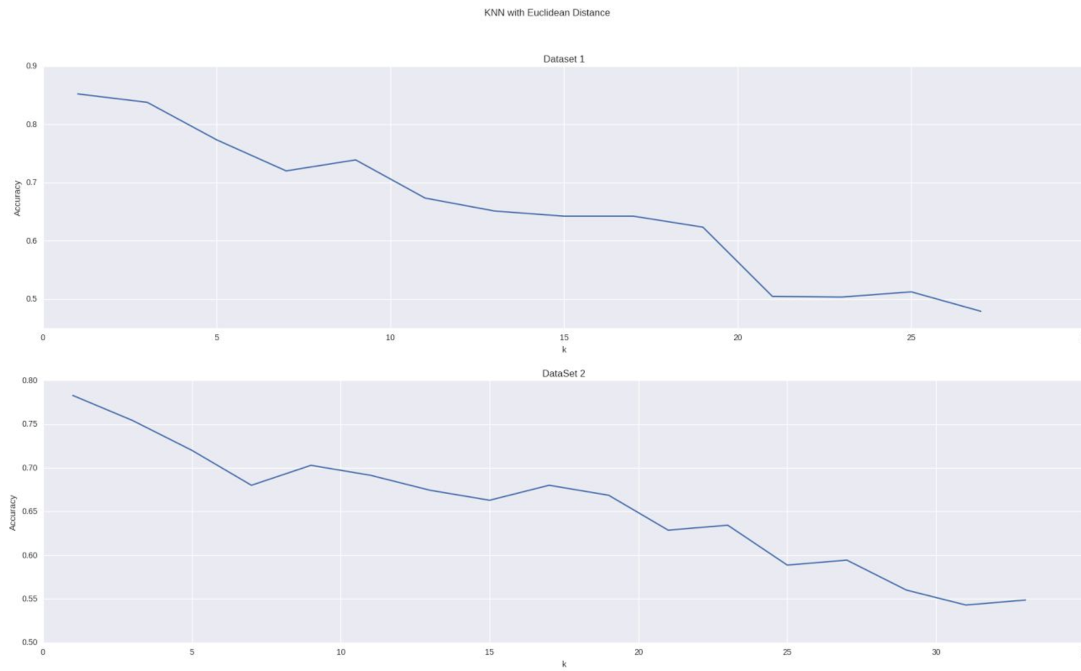

Fig. 1 a) KNN with euclidean distance on datasets 1 & 2

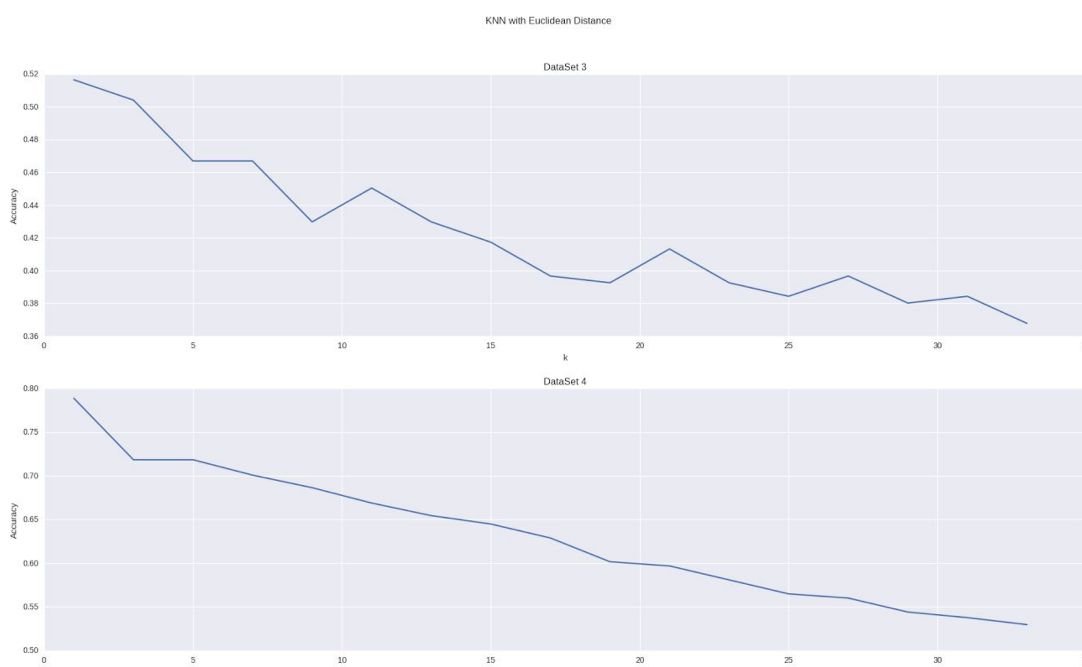

Fig. 1 b) KNN with euclidean distance on datasets 3 & 4

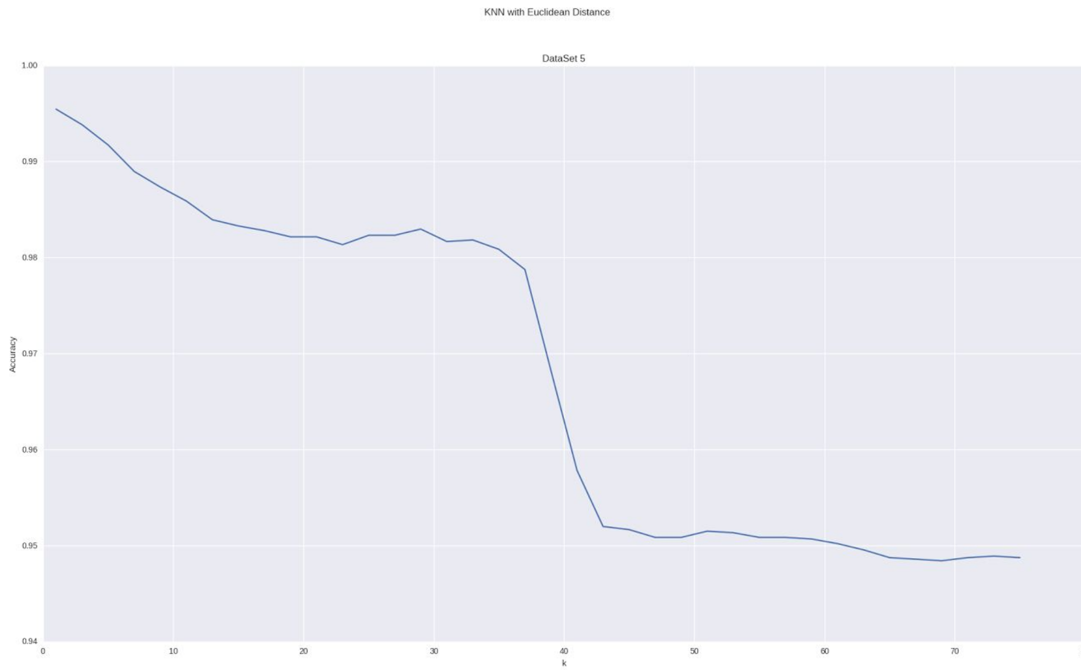

Fig 1 c) KNN with euclidean distance on dataset 5

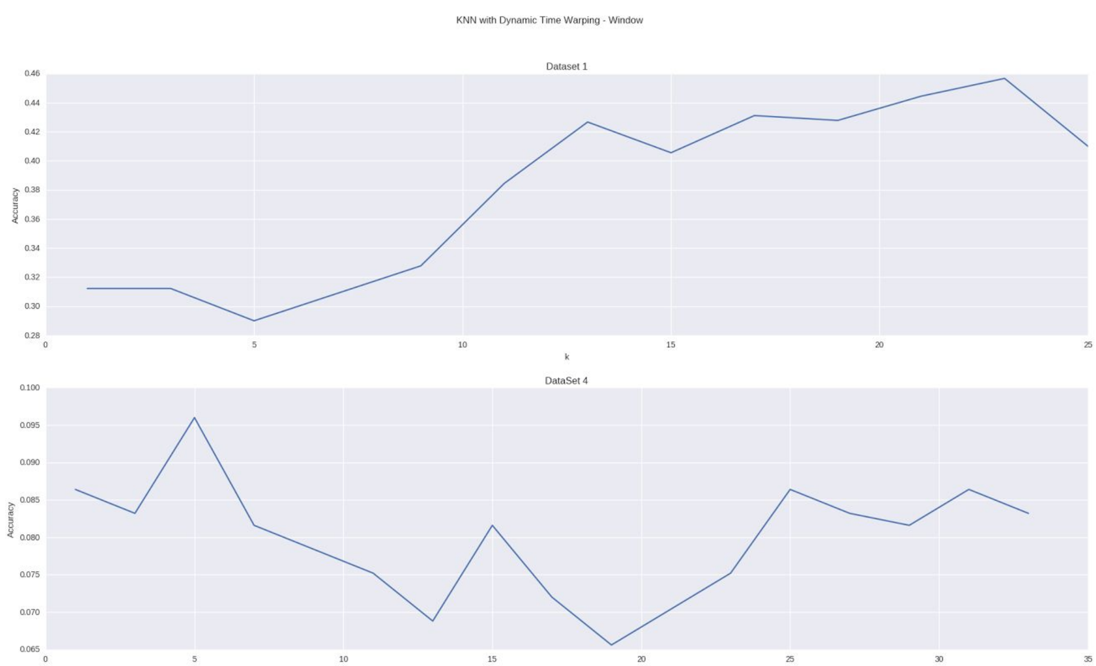

Fig. 2 KNN With Dynamic Time Warping - Window - datasets 1 and 4

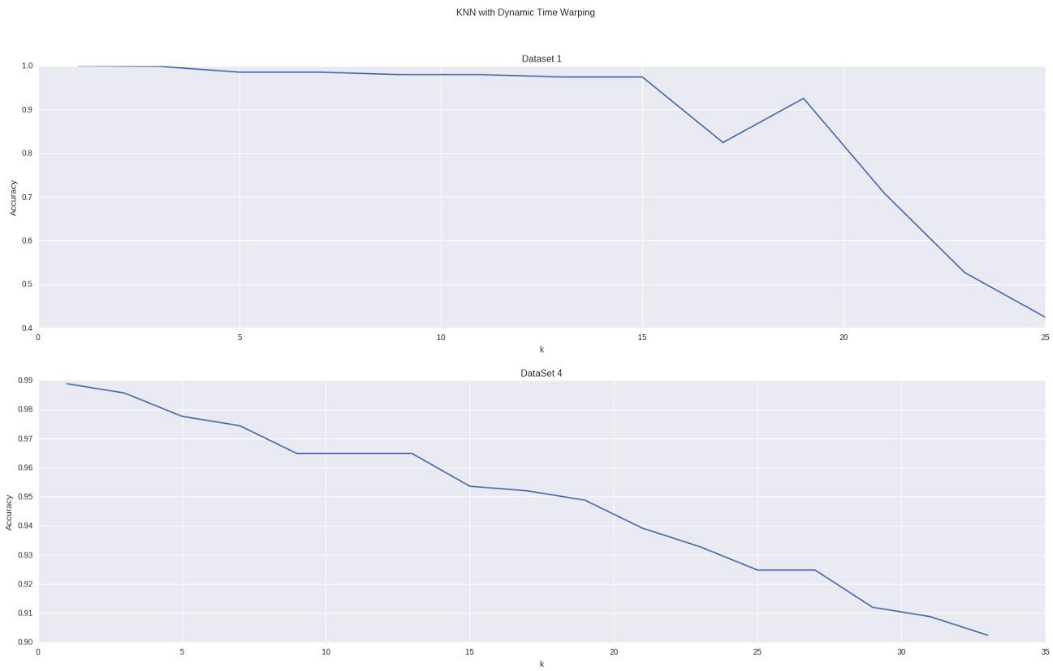

Fig. 2 KNN With Dynamic Time Warping - datasets 1 and 4

As we can see from Fig 1 a, b and c, the best accuracy is achieved when k = 1 for time series data. Additionally, dynamic time warping with a 20% warping window does not seem to give great results. In fact, simple DTW seems to be doing much better on datasets 1 & 4.

Improvement: LB Keogh

We implement LB_Keogh to speed up DTW to attempt an improvement over DTW. On dataset 1, we achieve an accuracy rate of 85.22% and it is on par with Euclidean distance and 1NN.

LB Keogh has been known to improve DTW performance to O(n) running time. We are unable to reproduce this result or comment on the improvement in the running time as it hasn't finished executing on datasets 2 through 5 as yet.

Pseudocode for LB_Keogh:

function LB_Keogh(ts1, ts2, r)

    LB_Sum = 0

    for idx, elem in ts2

        upper_bound = max($ts2_{i-r}$, $ts2_{i+r}$)

        lower_bound = min($ts2_{i-r}$, $ts2_{i+r}$)

    if elem > upper_bound

        LB_Sum = LB_Sum + $(\text{elem} - \text{upper\_bound})^2$

```
else if elem < lower_bound
        LB_Sum = LB_Sum + (elem - lower_bound) ^2

return LB_Sum ^0.5
```

Pseudocode 1-NN DTW:

```
function knn(ts1, ts2, w):
    Test_labels = []
    best_so_far = infinity
    for all in ts2:
        nearest_neighbour = []
        for all in ts1:
            If LB_keogh(j, k, w) < best_so_far:
                D = dtw (k, j)
                If D < best_so_far:
                    best_so_far = D
                    Add train example to nearest_neighbour
                    label = truth label for current train example
        Add label to Test_labels
    return Test_labels
```

References

[1] Ralanamahatana, Chotirat Ann, Jessica Lin, Dimitrios Gunopulos, Eamonn Keogh, Michail Vlachos, and Gautam Das. "Mining Time Series Data." In *Data Mining and Knowledge Discovery Handbook*, edited by Oded Maimon and Lior Rokach, 1069–1103. Boston, MA: Springer US, 2005. http://dx.doi.org/10.1007/0-387-25465-X_51.

[2] Rakthanmanon, Thanawin, Bilson Campana, Abdullah Mueen, Gustavo Batista, Brandon Westover, Qiang Zhu, Jesin Zakaria, and Eamonn Keogh. "Searching and Mining Trillions of Time Series Subsequences under Dynamic Time Warping." In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 262–270. ACM, 2012. http://dl.acm.org/citation.cfm?id=2339576.

[3] Ratanamahatana, Chotirat Ann, and Eamonn Keogh. "Three Myths about Dynamic Time Warping Data Mining." In *Proceedings of the 2005 SIAM International Conference on Data Mining*, 506–510. SIAM, 2005. http://epubs.siam.org/doi/abs/10.1137/1.9781611972757.50.

[4]"The LB_Keogh Page." Accessed February 22, 2017. http://www.cs.ucr.edu/~eamonn/LB_Keogh.htm.