Dynamic Time Warping

Note: Read entire document first and start coding

In this lab, we are going to do digit recognition from spoken recordings using DTW.

Data details:
You are given 300 spoken digits: 30 utterances for each digit.
Each digit is spoken by 10 times by each of the speakers Jackson, Nicolas, Theo.
Notation: 0_nicolas_2.mat means second instance of digit 0 when speaker is Nicolas
Each mat file contains M X 13 matrix. M denotes number of frames that utterance has and 13 is dimension of each frame.

Task:
You need to take each utterance (call it test utterance, T) and align with all other (299) utterances (call it set of reference utterances). For each alignment (T vs one of reference utterance(R)), you will get one number (call it alignment cost) which denotes how much dissimilar T and R are, i.e., the higher the number the more different those two utterances. Lower value denotes that both of those utterances are more similar

So for each utterance (T), you will have 299 values. Sort them and take least 29 values (because you have 29 utterances of same digit). Now, check if the utterances corresponding to those 29 values are of same digit and note down number of comparisons that yielded the same digit as T.

Repeat the above step for all the utterances and compute accuracy for each digit.

DTW: Follow below procedure to compute alignment cost between T and R
Let M X 13 is size of representation of utterance T and N X 13 is for R.

1) Compute dissimilarity matrix(S) between T and R using Euclidean distance. This will be M X N matrix
2) Compute accumulated DTW matrix (D) from the above dissimilarity matrix. To compute accumulated value at a given position, consider 3 previous elements i.e., for (m,n) position consider (m-1, n), (m, n-1) and (m-1, n-1) elements
3) So cumulative value at (m,n) is calculated as D(m,n) = S(m,n) + min(D(m-1, n), D(m, n-1), D(m-1, n-1))
4) At each step, keep track of the path (Store the index of the minimum) i.e., how D(m,n) is calculated, what is the minimum of D(m-1, n), D(m, n-1) or D(m-1, n-1) ?
5) Do this iteratively until you calculate D(M,N) element in D matrix.
6) Now, do backtracking through those stored indices until (0,0) position in D matrix to get optimal path. Optimal path is the one with minimum cost
7) Add all the dissimilarity values along the optimal path and normalize with length of the path

8) Now, this is the number that denotes the alignment of T and R

Compute confusion matrix. Confusion matrix will have information of how many times each digit is confused with other digits. Display it by the end of the notebook.

Deliverables:

If you use Python, use the provided Jupyter Notebook template for all problems, which is also available at following link:

https://colab.research.google.com/drive/1Y_KIgKz8k6wCJfK8a5LZQA0iyG4fkQwU