Noushin Islam Student No. 201508438

Assignment 1 – COMP336

Part 2 Report

Assignment Brief: Come up with an idea for a measure that can quantify how much any two given trajectories differ from each other. To simplify matters you can interpret (longitude, latitude, altitude) as a (x, y, z) point in 3D space and calculate the distance between two such points using the standard Euclidean distance.

<u>Task 7</u>

Brief: Design a formula to calculate the dissimilarity between two day-trajectories.

Symbols:

Timestamp	t
Longitude	x
Latitude	y
Altitude	z
Trajectory	T
Euclidean Distance	d_{eucl}
Trendline	L
Dissimilarity measure	D

Requirements:
d(T,T')=d(T',T) for any trajectories T and T' (symmetric)
d(T,T')=0 if and only if T=T'

Steps:
What we are essentially trying to find is a formula to measure how different two trajectories are from 0 (same) to ∞ (different).
The Euclidean distance, is a dissimilarity measure that has the following properties:

- $d_{eucl}(p,q) = 0$ if and only if p = q (non-degenerated)
- $d_{eucl}(p,q) >= 0$ for all p and q (positive)
- $d_{eucl}(p,q) = d_{eucl}(q,p)$ for all p and q (symmetric)
- $d_{eucl}(p,r) <= d_{eucl}(p,q) + d_{eucl}(q,r)$ for all p, q and r (triangle inequality)

Thus, we can derive that if we use the Euclidean distance for our formula, then two of the given requirements are met, such that it's symmetric, because regardless of whether $d_{eucl}(p,q)$ or $d_{eucl}(q,p)$ the result will be the same. Also, if p=q then d=0 as they are the same and there is no such thing as dissimilarity or distance then.
The Euclidean distance between two points p and q is calculated with the following formula:

$$d_{eucl}(p,q) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}$$

For simplicity matters, we will assume that the number of observations and the timestamps are the same.
To find the Euclidean Distance of two points from our trajectories T and T' with values <x,y,z> and the same timestamp we have the following formula:

$$d_{eucl}(T,T') = \sqrt{\sum_{i=1}^{n}(T_i - T_i')^2} \qquad d_{eucl}(T,T') = \sqrt{(x'-x)^2+(y'-y)^2+(z'-z)^2}$$

This gives us so far only the dissimilarity measure for data from a specific timestamp, however, to calculate the overall dissimilarity measure between T and T' we can do so by first multiplying all Euclidean distances that have been calculated for each timestamp. Following by taking the squared root of the combined distance results in the overall dissimilarity D:

$$D = \sqrt{d_{eucl}(T_1,T_1') \cdot d_{eucl}(T_2,T_2') \cdot \ldots \cdot d_{eucl}(T_{k-1},T_{k-1}') \cdot d_{eucl}(T_k,T_k')}$$

The letter k represents the number of timestamps at which there is data for both T and T'. If the trajectories are the same, such that T=T', then every $d_{eucl}(T,T')$ calculated for each timestamp t will result in 0, thus the

overall dissimilarity D will result in 0 too (0 by the power of any number results in 0 as well as the square root of 0 is 0).
The space complexity of this calculation in 3D space is O($ß^3$).

References:
D'Urso, P. (2000). Dissimilarity measures for time trajectories. Journal of the Italian Statistical Society, [online] 9(1-3), pp.53–83. doi:10.1007/bf03178958.

Sharma, P. (2020). Distance Metrics | Different Distance Metrics In Machine Learning. [online] Analytics Vidhya. Available at: https://www.analyticsvidhya.com/blog/2020/02/4-types-of-distance-metrics-in-machine-learning/ [Accessed 8 Nov. 2022].

Lee, W.-J., Duin, R., Alessandro Ibba and Loog, M. (2010). An experimental study on combining Euclidean distances. [online] ResearchGate. Available at: https://www.researchgate.net/publication/224183279_An_experimental_study_on_combining_Euclidean_distances [Accessed 9 Nov. 2022].

Task 8

Brief: Prove whether the formula is a distance metric, i.e., satisfies the triangle inequality ($d(T_1,T_2) <= d(T_1,T_3) + d(T_3,T_2)$ for any trajectories T_1, T_2 and T_3)?

Steps:
Triangle inequality in this case means that the distance from T_1 to T_3 via T_2 cannot be shorter than the distance between T_1 and T_3.
The Euclidian distance is a distance in a metric space and obeys all the properties of such including triangle inequality, however, only the squared Euclidean distance satisfies triangle equality.

Proof:

On $\mathbb{R}^n$ $\quad d(T_1,T_2) = \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2}$

satisfies triangle inequality.

Let $T_1, T_2, T_3 \in \mathbb{R}^n$:

$$d(T_1,T_2)^2 = \sum_{i=1}^{n} |T_{1i} - T_{2i}|^2$$
$$= \sum_{i=1}^{n} |(T_{1i} - T_{3i})|^2 - 2\sum_{i=1}^{n} (T_{1i} - T_{3i})(T_{2i} - T_{3i}) + \sum_{i=1}^{n} |T_{2i} - T_{3i}|^2$$
$$\leq d(T_1,T_3)^2 + 2d(T_1,T_3)\,d(T_2,T_3) + d(T_2,T_3)$$
$$= (d(T_1,T_3) + d(T_2,T_3))^2.$$

Thus, $\quad d(T_1,T_2) \leq d(T_1,T_3) + d(T_2,T_3)$

it is the same as $d(T_1,T_2) \leq d(T_1,T_2) + d(T_2,T_2)$

as the Euclidean distance is symmetric.

References:
Wikipedia Contributors (2022). Euclidean distance. [online] Wikipedia. Available at: https://en.wikipedia.org/wiki/Euclidean_distance [Accessed 9 Nov. 2022].