Part I: Time Series

Excercise 1

Excercise 1.a

The following results were obtained after running my script (which took about 15 minutes):

Pair of classes	Manhattan	DTW, w = 0	DTW, w = 10	DTW, w = 25	DTW, w = inf
abnormal:abnormal	67.77	67.77	38.65	26.48	25.37
abnormal:normal	67.52	67.52	34.2	26.94	26.35
normal:normal	45.65	45.65	24.42	22.17	21.87

Excercise 1.b

Compare and discuss the results of the DTW and Manhattan distances on separating abnormal and normal heartbeats (i). Which one would you choose in practice and why (ii)?

(i)

- Manhatten distance and DTW with $w = 0$ deliver exactly the same results.
- As the window parameter w increases the DTW distances decrease.
- Most interestingly, Manhatten distance and DTW with $w = 0$ seem **not** to seperate the groups like one would expect: "abnormal:abnormal" are further apart on average (67.77) then "abnormal:normal" (67.52). For DTW with $w = 10$ this phenomenon is even more pronounced.
- DTW with $w = 25$ & $w = \infty$ separate the groups like one would expect: "abnormal:abnormal" and "normal:normal" are closer/more similar on average compared to itself ((26.48, 22.17) for $w = 25$ and (25.37,21.87) for $w = \infty$) then when compared to each other, i.e. "abnormal:normal" (26.94 for $w = 25$ and 26.35 for $w = \infty$). Given the results, DTW with $w = \infty$ seems to separate better than DTW with $w = 25$.

(ii)

Eventhough the empirical results indicate that DTW with $w = \infty$ separates the groups better, I would choose DTW with $w = 25$ in practice for the following reasons:

1. $w = \infty$ implies that every element in one time series can be aligned with every element in the other time series, i.e. it's like not having any constrained at all. This can cause pathological warping (one element gets aligned with many elements from the other time series) which is not in line with an expected, intuitive "feature to feature" alignment (Zhang et al., 2017, p. 2). Therefore, we need some sort of window constrained, i.e. $w = 25$.
2. Choosing a window constrained will speed up the computations (Ratanamahatana & Keogh, 2005, p. 507). Hence, $w = 25$ should also be preferred over $w = \infty$ in practice, especially with large data sets.

Excercise 1.c

Discuss the effect that hyperparameter w has on the DTW distance and its ability to separate abnormal and normal heartbeats.

As mentioned in the previous excercise, the distances get smaller and the separation ability of DTW gets better with increasing w. See 1.a for the results.

Excercise 1.d

Is the DTW distance a metric? If not, give examples showing which conditions are not satisfied.

No, DTW is not a metric since it doesn't satisfy the triangle-inequality. The following example is given by Wang (2006, p .13):

> Suppose there are three time series data x, y, z where
>
> $x = 1, 1, 1, 2, 2, 2$
>
> $y = 1, 1, 2, 2, 2, 2$
>
> $z = 1, 1, 1, 1, 2, 2$
>
> The local DTW distances between them with 5%-warping are $D(x, y) = 0$, $D(x, z) = 0$ and $D(y, z) = \sqrt{2}$. Thus $D(y, z) > D(x, y) + D(x, z)$ where it does not obey the triangle inequality.

Excercise 1.e

What is the runtime complexity of computing the DTW distance with w-constrained warping? You may assume that m = n. By contrast, what is the runtime complexity of DTW (without any constraints)?

Runtime complexity without constraints: $\mathcal{O}(m^2)$, given that $m = n$.

Runtime complexity with w-constrained warping: $\mathcal{O}(mw)$, where w is the window-paramter. The reason for this is that the algorithm's inner loop goes now only over a part of the second time series, determined by the window size w.

Part II: Graphs

Excercise 2

Excercise 2.a

See the code included in the zip file.

Excercise 2.b

See the code included in the zip file. The generated output looked as follows:

Pair of classes	SP
mutagenic:mutagenic	5303.65
mutagenic:non-mutagenic	2706.78
non-mutagenic:non-mutagenic	1428.25

Excercise 2.c

What is the runtime complexity of Floyd-Warshall's algorithm? What is the runtime complexity of SP? How would you improve the runtime?

- Floyd Warshall's runtime complexity: $\mathcal{O}(n^3)$
- SP runtime complexity: $\mathcal{O}(n^4)$.
- One could try to improve the runtime by parallelization of the shortest path kernel like Xu, Wang, Alvarez, Cavazos, and Zhang (2014) propose in their paper.

References

Ratanamahatana, C. A., & Keogh, E. (2005). Three myths about dynamic time warping data mining. In *Proceedings of the 2005 siam international conference on data mining* (pp. 506–510). Retrieved from https://epubs.siam.org/doi/pdf/10.1137/1.9781611972757.50

Wang, Z. (2006). *Time series matching: a multi-filter approach* (Doctoral dissertation, New York University, Graduate School of Arts and Science). Retrieved from https://cs.nyu.edu/media/publications/wang_zhihua.pdf

Xu, L., Wang, W., Alvarez, M., Cavazos, J., & Zhang, D. (2014). Parallelization of shortest path graph kernels on multi-core cpus and gpus. *Proceedings of the Programmability Issues for Heterogeneous Multicores (MultiProg), Vienna, Austria*.

Zhang, Z., Tavenard, R., Bailly, A., Tang, X., Tang, P., & Corpetti, T. (2017). Dynamic time warping under limited warping path length. *Information Sciences*, *393*, p. 91–107. Retrieved from https://hal.archives-ouvertes.fr/hal-01470554/document