Part 2 Exercise 2

Group 10: Till Kisling, Samara Jansen, Valton Lohaj

We used three dataset:

Dataset	Number of classes	Number of features	Number of samples
Olivetti faces dataset	40	4069	400
The Insurance Company Benchmark	2	85	4000
Reuters Corpus Volume I	103	47236	804414 (used 1000)

We used three DR methods: Isomap, Locally linear embedding and Laplacian eigenmap.

We trained all methods on all datasets and evaluated them with trustworthiness, continuity and LCMC.

Olivetti faces:

Here the best performing method is the Laplacian eigenmap, it has the best values for every measure. But even though it's the best performing method you can't really see the different clusters in the plot and the LCMC is also very low, so we have to say that no method performs extraordinarily well on this dataset.

The Insurance Company Benchmark:

Based on the measures there is no real best method for this dataset. Looking at the plots there are visible clusters, but looking at the measures especially the LCMC, we have to assume that these clusters aren't really good.

Reuters Corpus Volume I:

On this dataset all Methods perform the worst. This can of course be explained by the high number of classes with a particularly small dataset of only 1000 samples. In the LLE and the Laplacian eigenmap there are some visible clusters, but for the Isomap there isn't anything that could be considered a cluster.

Overall, we can say that all three methods performed pretty poorly.

---

https://github.com/xValo/ml_in_web/blob/master/part2_2.ipynb