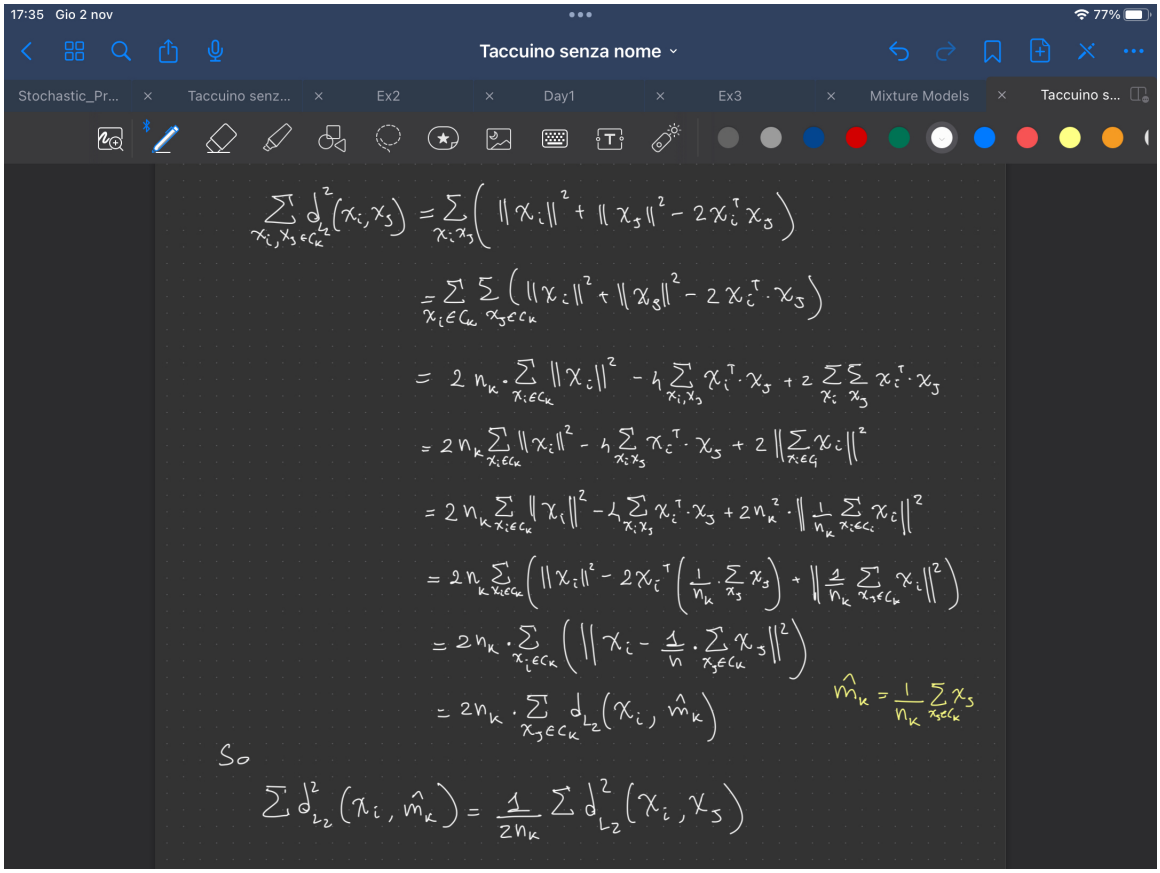
Exercise 3 Trashaj Alberto 1075402

alberto.trashaj

October 2023

1 Point 3



2 Point 4

2.1 Point 4.a

In the experiments conducted, the best distance metric for Complete Linkage clustering is L1 distance in most cases. The only exception is the "simple normal 0.99 noise" case, where L1 distance is still the best, but not significantly better than L2 distance.

The assessment of the best distance metric is based on the Adjusted Rand Index (ARI) which measures the similarity between the true clustering and the clustering obtained from the algorithm. In all cases, L1 distance yielded the highest ARI score, indicating that it provided the most accurate clustering.

Regarding the plots, they show the impact of different data standardization/transformations on the clustering results. On the x-axis, you have the type of transformation applied, and on the y-axis, you have the ARI index. This means that not only is L1 distance the best among all distance metrics, but it also outperforms all other possible data standardizations.
In summary, based on the experiments and ARI scores, L1 distance appears to be the most suitable metric for Complete Linkage clustering in these cases. This conclusion is supported by the fact that, across different data transformations, L1 consistently yields the highest ARI scores, indicating the highest quality of clustering results.

2.2 Point 4.b

In both algorithms, L1 distance consistently leads to a higher ARI for any type of transformation. However, in many cases, the difference in ARI scores is much more pronounced in Complete Linkage clustering. Essentially, L1 distance outperforms all others in Complete Linkage, whereas in PAM, while it does result in a higher ARI, other distances also achieve similar ARI levels.