

Intelligent Systems - Homework *until 06-05-2015*

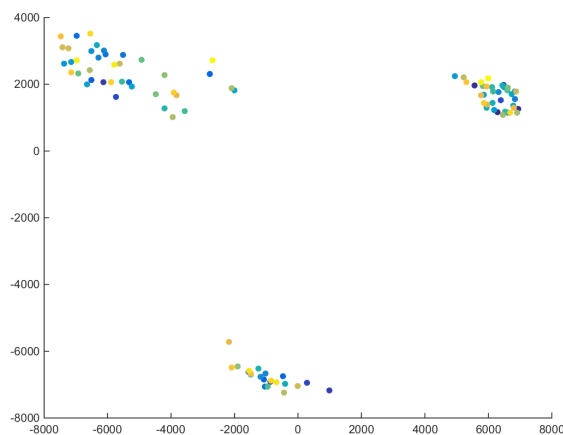
Task 1

features_A.txt

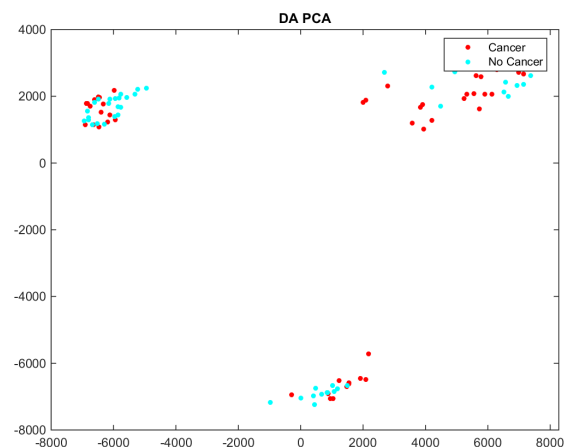
The assumption we made for the features_A.txt file were pretty bad. As we can see, features_A.txt is all about “truthy” or “falsy” values. Furthermore we made an automatically coloured plot which went pretty wrong as one can see. We could’ve assumed that there are three clusters with one feature per cluster.

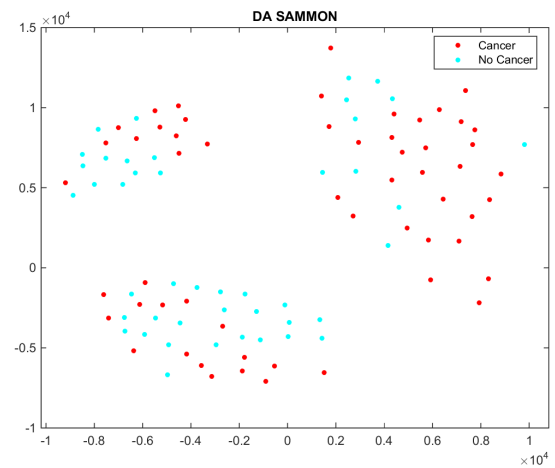
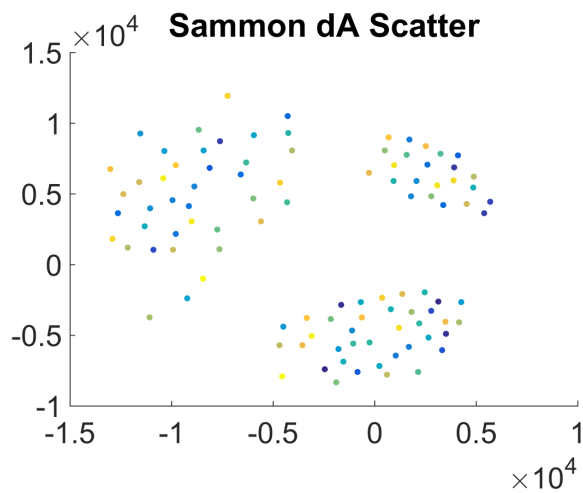
In the new versions there are still the same clusters identified. However, the clusters do not correspond to the classes, as both classes (Cancer and No Cancer) are present in all clusters.

PCA Old Version

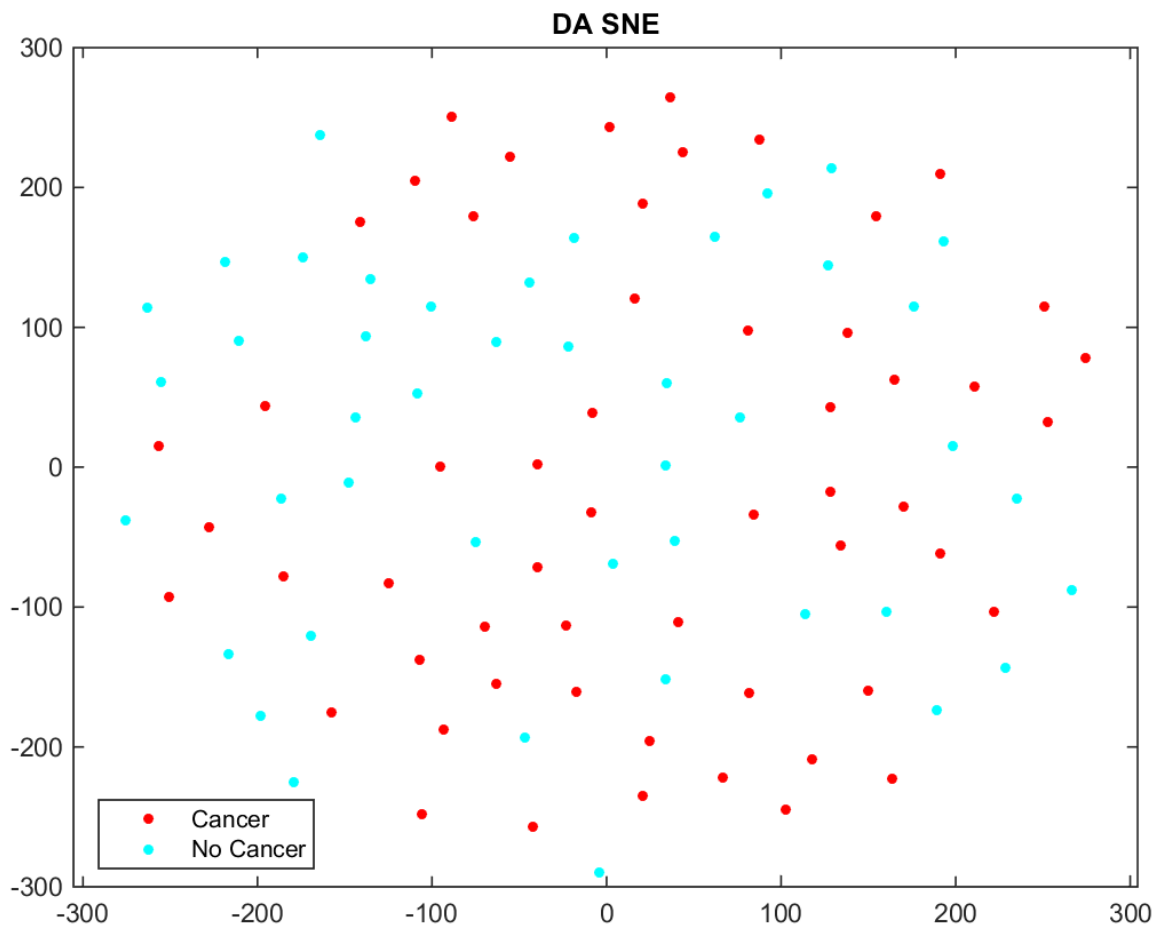


PCA New Version

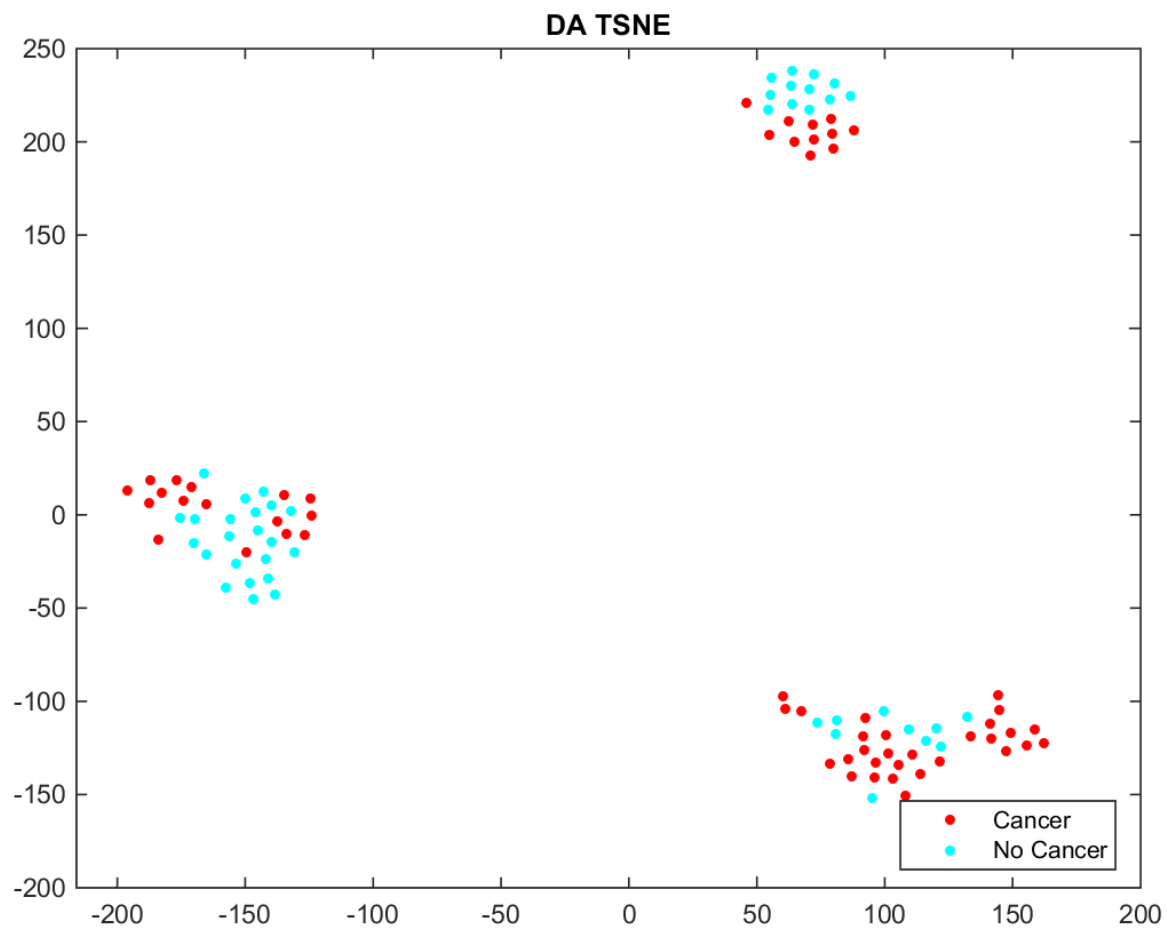




The example for SNE was tested with different parameters. We used a perplexity from 5 to 50 but sadly it had little to no effect on the mapping of our dataset.



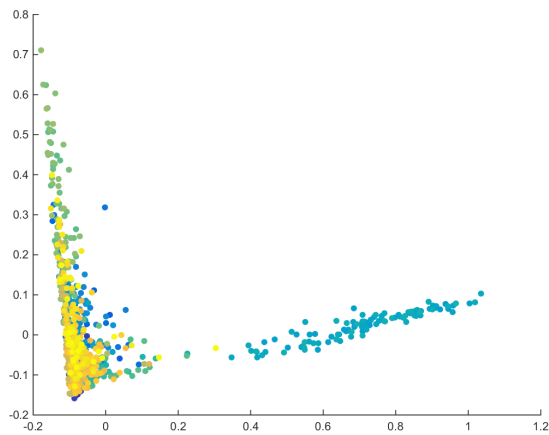
For T-SNE we tested the same range of perplexity values as for SNE. The following figure shows the best result for `features_A.txt`. Here, we can at least see the three different clusters compared to SNE.



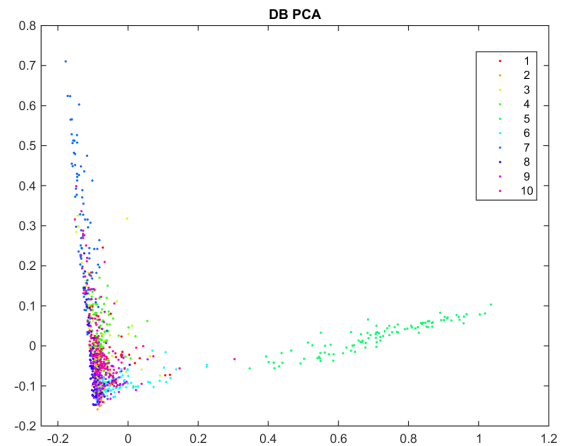
features_B.txt

The assumption we made for the features_B.txt file were pretty accurate. As we can see, features_B.txt has a lot of different features which we could identify pretty well with our automatic colour mapping.

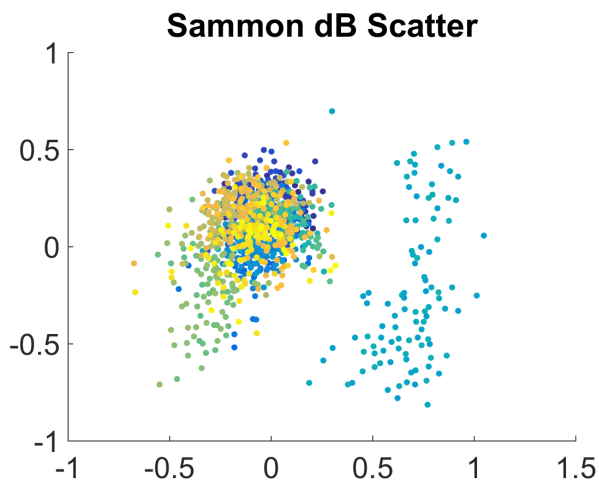
PCA Old Version



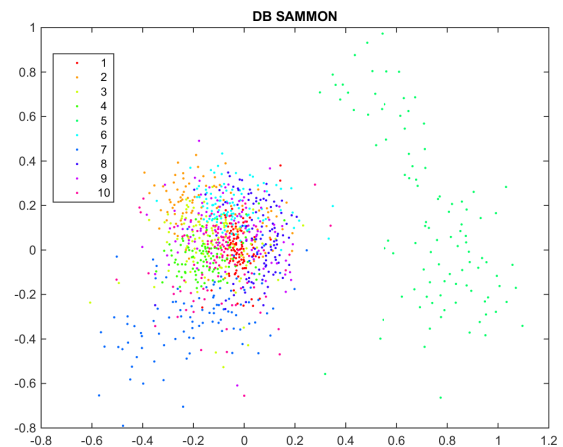
PCA New Version



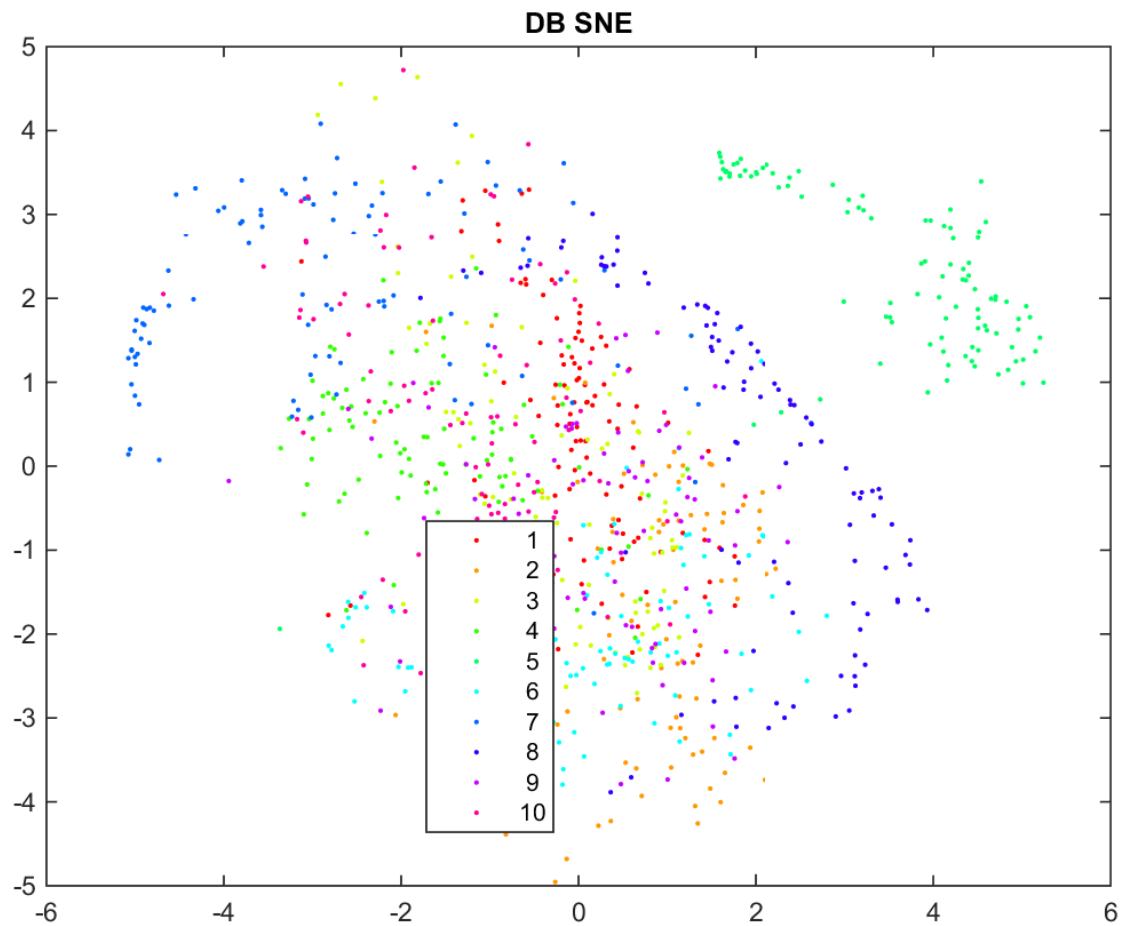
Sammon Old Version



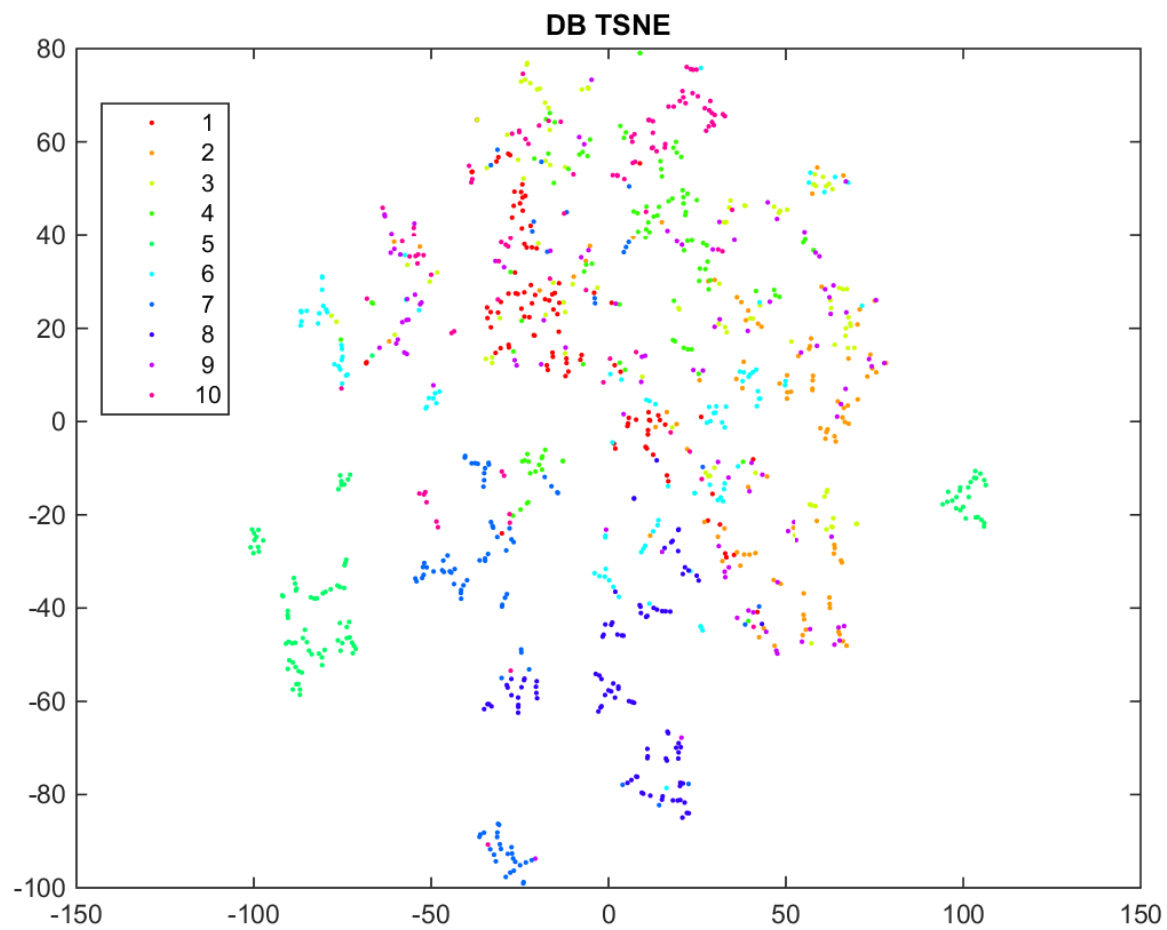
Sammon New Version



The example for SNE was tested with different parameters. We used a perplexity from 5 to 50 but sadly it had little effect on the mapping of our dataset. We can vaguely identify 3 clusters (blue, green and the rest).



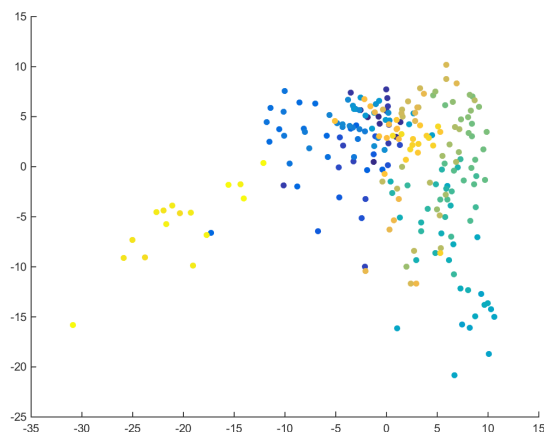
For T-SNE we tested the same range of perplexity values as for SNE. The following figure shows the best result for `features_B.txt`. A high number of clusters can be identified, which always consist of a small amount of points. Points and cluster correspond to each other. The small clusters can be spliced to create larger clusters.



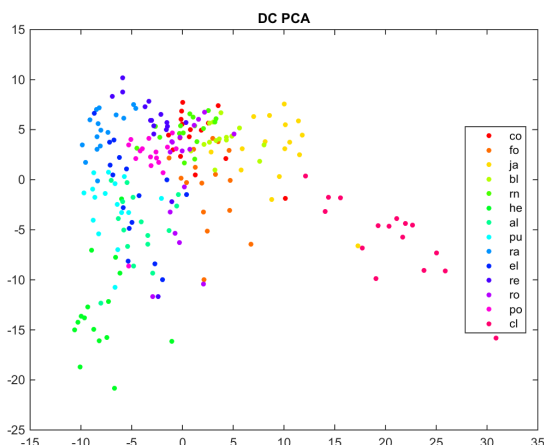
features_C.txt

The assumption we made for the features_C.txt file were pretty accurate again. As we can see, features_C.txt has a lot of different features which we could identify pretty well with our automatic colour mapping.

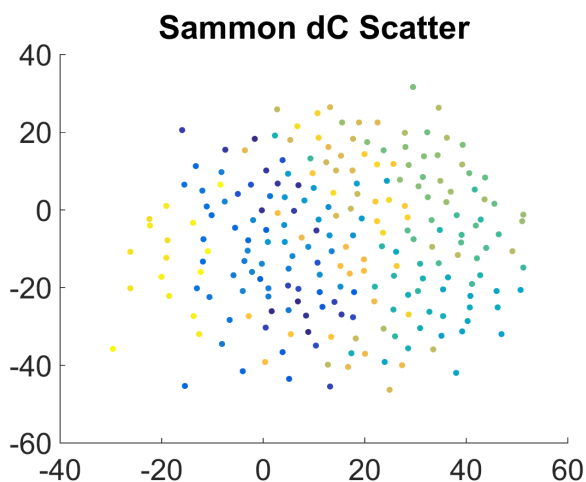
PCA Old Version



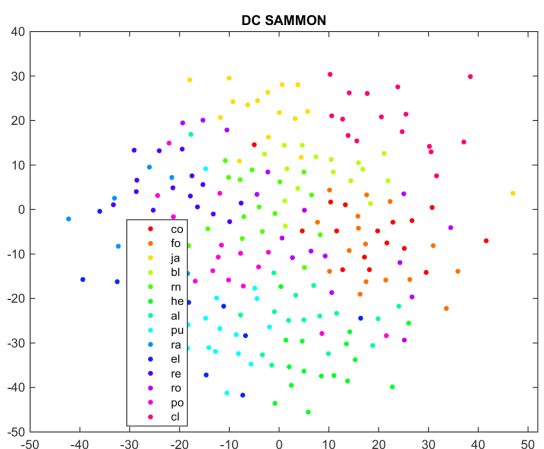
PCA New Version



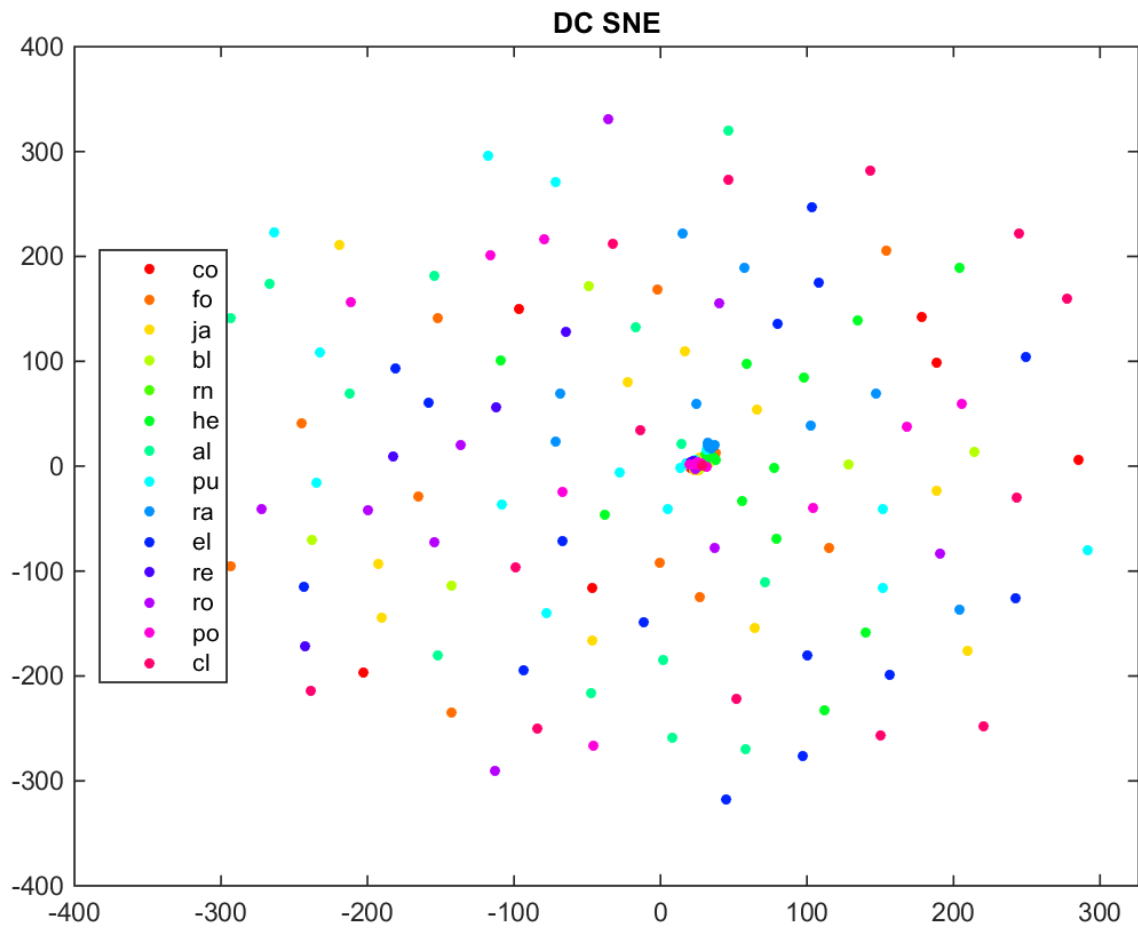
Sammon Old Version



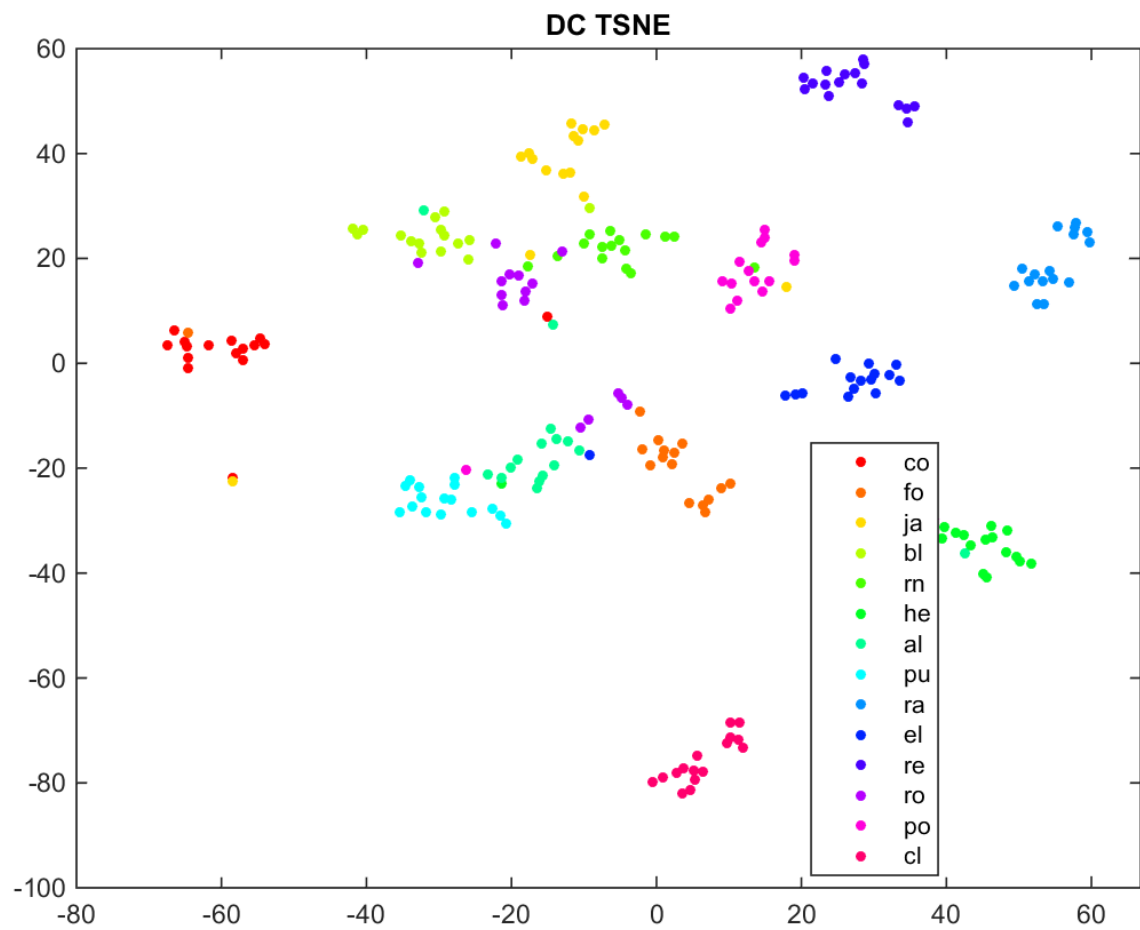
Sammon New Version



The example for SNE was tested with different parameters. We used a perplexity from 5 to 50 but sadly this time it had literally no effect on the mapping of our dataset.



For T-SNE we tested the same range of perplexity values as for SNE. The following figure shows the best result for `features_C.txt`. A high number of clusters can be identified, which always consist of a small amount of points. Points and cluster correspond to each other.



Best mapping for each feature-set

features_A.txt

The best result for features_A.txt was achieved by using t-SNE with a perplexity of 18. The result were three clusters. One that has a high number of truthy values (positive cancer results), and the others pretty mixed.

features_B.txt

The best result for features_B.txt was also achieved by using t-SNE with a perplexity of 50. Clusters are easily distinguishable and separated in classes.

features_C.txt

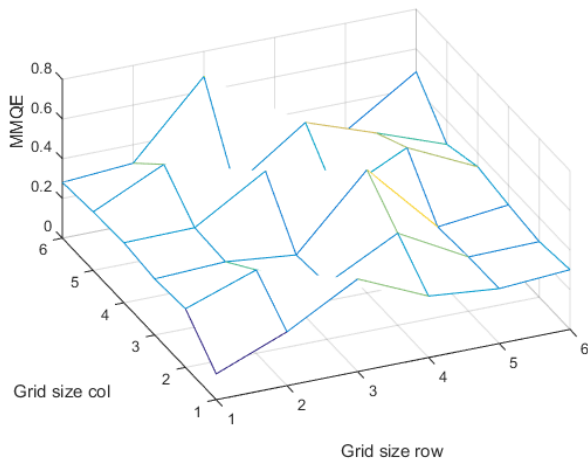
Again the best result for features_C.txt was achieved by using t-SNE with a perplexity of 5. Clusters are easily distinguishable and separated in classes.

Task 2

Our evaluations show that with increasing grid size the MMQE also linearly increases. Even though our own implementation of SOM has a lot of noise, the assumption is proved by the SOM-Toolbox implementation. In our Implementation one could see at least minor increases on the error value at increased grid size.

We tested with grid sizes from 1x1 to 5x5.

MMQE - Our Implementation



MMQE - Toolbox Implementation

