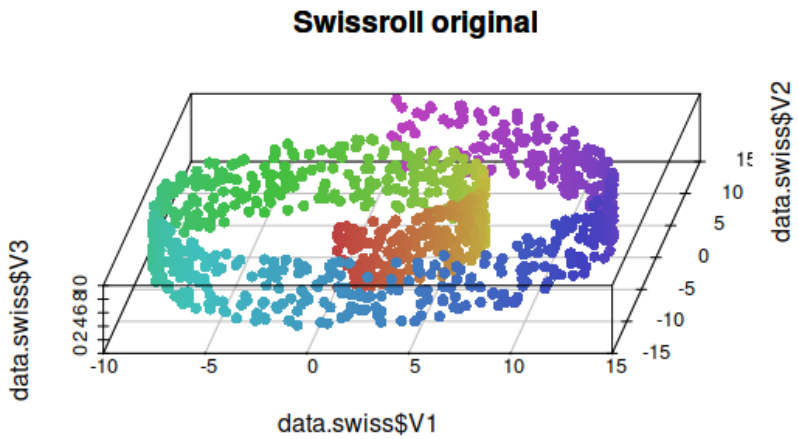
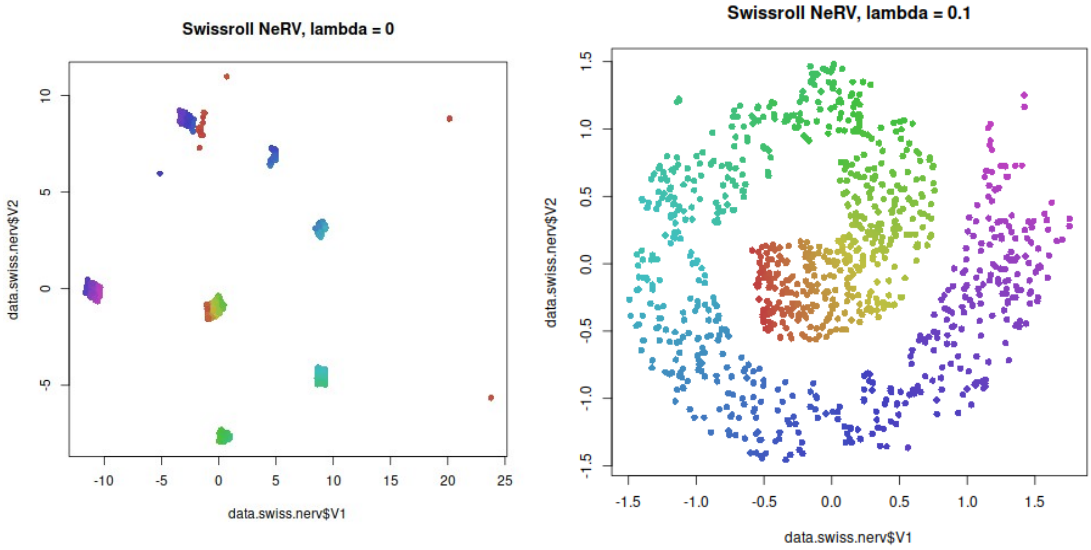
DATA.STAT.770

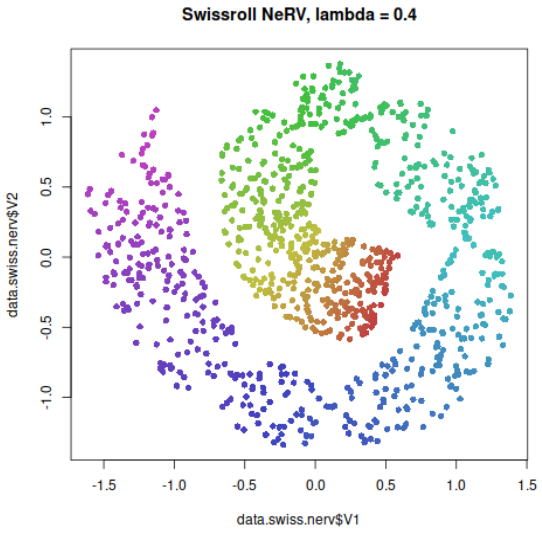
Exercise set 12

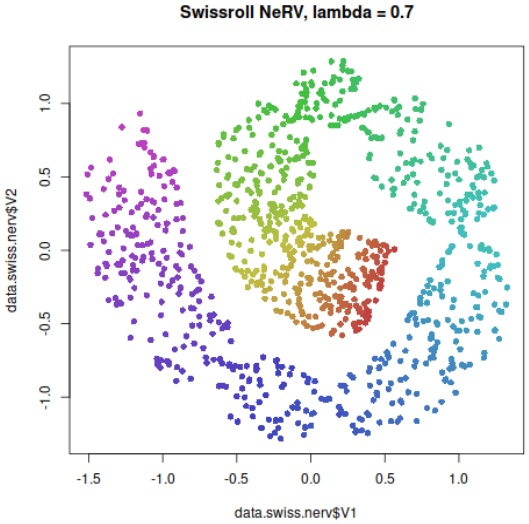
G2.

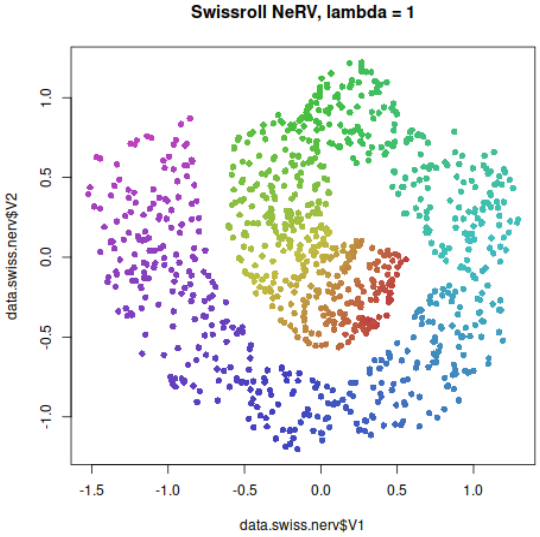
Swissroll dataset visualization:



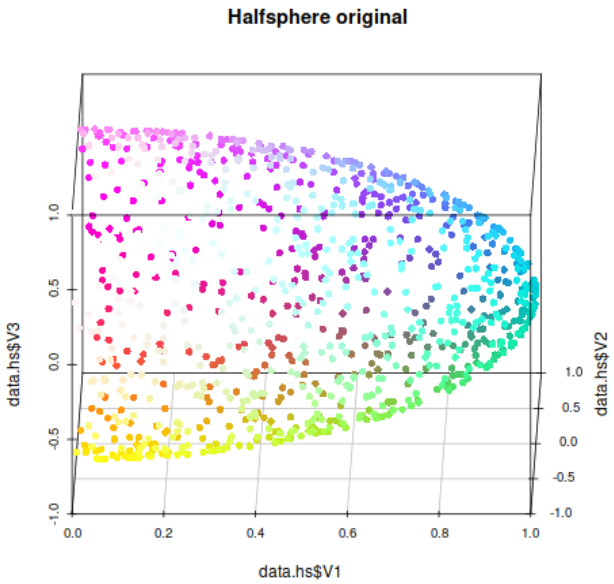


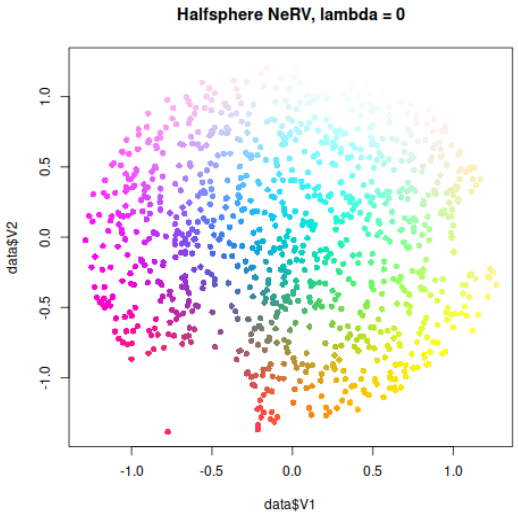


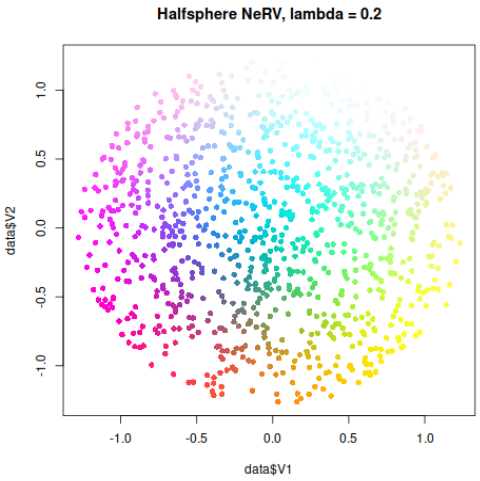


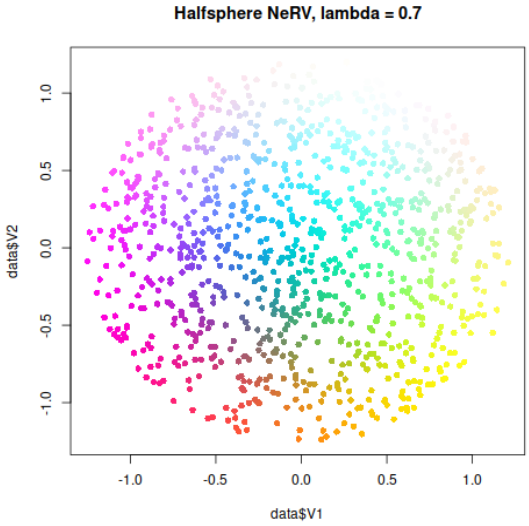


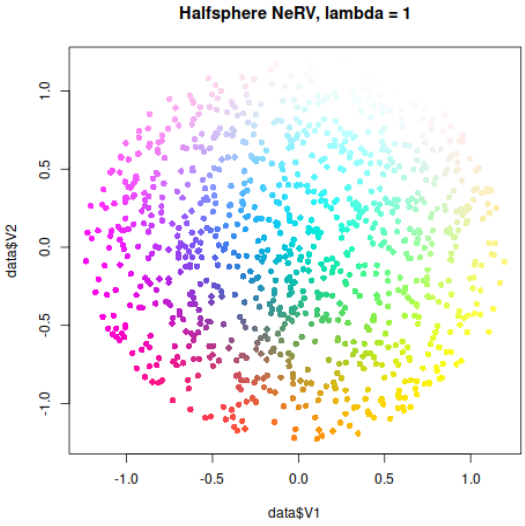
Halfsphere dataset:



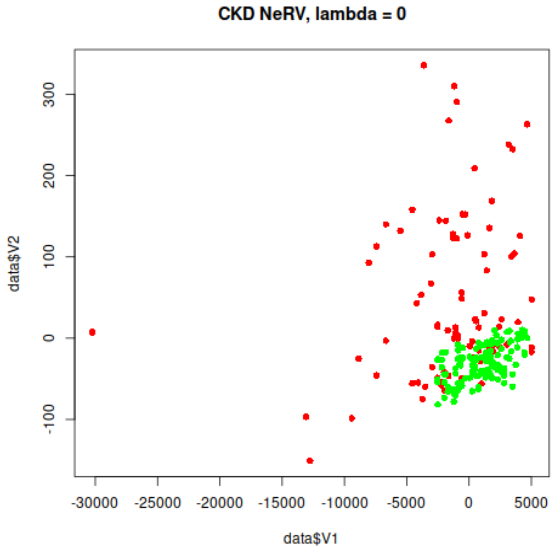

Halfsphere NeRV, lambda = 0
data$V2
data$V1
-1.0
-0.5
0.0
0.5
1.0

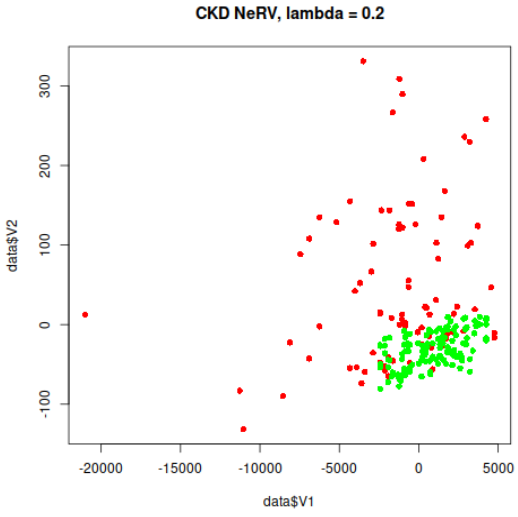

Halfsphere NeRV, lambda = 0.2
data$V2
data$V1
-1.0
-0.5
0.0
0.5
1.0

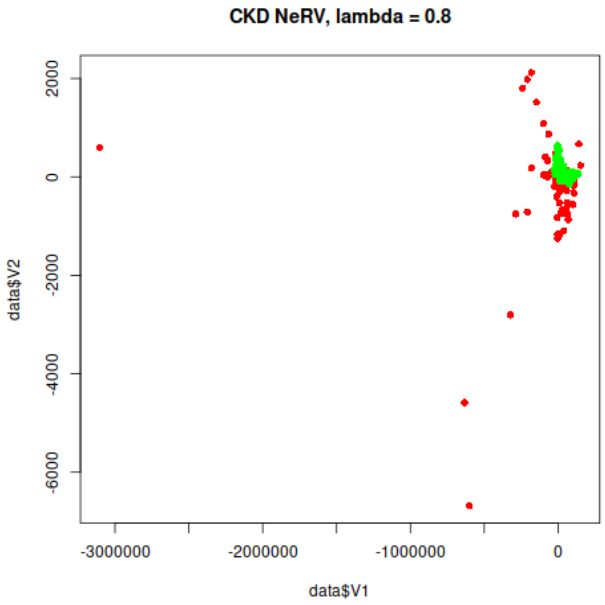

Halfsphere NeRV, lambda = 0.7
data$V2
data$V1
-1.0
-0.5
0.0
0.5
1.0

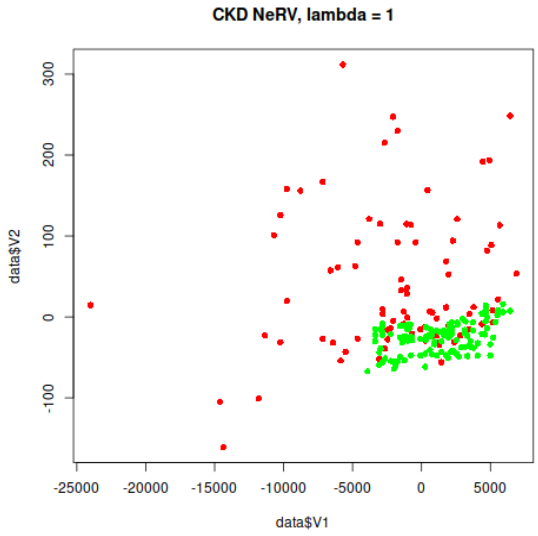

Halfsphere NeRV, lambda = 1
data$V2
data$V1
-1.0
-0.5
0.0
0.5
1.0

CKD dataset:


CKD NeRV, lambda = 0
data$V2
data$V1
-100
0
100
200
300
-30000
-25000
-20000
-15000
-10000
-5000
0
5000

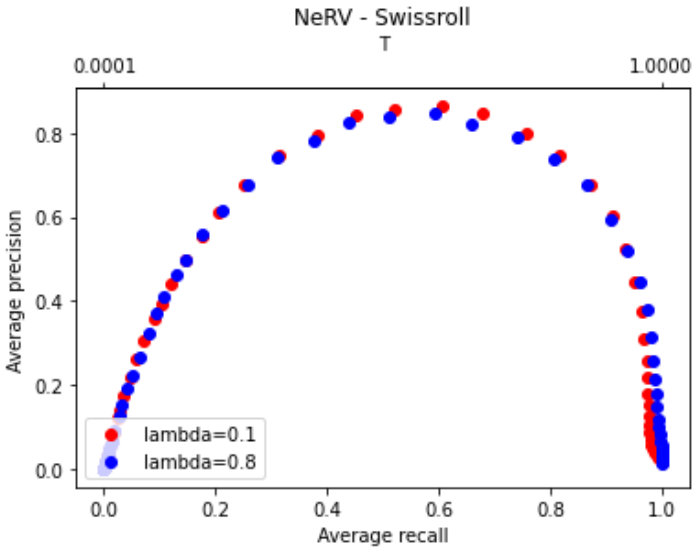

CKD NeRV, lambda = 0.2
data$V2
data$V1
-100
0
100
200
300
-20000
-15000
-10000
-5000
0
5000

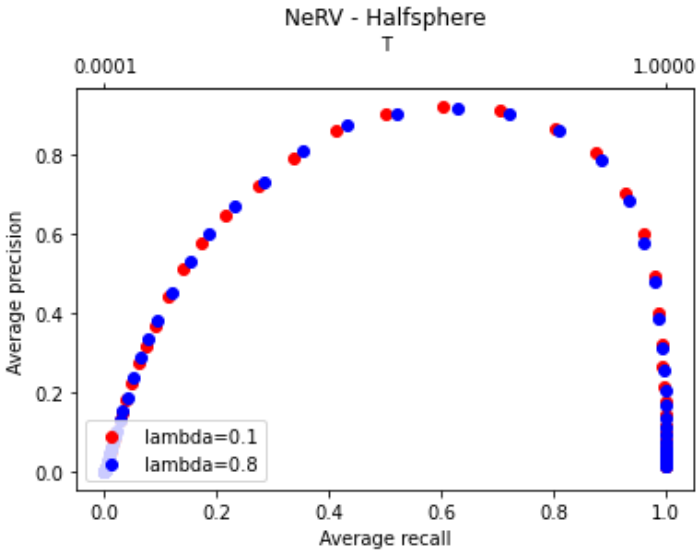


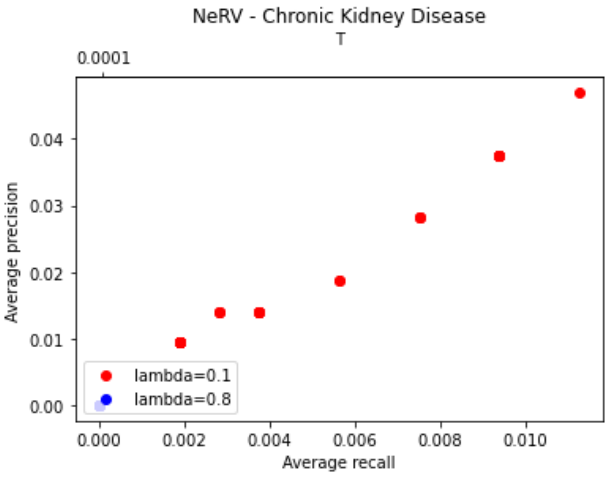


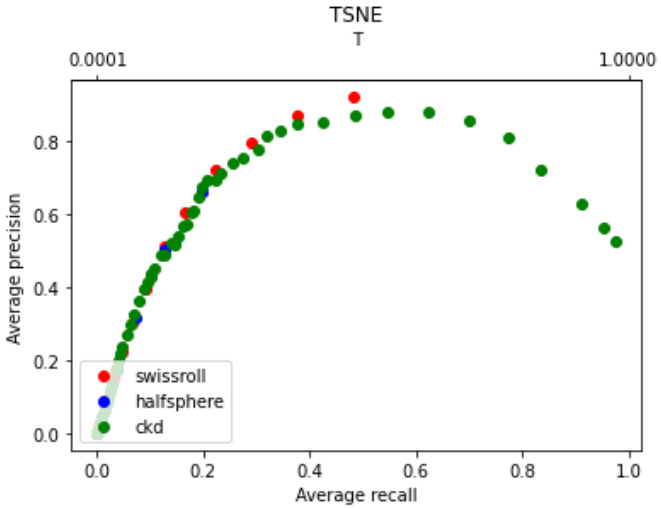
In most cases the small lambda parameter adjustment doesn't affect dramatically to the output. The most interesting patterns maybe happened in swissroll dataset. In the case of lambda=0 (full precision) the visualization merely produces individual clusters instead of unfolding.

Precision and recall curves. In cases where 0 neighbours were retrieved, I set precision to 0. T-values (neighbourhood sizes) are from [0.001, 1] and they are decreased iteratively by multiplying by 0.9.

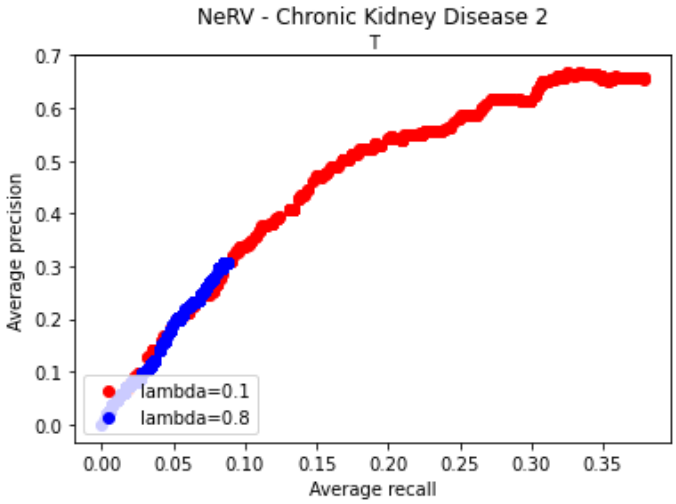






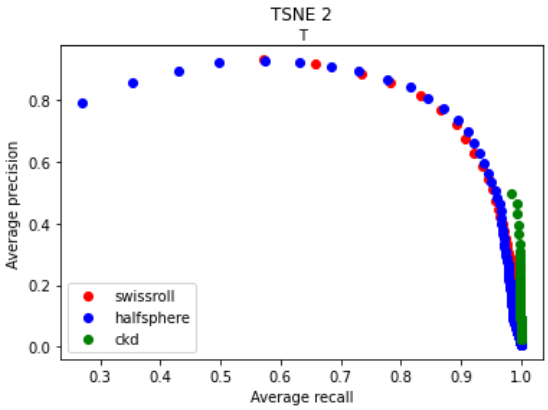


It seems that NeRV method performed pretty well in terms of precision and recall at swissroll and halfsphere datasets. However, originally 14-dimensional CKD dataset causes some difficulties. Overall it seems that projection produced by NeRV is rather sparse (as can be seen from point clouds). Especially when lambda=0.8, both precision and recall stayed at 0. I changed the neighbourhood size from [0.001,1] to [1.1, 50], with step size of 1, and obtained better results as seen below. I had some issues with top T-axis, but the left side means T=1.1 and right side T=50.



The performance is still poorer compared to swissroll and halfsphere, but better than previously.

Lastly, I used TSNE instead of SNE. Compared to the NeRV, this method performed better at CKD data but worse at swissroll and halfsphere (lot of false negatives compared to true positives). I did the same trick mentioned previously (changed T-scale) and obtained following results



The results were improved for swissroll and halfsphere, but CKD preformance in precision decreased remarkably. Based on these results we can deduce that TSNE and NeRV projections operate very differently: NeRV projections for swissroll and halfsphere are quite dense but TSNE projections for these datasets are sparse. Alternatively NeRV projection for CKD is sparse but corresponding TSNE projection is quite dense.