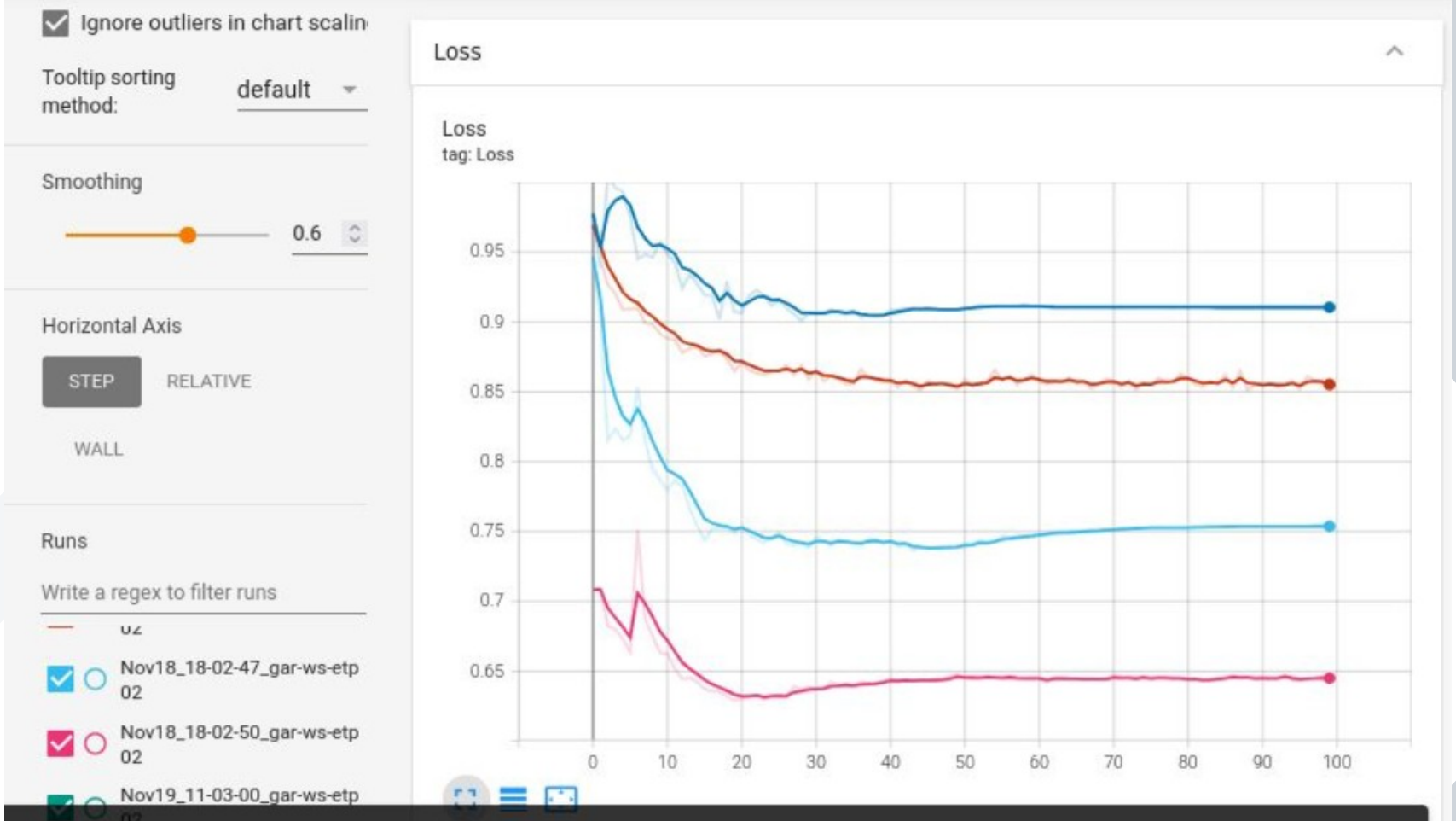
Bao Tai Le

Masterthesis update

WiSe 2024/25
19.11.2024

So where am I today

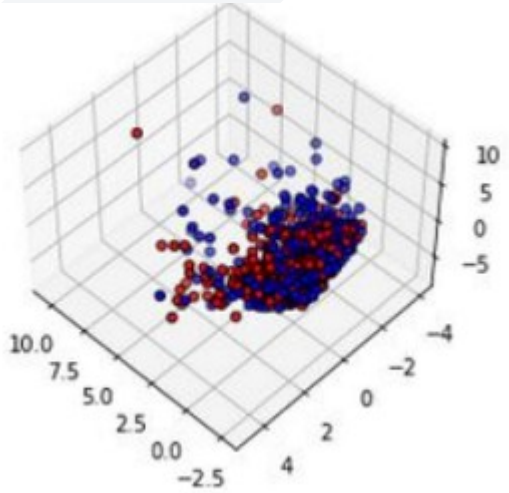


	Smoothed	Value	Step	Time	Relative
s2000.0_with_numhard_1000.0/TriplettVO_Simple_Embedding_V1/100_epochs/Loss_test_loss	0.9105	0.9105	99	Mon Nov 18, 16:50:44	25m 50s
s2000.0_with_numhard_1000.0/TriplettVO_Simple_Embedding_V1/100_epochs/Loss_train_loss	0.8552	0.8522	99	Mon Nov 18, 16:50:44	25m 50s
s20000.0_with_numhard_10000.0/TriplettVO_Simple_Embedding_V0/100_epochs/Loss_test_loss	0.7538	0.7538	99	Mon Nov 18, 18:54:27	3h 8m 47s
s20000.0_with_numhard_10000.0/TriplettVO_Simple_Embedding_V0/100_epochs/Loss_train_loss	0.645	0.6447	99	Mon Nov 18, 18:54:27	3h 8m 47s

```
array([      -0.37087,       -0.54345,        0.66894,       -0.00453,
              0.74774,       -0.41306,       -0.50472,        0.78362,
              0.49711,       -0.32627,       -1.13788,       -1.42681,
             -0.87503,       -1.22414])
```

PCA

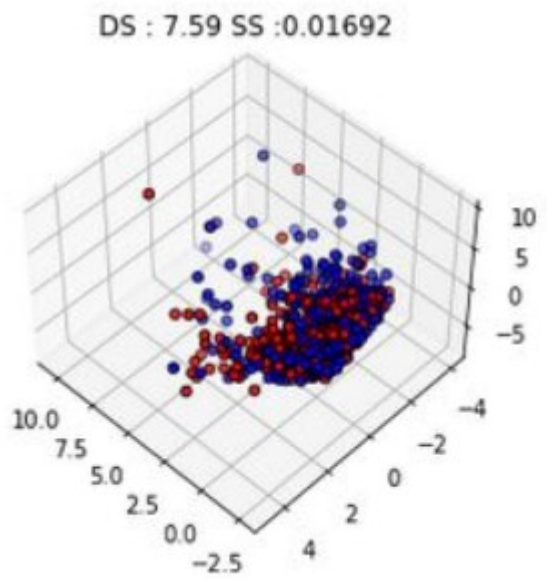
```
array([      -1.62946,        1.42960,        0.17960])
```


10
5
0
−5
−4
−2
0
2
4
10.0
7.5
5.0
2.5
0.0
−2.5

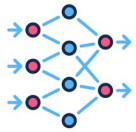
Davies Bouldin Score :
Lower index values indicate a better clustering result. The Index is improved (lowered by increased seperation between clusters and decreased variation within clusters)

Silhoutte Score :

The silhoutte score is a measure of how similar an object is to its own cluster compared to other clusters. The silhouette score ranges from -1 to +1 wherea high value idicates that the object is well matched to its own cluster and poorly matched to a neighbouring cluster. A clustering with an average score 0.7 is considered to be „strong“ ,0.5 „reasonable“ and 0.25 „weak“.


DS : 7.59 SS :0.01692
10
5
0
−5
10.0
7.5
5.0
2.5
0.0
−2.5
−4
−2
0
2
4

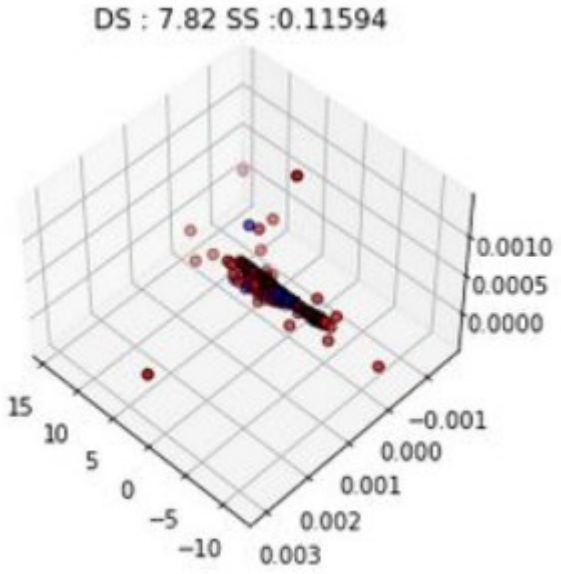
```
array([ -0.37087,  -0.54345,   0.66894,  -0.00453,
         0.74774,  -0.41306,  -0.50472,   0.78362,
         0.49711,  -0.32627,  -1.13788,  -1.42681,
        -0.87503,  -1.22414])
```

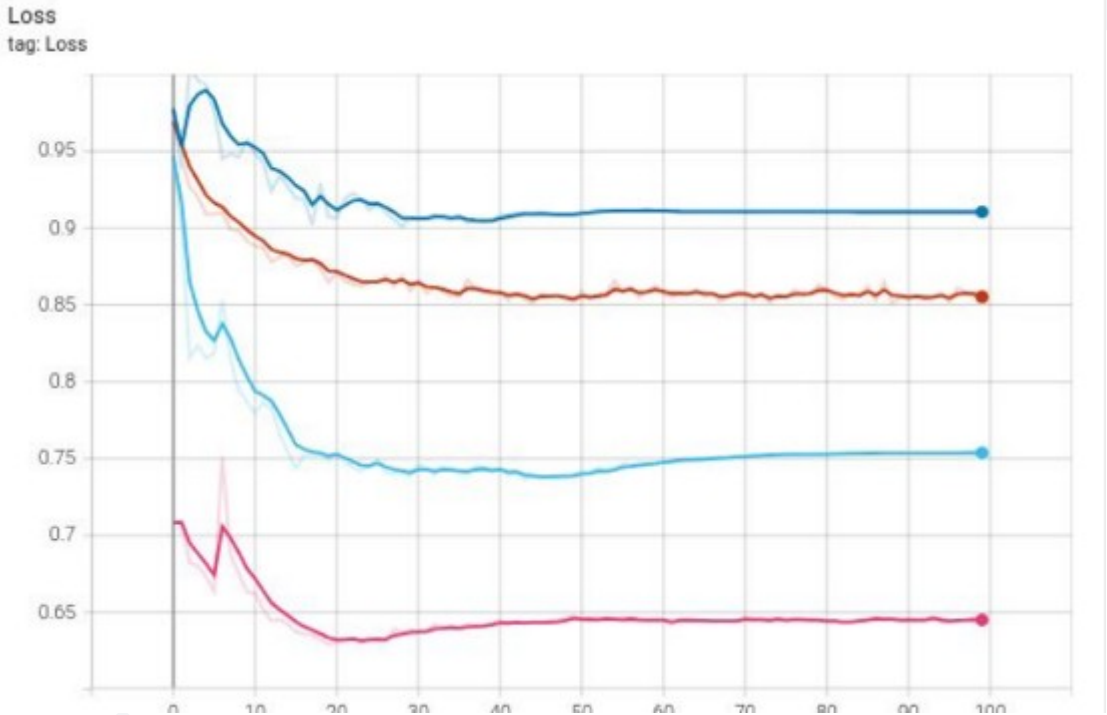


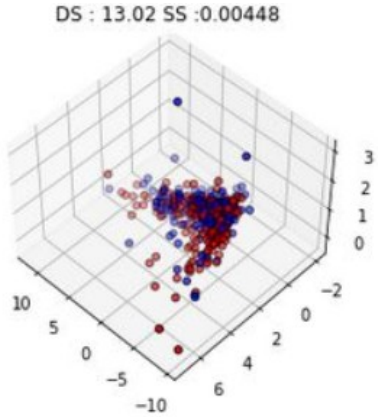
```
array([  0.09131,  -0.07131,  -0.07131,   0.09132,
         0.09132,   0.09132,   0.09131,   0.09132,
        -0.07132,  -0.07132])
```

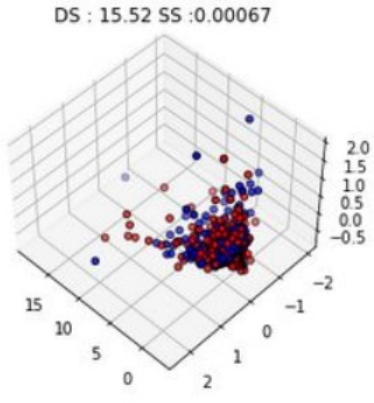

```
[  -0.77328   -0.00001   -0.00000]
```

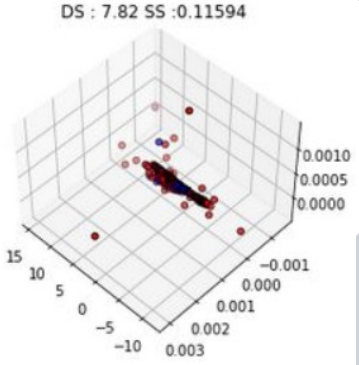
```
[      -0.77328        -0.00001        -0.00000]
```

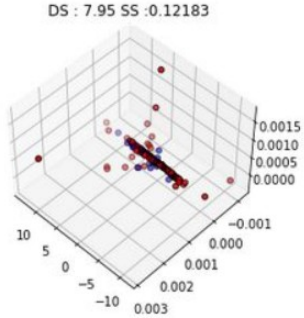



Loss
tag: Loss
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0
10
20
30
40
50
60
70
80
90
100


DS : 13.02 SS :0.00448


DS : 15.52 SS :0.00067


DS : 7.82 SS :0.11594


DS : 7.95 SS :0.12183

Dataset

Variables are found and looks promising

Just need a bigger dataset

Create the batches

The functions are created and work as intended

Bigger batch sizes

May need to do some unittesting to be 1000% sure

Train the Embedding Layer

The functions is created and works as intended with a smaller sample

PCA is set up

Model Loading and saving is setup

Needs optimization
- regularization
- different lossfunctions
- deeper layers
Etc.

Bigger batchsizes

Train the Classification Neural Network

Needs to be set up