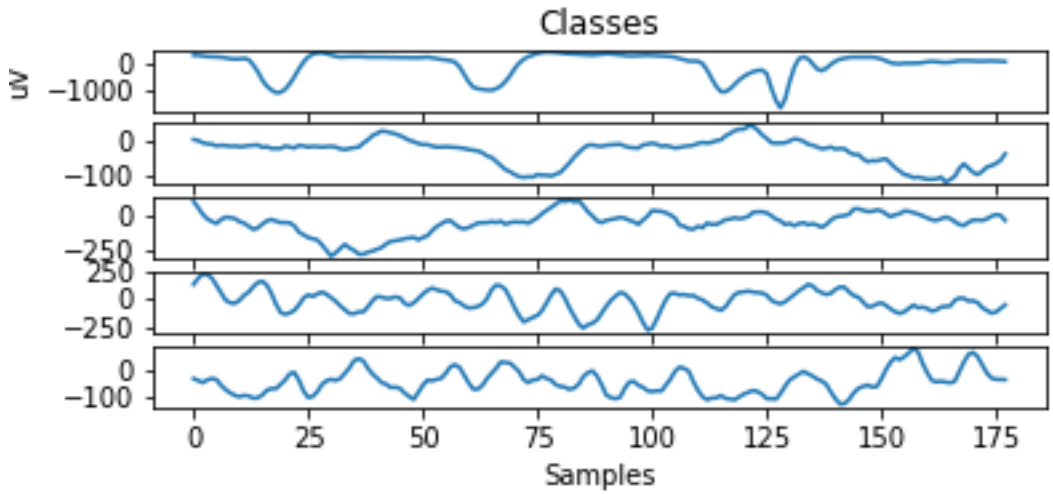
Solim LeGris
November 13, 2020

MAIS 202 - PROJECT DELIVERABLE 3

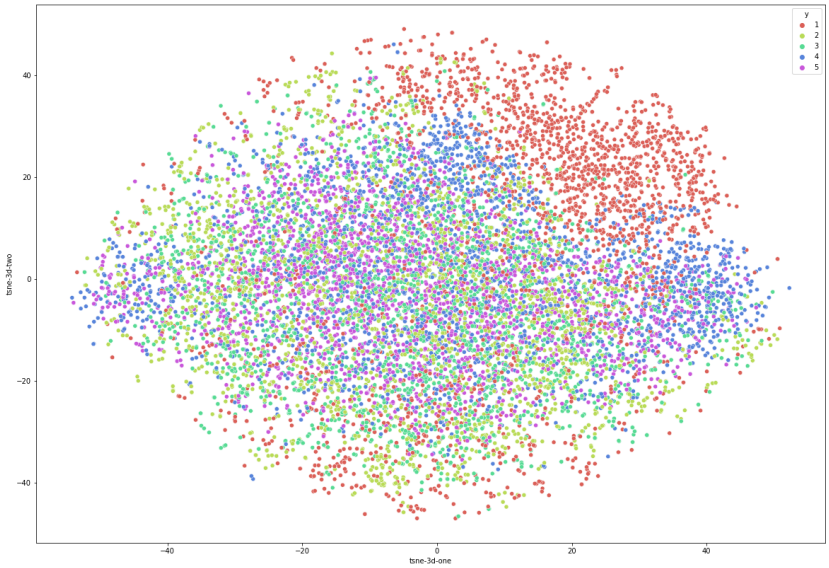
Further Data Analysis

As mentioned in my previous deliverable, I was getting extremely good results on both classifiers that I trained so I attempted to decipher my data furthermore.

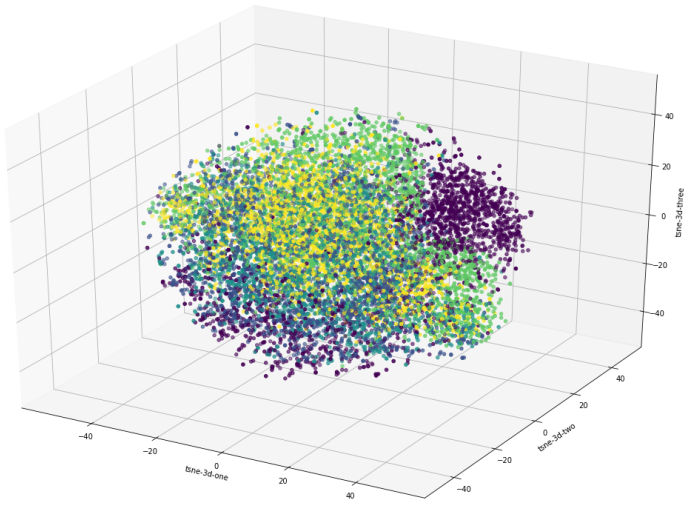
I plotted the 5 classes' EEG time-series to visualize the averaged electrical activity of the patients over the 1s recordings for each condition.



I also applied dimensionality reduction using PCA to my dataset (for visualization) and used t-distributed Stochastic Neighbor Embedding to visualize my data in 2D and 3D. I plotted the graphs for the data with the 5 original labels and with the 2 target labels as well.

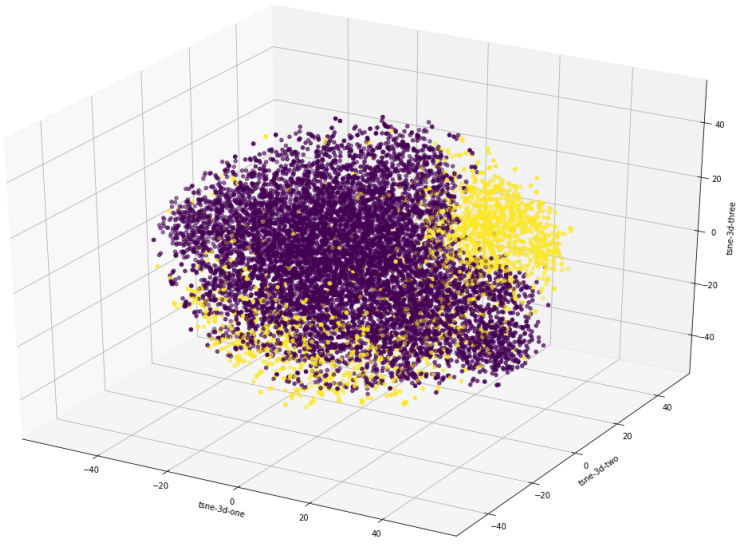


2D Scatter Plot with Multiclass Coloring



3D Scatter Plot with Mutliclass Coloring

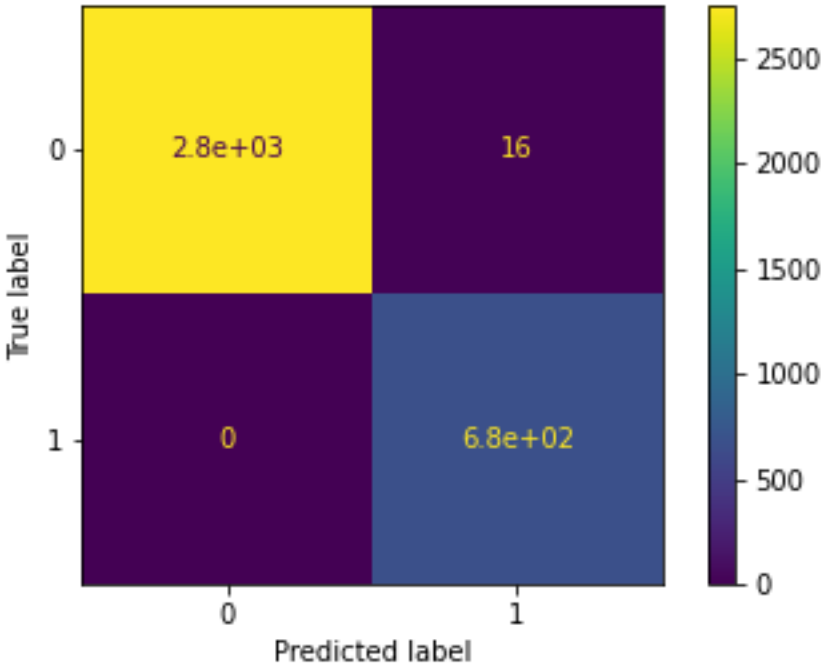
2D Scatter Plot with Binary Coloring



3D Scatter Plot with Binary Coloring

Final Training Results

Given that I was getting very accurate results, I reduced the size of the training set and increased the size of the test set, using 70% of my data as training set. I also standardized my data and introduced a Gaussian noise factor with $\mu = 0$ and $\sigma = 0.5$ to my data and trained a Support Vector Machine using that noisy data. I obtained an accuracy of 0.9953, a log loss of 0.1601 and the following confusion matrix.



Final Demonstration Proposal