Classification of Sleep Stage using Machine Learning Algorithms

1. Visualization of Raw EEG data

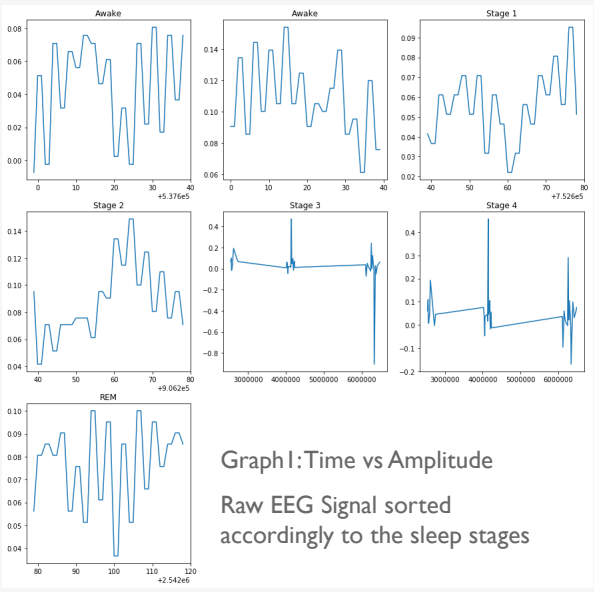
Raw EEG data has 8 channel.

CH1 Samples
CH2 A1
CH3 A2
CH4 C1
CH5 C2
CH6 O1
CH7 O2
CH8 Sleep Stage (1, 0, -1, -2, -3, -4, -5)

CH2 to CH7 are amplitude of EEG signal
Sampling frequency of the signal is 256 Hz.

amplitude of 0 or 1 = awake stage
amplitude of -1 = stage 1
amplitude of -2 = stage 2
amplitude of -3 = stage 3
amplitude of -4 = stage 4
amplitude of -5 = REM

Stage	# of Samples
2.0	2375229
-1.0	2157673
1.0	1976131
-5.0	303523
-3.0	255806
0.0	133029
-4.0	2448



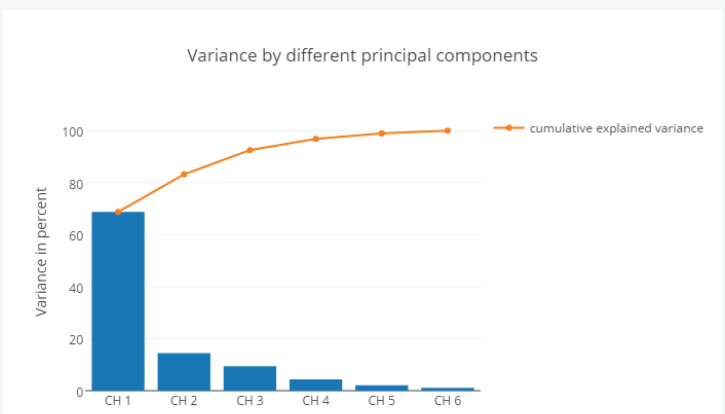
Graph I: Time vs Amplitude

Raw EEG Signal sorted accordingly to the sleep stages

2. Standardisation of the Data

mean = 0 ; variance = 1

Stardardized the whole data set assuming the mean as 0 and variance as 1.

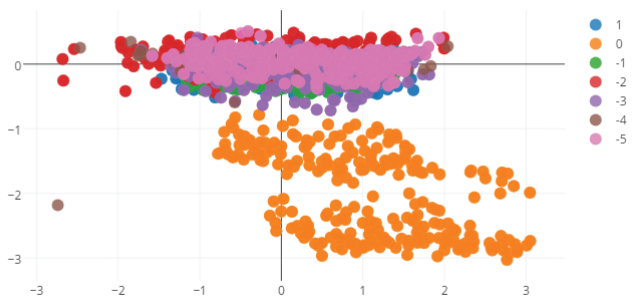


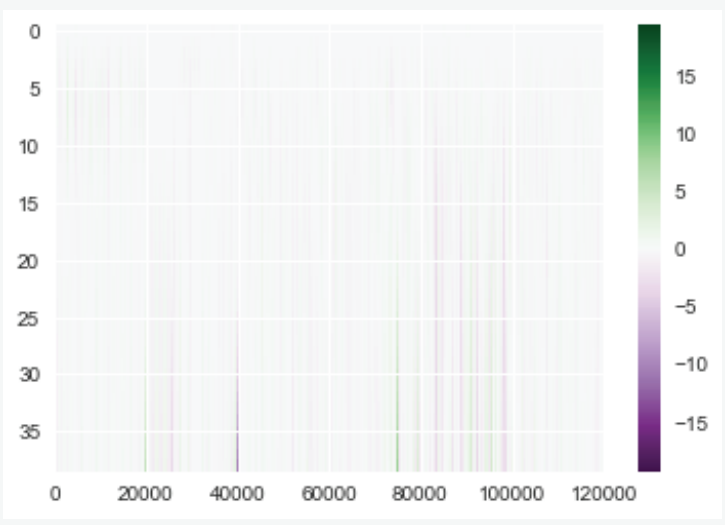
3. Principal Component Analysis

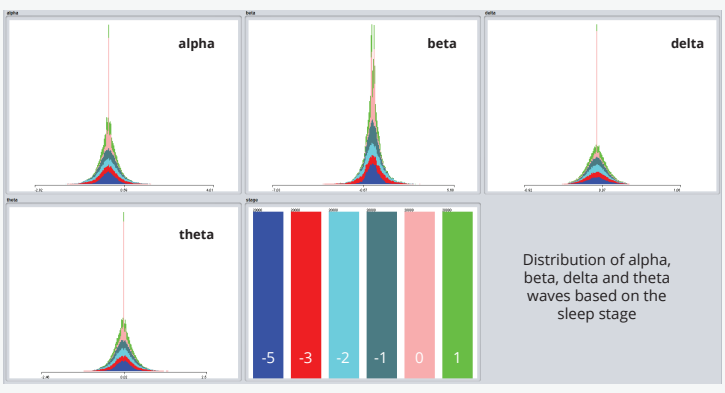
Goal is to reduce the dimensions of a d-dimensional dataset by projecting it onto a (k)-dimensional subspace (where k<d) in order to increase the computational efficiency while retaining most of the information.

Comparing all channel only one channel has high impact. So, I have extracted only one major component out of 6.

Scatter plot of amplitude vs sleep stages







Distribution of alpha, beta, delta and theta waves based on the sleep stage

4. Wavelet Transform

* Selected only 20000 samples from each stage
* Neglected -4 sleep stage since it has contributed to only around 4000 samples
* Did continuous Wavelet Transform to obtain Time vs Frequency matrix for a frequency range of 0-40.

Obtained matrix is 12000X40

Averaged based on the spectrum of different EEG Frequencies

Alpha - 8-12 Hz
Beta - 12-35 Hz
Gamma - 0-4 Hz
Delta - 4-7 Hz

Final Matrix used to Machine Learning is 5X12000

stage	alpha	beta	delta	theta
1.0	-0.061114	-0.288412	0.029730	0.054253
1.0	-0.113436	-0.311923	-0.010910	0.009674
1.0	-0.166980	-0.334301	-0.026620	-0.038508
1.0	-0.220323	-0.355295	-0.035942	-0.088563
1.0	-0.271949	-0.374668	-0.055139	-0.138784

5. Classifiers

Classifier	% of correctness
K - Nearest Neighbour	60 - 82 %
DecisionTreeClassifier	40 - 79 %
Multi-Layer Perceptron	35 - 77 %

The percentage variation is from all levels of stages of sleep to 2-3 levels of stages of sleep. The reason for choosing few level of sleep stage is that data points of stage 2 and 3 lies almost in the same region so I skipped in between stages.

Future improvements:
* Still facing difficulty in implementing STFT programmatically. I will do that in my preprocessing and again start with the prediction
* Implementing Deep Belief Network classifier for 6 million data points, still a difficult task for me with my system configuration though it is a higher end (i7 and 8 GB RAM). But I will try to reproduce as much result possible with my machine.
* Again selection of data points for different sleep stage is a difficult task. Because, transition from Stage 1 to 2 with respect to data point its a immediate sample point but I also need to figure out a stable sleep stage.