**Project/Lab Proposal**
**Tentative Title: Prediction of alcoholism based on EEG data**
**Team Members:Xiaojue Zhou , Smita Bagewadi**
**Brief Overview of Topic and Motivation:**
The project involves analysis of EEG data to predict predisposition to alcoholism.
Challenges:
Extracting features from time series data is challenging.
Machine learning/data analysis tasks involved in the project:
Classification based on least squares.
Principle component analysis to compress the data.
**Core Concepts:**
Concepts and tools may be involved in your project:
Knowledge of signal processing for extracting features
principle component analysis
classification

Since the data is in the form of time-series from different channels , we will first extract relevant features and form feature vectors for each subject.

We can use SVD to compress the data after getting features from time series. We can use SVD both separately for each feature of all regions and on all features of all regions. After getting the principle components, we can look at the singular values and see which regions are relevant. We can consider different features and analyze which of them are relevant based on the error rate of classification.

Also, we can use least square with regularization as above to see if the results are congruent with SVD or not.

On the other hand, we can perform SVD on subjects of alcoholic or control group and find the typical subject from each group.

In general, dealing with time series data is critical in the project. Some of the features include energy, entropy, mean, maximum and minimum values, other features based on fourier transform, wavelet transform and autocorrelation.

**Related Papers, Datasets, or Resources:**

Database of EEG data. It contains data collected from 122 subjects from 64 electrodes placed on different regions of the brain:
https://archive.ics.uci.edu/ml/datasets/EEG+Database