<u>Mid – Way Project Report: CS771A</u>

- **<u>Problem Statement:</u>** There is a dictionary of 60 words and there are 300 persons. Each word has an associated 218 – D feature vector where the features describe some human defined attributes. Each person is being shown a word from the dictionary along with its "Line Drawing" and then FMRI image of the brain is captured which is represented as a 21764 – D feature vector of voxel intensities.

- **<u>Literature Survey:</u>** We have studied different papers regarding the project and following is the summary.

<u>Predicting the brain activity for the given noun: Generation of FMRI</u>

This is a single-subject model which operates under the crucial assumption that **the brain activity for concrete nouns is a linear combination of contributions from each of its semantic features.**

To generate the FMRI for a noun, the model first maps the noun to a set of intermediate semantic features which are obtained from a Google text corpus where each feature is defined as the number of co-occurrences of that verb with the input noun in the text corpus. The final step generates FMRI as a weighted sum of brain images contributed by each of the semantic features. Specifically, the predicted activation *Av* at voxel *v* in the brain image for word w is given by:

$$A_v = \sum_{i=1}^{n} c_{vi} f_i(w)$$

fi(w) is the Co-Occurrence value of the ***i*th** semantic feature with the input word ***w, n*** is the number of features and ***Cvi*** is a learned parameter that specifies the magnitude of activation of the **ith** intermediate semantic feature contributes to voxel v.

This way, there is a significant difference of accuracies between the word-only and word-picture experiment.

- **<u>Approach and Experimental Results:</u>**

<u>Low Variance Filter:</u> We removed all data columns with variance lower than a given threshold, since the columns with little changes in the data, carry little information. By choosing the threshold as 1, we reduced the attributes to 11562.

<u>High Correlation Filter:</u> We calculated the Pearson product-moment correlation coefficient matrix for all the attributes. Pairs of columns with correlation coefficient higher than a threshold are reduced to only one. Using this method, attributes were reduced to around 4000.

<u>Principal Component Analysis (PCA):</u> PCA is equivalent to the best rank-K SVD after the data being centred. PCA in essence, minimizes reconstruction of data from the selected number of basis vectors. Data column ranges need to be normalized before applying PCA. We tried to apply PCA on the dataset but it wasn't effective due to large number of attributes. Hence, we plan to apply functional PCA using Kernel function.

- **<u>Roadblocks:</u>**

  1. We haven’t got a way for effective dimensionality reduction yet, as the best number of attributes that we have reduced the data to, is approximately 4000.

  2. When applied to magnetic resonance images, ordinary PCA runs into serious difficulties because of the extremely high number of dimensions in the data relative to the number of observations.

  3. The number of training data is very less as compared to the number of attributes, this leads to an ineffective implementation of any selected model leading to significantly lower accuracies.

- **<u>Tentative Future Plans:</u>**

  1. We intend to reduce the number of voxels to around 400 using kernelized functional PCA. Regular PCA doesn’t work for high dimensional data without regularization while functional PCA has a built in regularization due to the smoothness of the functional data and truncation to a finite number of included components.

  2. After the number of attributes has been considerably reduced, we can apply Gaussian Naive Bayes which uses the training data to estimate the probability distribution over FMRI observations, conditioned on the subject’s cognitive state. Neural network model can also be applied.

  3. Sparse model, LASSO (Least Absolute Shrinkage and Selection Operator) is a regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and will select similar features via convex optimization.

  4. Ensemble methods such as bagging and boosting can be applied to further increase the accuracy of the selected model.

- **<u>References:</u>**

  1. Blankertz, B., Curio, G., & Mller, K. R., Classifying Single Trial EEG: Towards Brain Computer Interfacing, Advances in Neural Information Processing Systems (NIPS 2001), vol. 14, 157-164, MIT Press, 2002

  2. C. Davatzikos and et al. Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. Neuroimage, 28(1):663–668, 2005

  3. Nathan Eagle. Feature selection analysis. Technical Report TR-557, MIT Media Lab Vision and Modeling, 2002

- **<u>Group Members:</u>**

  1. Aaditya Aanand – 14002
  2. Aryaman – 14139
  3. Pulkit Sharma – 14504