

# **Mid Term Project Report**

## **Members:**

Bhanu Yadav(13198), Hitesh Menghwani(13310), Prashant Kumar(13496)  
Prince Kumar(13511), Sawan Kumar(13637)

## **Project title: Predicting the word from brain activity**

### Work done

- Approach: 1-Nearest Neighbor
  - Applied 1-NN to classify the given test example into one of the two classes of words.
  - Used only those FMRI training images for classification corresponding to the two words (one correct and other incorrect) given for a test example.
  - Achieved 63.33 % accuracy on the test data.
- Approach: 3-Nearest Neighbour
  - Same as previous approach but used  $k = 3$  instead 1 i.e three nearest neighbours for classification of test example.
  - Makes sense to use higher value of  $K$  as the amount of training data is small.
  - Achieved a higher accuracy , 68.33%, compared to previous approach.
- Approach: Distance from means
  - As in first approach used FMRI training images for classification for which label is one of the given two words.
  - Achieved accuracy of 68.33% on the test data.
- Approach: Feature Selection using Mutual Information
  - Using 1-NN classifier we achieved an accuracy of 65% on the test data.
  - Using distance from means classifier we achieved an accuracy of 70% on the test data.

In all the above methods we see that the accuracy that is attained on the test data is significantly higher than random guessing (i.e. 50%).

### Things yet to be done

- Appropriate feature selection and/or feature extraction techniques have to be applied on this high dimensional data to avoid overfitting.
- Feature selection algorithms like Filter Methods ( eg chi-squared ) , Wrapper method , and Embedded method ( which uses regularisation algorithms like LASSO ,Elastic Net and Ridge Regression)
- More advanced and robust classification techniques like support vector machine, logistic regression, etc. have to be applied to achieve better results.
- The feature matrix used to describe the words is yet to be utilized.

### Problems being faced

- Feature extraction techniques like PCA when applied on the training and test data are computationally intensive i.e. it takes very long time as the dimension of training data is very large.
- Problem in finding suitable feature selection technique out of various available techniques.
- Problem in implementing the Mathematical formula of various feature selection methods.

## References

- Cox, David D., and Robert L. Savoy. "Functional magnetic resonance imaging (fMRI)"brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex." *Neuroimage* 19.2 (2003): 261-270.
- Pereira, Francisco, Tom Mitchell, and Matthew Botvinick. "Machine learning classifiers and fMRI: a tutorial overview." *Neuroimage* 45.1 (2009): S199-S209.
- Wang, X., Hutchinson, R., and Mitchell, T., "Training fMRI Classifiers to Detect Cognitive States across Multiple Human Subjects". Submitted to Neural Information Processing Systems 2003
- Mitchell, T., Hutchinson, R., Just, M., Niculescu, R., Pereira, F., & Wang, X., Classifying Instantaneous Cognitive States from fMRI data, The American Medical Informatics Association 2003 Annual Symposium, submitted, 2003