MMT Project Report

Aditya Khandelwal 20171117
Anushka Wakankar 20171112

Link to colab notebook: Code link

Class Participation: Both of us attended each lecture in college, as well as the online classes. We actively participated in most of the classroom discussions and assignments. Hence we feel that both of us deserve a 10 for our class participation.

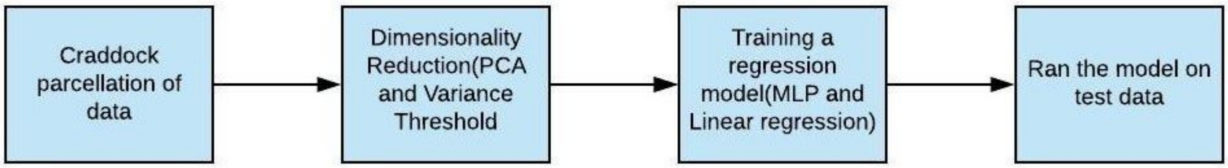
Introduction: Functional magnetic resonance imaging or functional MRI (fMRI) measures brain activity by detecting changes associated with blood flow. Hence, we can use fMRI to gauge the reactions going on inside the brains of participants as they listen to music. But fMRI doesn't actually provide detail at the level of a cell. The 3-dimensional image it provides is built up in units called voxels. Each one represents a tidy cube of brain tissue—a 3-D image building block analogous to the 2-D pixel of computers screens, televisions or digital cameras. Each voxel can represent a million or so brain cells. Using the data we receive from fMRI's and the individual voxels of each region of the brain, we can make a number of observations, like the ones we have made in this project.

Problem statement: Predicting Musical Aptitude scores using brain responses to music.

Dataset: fMRI scans of 38 participants while they listened to the following piece of music

Piazzolla Piazzolla, A. (1959). AdiÓs Nonino.

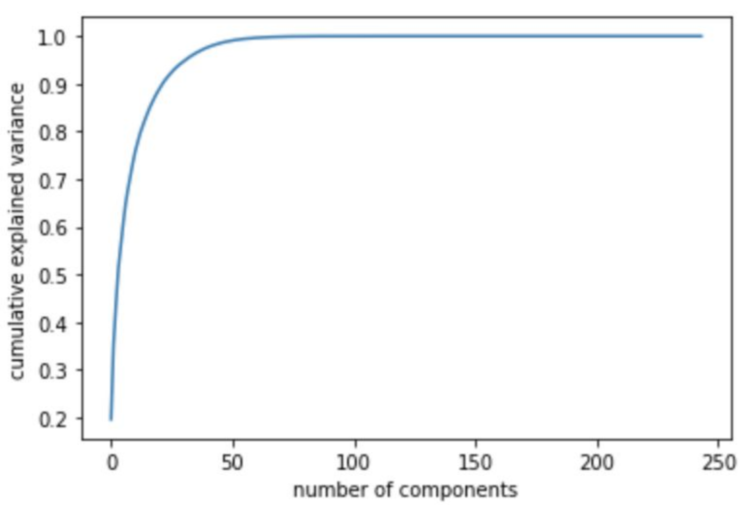
Pipeline:



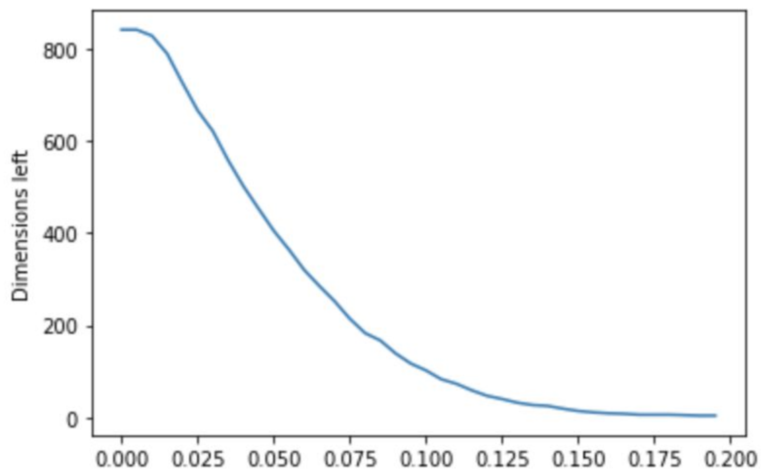
Procedure:

1. **Parcellation:** Defining distinct partitions in the brain, be they areas or networks that comprise multiple discontinuous but closely interacting regions. We used Craddock parcellation to reduce the data from (228453 x 244) to (841 x 244) for each participant i.e. it clustered voxels into 841 different regions.

2. **Dimensionality Reduction:** As the number of features increase, the number of samples also increases proportionally. The more features we have, the more number of samples we will need to have all combinations of feature values well represented in our sample. But we have only 38 samples so we need to reduce the dimension even further from 841. We used two techniques for the same purpose :
    a. **PCA:** PCA is a projection based method which transforms the data by projecting it onto a set of orthogonal axes. We used the following plot to know how many components to keep. From the plot we can see around 50 components capture more than 90% of the cumulative variance.



b. **Variance Thresholding:** Feature selector that removes all low-variance features. Here the key idea is that features with low variance contain less information. In our case variance relates to activity in that particular region of the brain. So if a region has low variance implying have lower activity and hence not useful for our purpose. This is the plot showing the number of dimensions left vs the threshold. So keeping in mind the curse of dimensionality, the threshold was kept at 0.1



3. **Regression Model:** Regression models are used to predict a continuous value. In our case, we have to predict the aptitude scores which can take any value. First, we did 80:20 data split for training and testing sets. We used the following regression models :
   a. **Linear regression:** This is one of the most common and interesting types of Regression technique. But only works when there is a linear relationship between target variable Y and the input variable X. In our case, there is clearly no linear relation. So the results were as expected poor.
   b. **MLP:** A multilayer perceptron (MLP) is a deep, artificial neural network. It can find non-linear relations also between target variable Y and the input variable X. The architecture we used :
      i. Input layer size : 50*244 = 12200
      ii. Hidden layer size: We tried a bunch of different hidden layer sizes. 10000*10000 worked pretty well relatively.
      iii. Output layer size: 1

   The R2 score ranged from -190 to -0.05 but we never got a positive score implying they were nothing better than spitting random outputs.

Why regression models won't work?

The total number of samples was 38. Out of which we made an 80:20 split to make the train and the test set. So we finally had 30 samples to train the model. A neural network requires a much higher number of samples to be trained effectively. Any regression model with high dimensional input value with low variance would require a much higher number of samples.
On the other hand, classification models can be trained with a much lower number of samples than regression models due to reduced complexity in the output. Hence, we converted the problem into a binary classification one. We applied a median split on scores so the task was converted into a binary classification task.

Binary Classification:

Now instead of regression, we tried binary classification. We tried the following models :

I. SVM
II. Logistic regression
III. Random forests
IV. MLP

So these classifiers were used to predict whether a person has a high(1) or low(0) aptitude score.

Experiments:

1. We tried all the above regression as well as classification models without parcellation as well. The reason being after parcellation many voxels merge to form a single region but there is a possibility that some voxels of a region are not active at all and others are. So considering this possibility we tried directly applying variance thresholding on voxels.
2. After the parcellation, the input vector size was (38x841x244). So for dimensionality reduction, we wanted to select some regions out of 841. So we used PCA in the following two ways :
    a. PCA was fit into a matrix of size (841x244) made from taking the mean of all samples. Then the input vector was transformed using this PCA.
    b. Taking mean across all timestamps for all samples (samplesx841x1) and then applying PCA/VT.

Key Results:

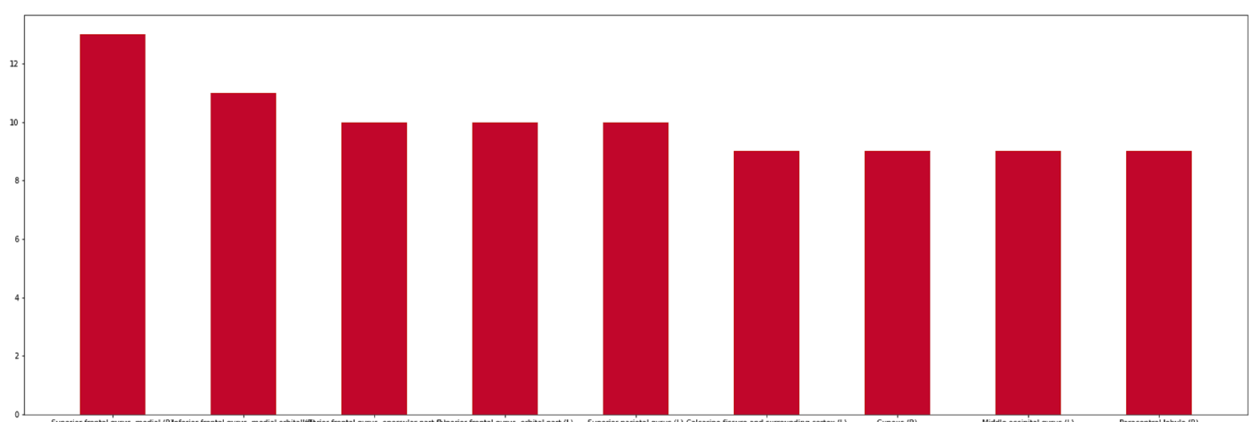
Model	Accuracy (5-fold cross-validation)
SVM with only variance thresholding	0.8
Logistic regression with only VT	0.6
Logistic regression with parcellation and VT	0.5
SVM with only parcellation	0.75
SVM with parcellation and PCA	0.8

Note: Here the accuracies are not a true measure of how good the model is because the test set consisted of only 8 samples. We need a much higher number of test samples for knowing how good these classifiers are.

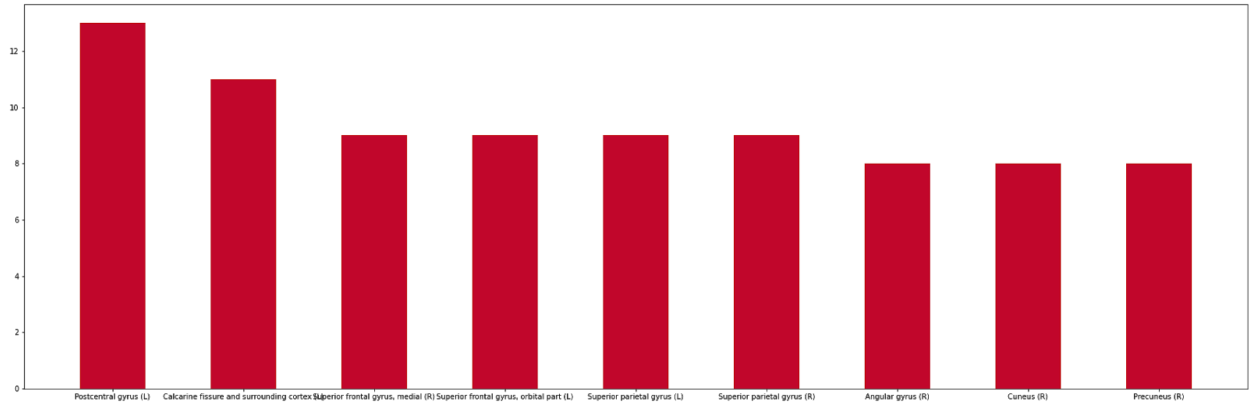
Backtracking:

The task is to find which regions of the brain are most active while listening to music for musicians and non-musicians. We used marsbar parcellation for the same purpose and extracted 117 regions. For each region, we check the number of participants for whom the particular region has exceeded the variance threshold. We then map these region numbers to the names of the regions to get two different dicts, one for the musicians and one for the non-musicians. Then we take the regions with the maximum frequencies(9 regions) and plot them to receive the following histograms.

For musicians



For non-musicians



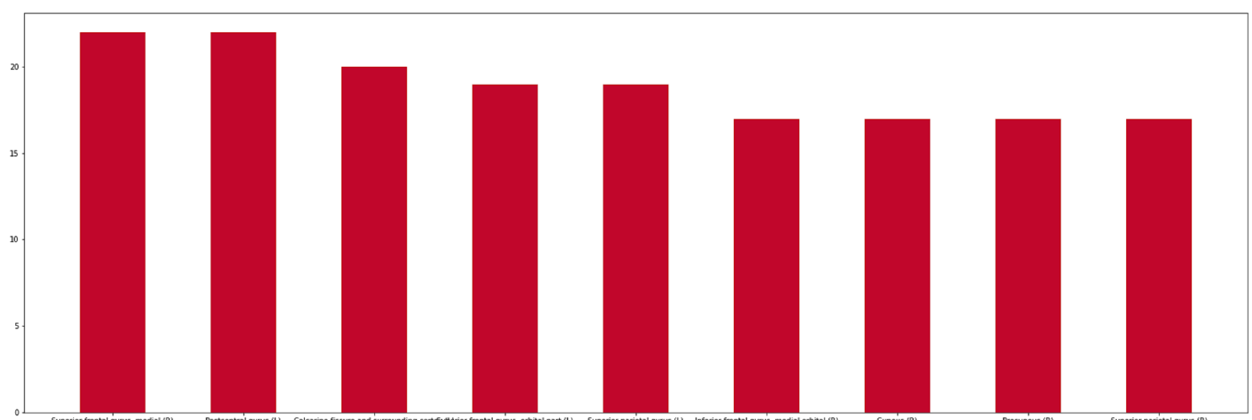
For musicians, the top 3 regions were:

1. Superior frontal gyrus, medial(R) - 13
2. Inferior frontal gyrus, medial orbital(R) - 11
3. Inferior frontal gyrus, opercular part (L) - 10

For non-musicians, the top 3 regions were:

1. Postcentral gyrus (L) - 13
2. Calcarine fissure and surrounding cortex (L) - 11
3. Superior frontal gyrus, medial(R) - 9

Next, we combined these results to see which regions were used the most overall, i.e in the case of musicians AND non-musicians. We got the following graph-



The top 3 regions overall were:

1. Superior frontal gyrus, medial(R) - 22
2. Inferior frontal gyrus, medial orbital(R) - 17
3. Inferior frontal gyrus, opercular part(L) - 16