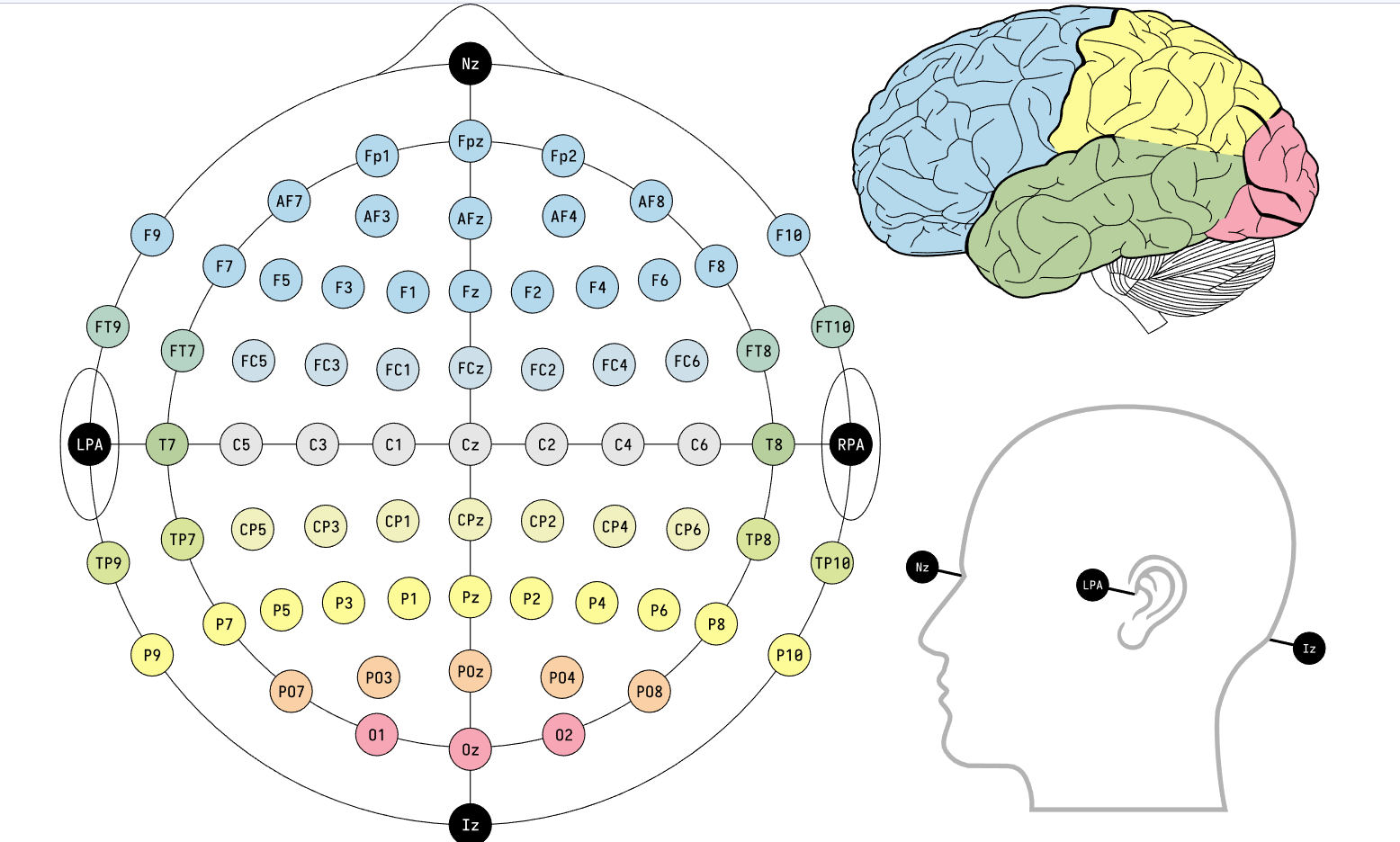
Data description -

We have the data from Muse which is an electrode containing device that has the capability to record brain activity(EEG). It is cheap and is used for research purposes as well as a commercial product.

!! image of Muse EEG 2

!! image of the electrode and there location on the brain

TP9, TP10, AF7, AF8, Fpz



The data was collected over one subject , who has been shown the various numbers on the mnist dataset. The stimuli were the numbers and the reaction was observed from the stimuli. Five electrodes were recording the signals from the brain. There locations are

Af7, af8, tp10, tp9, Fpz.

Out of these five , three were observed to have too much noise in them , so they were taken out of the data.

Therefore, the data consist of nearly 10K recordings of 2secs each ( at 256 Hz sampling frequency) , this means that we have 2 sec \* 256 Hz \* 2 electrodes = 1024 inputs for every line. Compared to that we have 10 K data points. Ideally speaking , this is a decent amount of data. However, depending upon the observations it could be way less too.

I might work on another data if I understand it better, the data is regarding the brain signals relating to emotions when subjected to various film clips that induce different emotions. For this experiment , six different film clips were used and there corresponding emotions were captured. Those clips are -

!! write the names of the clips and their corresponding emotions

This data has to be explored more. However , from what I believe it seems that emotions have a larger impact on brain signals than just numerical numbers. This means that the emotions will be far clearly visible in the EEG signals. One potential concern about this data is that it is a very much specific dataset. We know that emotions are very funzzy and very personal. They also may change from time to time , depending upon how the subject’s mental health is at the time of the experiment. Another concern is that if the data has been collected at about the same time, then it might have emotions that comes from the whole day , rather than just that moment. .

This data is collected using Muse 2 device which is the same as the device that has been used to collect the data for MNIST MIndBigdata dataset. However, there are far too many data points.

Learnings from the project -

1. Brain Signal Characteristics and familiarization
2. Machine Learning on brain EEG signal Data.
3. Neural Networks and transformations.

Justifications on the experiments performed -

1. No preprocessing done -

I believe that that there isn’t any need of preprocessing of data as that could lead to loss of information and neural network will learn the features even on the uncleaned data if given enough complexity

1. Usage of decision tree models -
   1. Decision tree classification will not work because of its re;ative simplicity but that it why we are using it, to show that brain signals are sequential data that needs to be dealt with models taht are highly complex.
2. Usafe of Neural Network -

We have implemented a fairly complex neural network model and that is not working , the reason for that could be that while neural networks are capable of finding pattern in the data that are complex, they are not the most optimal option when we know that the data is sequential and highly interdependent.

1. Usage of CNN - This is interesting because we want to use the CNN models to predict the number in the images. The rationale for trying this is that we believe that the CNN unlike the NN , can rather quickly learn the underlying patters relative to the adjacent data. The data in our case is adjacent wrt to time. Maybe it will give good outputs if ti is able to learn long term features in the dataset.
2. Usage of LSTM - we know that the signals are time varying and there changes through the time overall plays an important role in the identification of numbers or emotions. Therefore , using the LSTM seems like a good choice , considering it assumes the time varying nature fo the input. We will see the accuracy of that model also .

April 13 2023 , Researched LSTM

Tried to make an LSTM model from scratch on pytorch , could n’t do it .

Possible reasons for not being able to do it -

1. Incomplete understanding of the theory
2. Vector dimensions in the torch not clear
3. Vector dimensions in case of the LSTM not clear
4. Usage of LSTM not clear as of not
5. Not clear how LSTMs are used for voice recognition

Solutions

1. Research LSTM thoroughly , try to code it from scratch

Next trial -

1. Use nn.LSTM to get the outputs and then be done with it,

April 17 , 2023

Tried the cnn, didn’t work , peculiar pattern of loss observed, not overfitting, not training well at all, the reason that the model didn’t work could be that the more cnn is inhibiting the training

Things to try -

1. LeakyReLU()
2. Lesser CNN layers - not working , the model overfits without giving a good accuracy on the testing data.

Observations -

1. More model complexity will not help. We see that the model already overfits the data and adding more FC layers doesn’t benefit the training process at all.
2. More Convolutions layer might be a good alternative, but they take an insane amount of time to train.

Next endeavors -

1. LSTM - fit an LSTM
   1. Challenges - I am really not able to understand how LSTM works . ?
2. Transformers -
   1. Challenges - have no idea what it is ?

Apr 22, 2023:

Since, I couldn’t really get any good result on the dataset. It turns out that no one else was able to get any better results. The maximum reported were 12.5%, which is not greater that the random accuracy will fetch you.

Solution -

1. Find the reason behind why the data has not yeirlded any good accuracy ?
2. Start experimenting on the EEG dataset of Emotions when the emotions are evoked. The dataset on the kaggle.

Apr 23 , 2023 :

1. Started working on the dataset on kaggle. The dataset is fairly simple, it has many components in it but as it turns out it doesn’t really need a complex model to perform well. Although a neural network was sufficient to bring the right accuracy for the model.
2. Using CNN didn’t really give any additional benefits. And using LSTM network was not really useful as well.
3. I didn’t do enough scripting in this sheet which is not a good thing for presentation.
4. I have been able to achieve pretty high accuracy on this dataset using xgboost, and decision also . Even decision tree with a depth of 3 did a good job at this.

TODO :

Print the decision tree so we can see the results of the decision tree. Paste here.

How to show the model performance.? The accuracy could be the average of the f1 scores of different class labels.

This is the structure of a decision tree of depth three that gives an accuracy of 0.747.but i don't know what it does mean. I think it is the accuracy that the model performed on the dataset.

