Deep Brain Inference from Cortical Functional Magnetic Resonance Imaging Measurements

Marc R. Schlichting
Department of Aeronautics and Astronautics
Stanford University
Stanford, CA 94305
mschl@stanford.edu

Abstract

Getting insight in the mechanics of the brain is a crucial aspect for a numerous open research questions in the field of neuroscience and medicine. Unfortunately, getting access to the deep brain in a non-invasive manner requires bulky and expensive equipment in the form of functional magnetic resonance imaging. However, measurements of cortical regions can be easier obtained. In this paper a method using linear regression, k-nearest neighbor regression, and neural networks is proposed that successfully solves the problem of mapping cortical brain activations to deep brain activations.

1 Introduction and Related Work

Functional magnetic resonance imaging (fMRI) is one of the most powerful tools in neuroscience research when it comes to non-invasive data-acquiring techniques. fMRI measurements have the big advantage of being able to provide measurements from the entire brain, including subcortical regions with a frequency of about 0.5Hz. The measured quantity is the *blood-oxygen-level dependent* (BOLD) response, which is a quantity that can be taken to describe the activity of the brain. Compared to other non-invasive methods such as *functional near infrared spectroscopy* (fNIRS), fMRI can capture the activity of the entire brain. Unfortunately, fMRI scanners are not very portable, hence, experiments are restricted to a specific environment. For many applications of interest in neuroscience such as the effect of stress, it is required to take measurements in real-world situations such as for pilots during final approach.

Naturally, these measurements require more portable techniques such as fNIRS or *electroencephalography* (EEG). Decades of brain research indicate that there is a strong correlation between cortical brain activity and sub-cortical brain activity [1]. For stress, e.g., the subcortical region called Amygdala is essential, but the activity cannot be measured with cortical measurement devices such as fNIRS or EEG. The idea of this project is to apply machine learning methods to an fMRI dataset to predict subcortical activity from cortical activity. Instead of using actual congruent fNIRS and fMRI measurements, a combination where only very few datasets exist, the task will be split so that for this project only the mapping of the cortical fMRI data to the subcortical fMRI data is the goal. Therefore, in a first step the raw data shall be preprocessed. Once the time series for the different regions of interest are extracted, different machine learning methods can be tried. In this paper, linear regression, k-nearest neighbors, and neural networks are used. A similar idea has been explored using support vector regression [2]. However, the naive assumption to directly map the fNIRS signals to the deep brain regions has been used, a strong assumption which causes issues because of the different quantities measured by fNIRS and fMRI. Approaches for applying machine learning methods to fMRI data have been broadly outlined in [3], while the idea of applying machine learning to a regression problem with fMRI data has been been explored, however, only for resting-state networks and not for actual tasks [4]. The opposite task, predicting movements from

CS229, Spring 2021, Stanford University.

brain readings has also been attempted [5], but none of the papers so far, has actually dealt with mapping cortical fMRI measurements to subccortical fMRI measurements directly.

2 Dataset and Features

The dataset that is used for this paper originates from a 2019 study that examined inhibitory processes for a group of 150 children (mean age 11.6 ± 1.5 years), wherein 62% of the subjects had type 1 diabetes [6]. 47.4% of the participants were female while the rest were male. The subjects were asked, over the course of two runs (each run lasted $500\,s$, $\Delta t = 2\,s$), to perform the Go/NoGo task, wherein the participants are required to push a button based on visual stimuli. The study finds that children with type 1 diabetes show a higher activation of the prefrontal dorsolateral and supramarginal gyri regions. As this project does not aim to investigate the difference between the different subject groups, the aforementioned regions are excluded from the analysis for this paper. As not all subjects completed all the tasks, the final number of subject used for this paper is 145.

In addition to the two sequences captured in the fMRI scanner, there also exists a structural MRI image which is required for preprocessing. Preprocessing is done using FSL from the Oxford Centre for Functional MRI of the Brain [7]. During preprocessing, the data was motion corrected, spatially and temporally smoothed to reduced noise in the data, and only the brain was extracted. Additionally, the different brain regions were registered to the MNI512 standard space to allow for inter-subject comparison. For the registration, a 6 degree-of-freedom search to map the fMRI to the structural MRI image of the subject was used and a affine 12 degree-of-freedom search was used to map from the subject space to the MNI512 space.

For this paper, 20 different regions of interest are selected, 6 are subcortical regions and serve as targets, while the 14 cortical regions serve as features. The 6 subcortical regions which represent the targets are : left amygdala, right amygdala, left caudate, right caudate, left hippocampus, right hippocampus and the 14 cortical regions used as features are: frontal orbital cortex, frontal pole, inferior frontal gyrus, lingual gyrus, middle frontal gyrus, parahippocampal gyrus anterior, parahippocampal gyrus posterior, postcentral gyrus, precentral gyrus, superior frontal gyrus, superior parietal lobule, superior temporal gyrus anterior, superior temporal gyrus posterior, and temporal pole. Those features were selected to represent an uniform spread over the entire cortical region. For the extraction of the brain regions, the Harvard-Oxford cortical and subcortical probabilistic atlases were used [8]. While BOLD intensities still vary slightly from subject to subject, for ease of comparability, each individual subject and brain region has been normalized to zero mean and unit variance. For computational tractability, all the preprocessing was carried out on Sherlock, Stanford's high-performance computing cluster. The dataset was split into a training dataset (70%) and a test dataset (30%) for evaluation using the holdout method. A sample from the training dataset is depicted in Fig. 1.

3 Methods

Three different supervised learning methods are used, all tested with and without a rolling window to include time dependencies. Let the dataset be denoted as $\{x^{(i)}, y^{(i)}\}_{i=1}^{n}$, where x are the features and y are the targets. $\hat{y} = h_\theta(x)$ denotes the model's prediction. The design matrix $X \in \mathbb{R}^{n \times d}$ contains the samples row-wise, similarly, $Y \in \mathbb{R}^{n \times m}$ contains the row-wise sample of the targets.

3.1 Linear Regression

In linear regression, a hyperplane which is linear in the features is fitted using a least squares cost function. While possibilities exist for regularization (Lasso regression and Ridge regression), they are not explored in this paper. The hypothesis function in terms of the design matrix then is:

$$h_\theta(X) = XW^T + b,$$

where $W \in \mathbb{R}^{m \times d}$ are the weights and $b \in \mathbb{R}^m$ is the intercept term. Another way to interpret linear regression is as as finding an "optimal", linear combination of the features. For ordinary least squares, this problem can be solved using the normal equations:

$$W = (X^T X)^{-1} X^T Y$$

By prepending a column of ones to X, this equation can also be used to solve for the intercept term. However, as the data was normalized to zero mean, the intercept is expected to be zero.

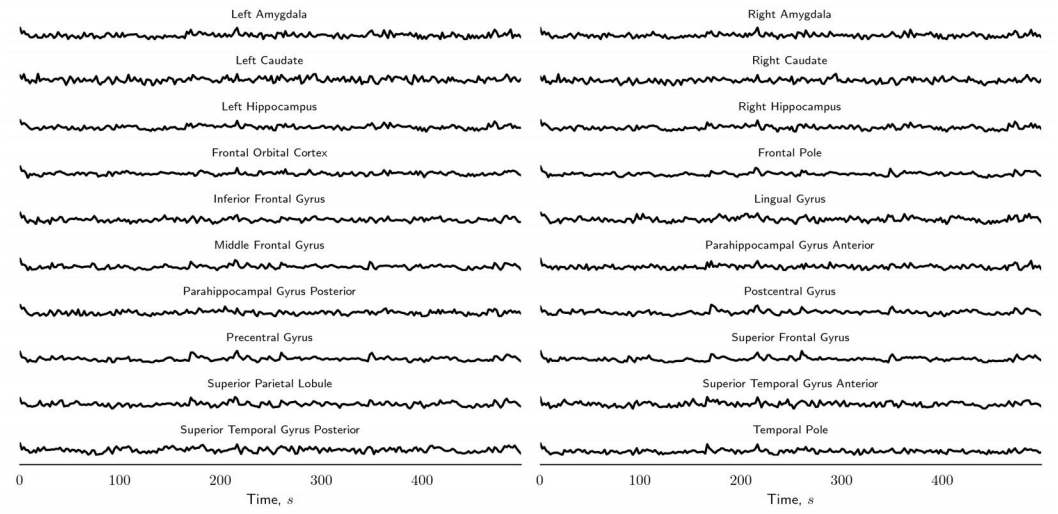

Figure 1: Normalized BOLD signals plotted for all 20 regions of interest. To improve readability, the scaling of the BOLD signal is not displayed.

3.2 K-Nearest Neighbors

The *K-nearest neighbor* method (KNN) is a parameter-free model as it only keeps the training data to make predictions. Inherently, it is a non-linear model, as for each input the output is calculated as the mean of the outputs of the k-nearest neighbors in the training set [9, 10]. The prediction can be formalized by assuming a set $\mathcal{S}$ which consist of pairs $(x^{(i)}, y^{(i)})$. Let $\tilde{x}$ be the feature in question for the prediction. The prediction can be obtained by ordering $\mathcal{S}$ in such a way that $\|\tilde{x} - x^{(1)}\|_2 \leq \|\tilde{x} - x^{(2)}\|_2 \leq \ldots$. Assuming such an ordered set, the prediction can obtained by:

$$h(\tilde{x}) = \frac{1}{k} \sum_{i=1}^{k} y^{(i)}$$

3.3 Neural Networks

A simple form of neural network that is used here is a *multi-layer perceptron* (MLP), which is an extension of the linear model described in section 3.1. There are two significant differences concerning the model architecture that distinguish the MLP from a linear model: As the name suggests, there are multiple layers that are concatenated, where the output of one layer is the input to the next layer. Secondly, to make the model non-linear, a non-linearity g is applied to each output of the linear layer with exception to the last layer to provide more flexibility to the output. The non-linearity is important to make the model more expressive. Each layer's output can be described as:

$$z^{[k]} = g\left(X\left(W^{[k]}\right)^T + b^{[k]}\right)$$

All intermediate values of $z^{[k]}$ that are not the output, are referred to as hidden layers (with k as enumerator) and the dimension of each state is colloquially referred to as the number of neurons. For training, gradient-based methods are used, as no closed-form solution like the normal equations exist. A typical choice is the Adam optimizer as it outperforms many other optimizers and delivers robust performance [11]. To prevent overfitting, the weights but not that biases, are L2-regularized.

3.4 Rolling Time Windows

So far, it was assumed that all the data points are independent of each other, however, since the goal is to analyze a time series, this cannot be assumed by default. For this reason, the features for each of the 14 cortical regions are augmented by previous values for the same region. E.g., instead of only using the measurement of the frontal orbital cortex at time t, the measurements up to $t - p$, where $p + 1$ is the window length, is also included as a feature.

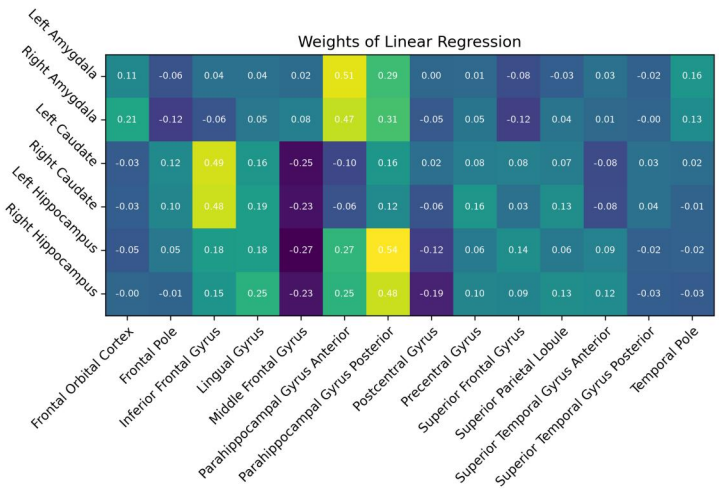

Figure 2: Weights for linear regression without rolling window.

4 Results

4.1 Metrics

To evaluate the performance of each algorithm, the *mean squared error* (MSE) is calculated for the test dataset for each each of the six subcortical regions (subscript s) separately. This is done to assess the prediction quality for each of the subcortical target areas individually:

$$MSE_s = \frac{1}{n}\sum_{i=1}^{n}\left(y_s^{(i)} - h_\theta(x^{(i)})_s\right)^2$$

For a baseline comparison, samples of the same size as the test dataset are drawn from the standard normal distribution, to match the normalized dataset. These results are tabulated in Table 1.

4.2 Linear Regression

Despite the simple nature of linear regression, even if only using point estimates, very good and accurate results can be achieved. With exception to the caudate, linear regression outperforms the KNN and MLP. Of special interest are the weights of the linear model as these shed light on how each feature influences the outcome. Fig. 2 shows three distinct cortical regions (in yellow) with high correlation to the respective subcortical region. It is also worth noting that the middle frontal gyrus appears to have an opposing role, which is indicated by the negative weight. Using a rolling window of size 3 ($p = 2$) which is equivalent to a 6 s window, slightly decreases the MSE as denoted in Table 1, however, evaluating the weights shows that there is no strong correlation between previous time steps and the predicted subcortical activation. Hence, it can be concluded that for this study, barely any connection beyond the scope of Δt could be detected.

4.3 K-Nearest Neighbors

Choosing k is a tradeoff between accuracy a smoothness. Choosing k too small results in the prediction being prone to noise, choosing k to large results in smoothing the result too much as samples that are farther away from the requested point influence the result too. Hyperparameter optimization yields an optimal $k = 20$ for the point estimates and $k = 10$ for the rolling window. The performance using the rolling window is significantly less accurate for the amygdala and hippocampus. Overall, the KNN method delivers inferior results compared to the linear regression and the MLP.

4.4 Neural Networks

Experimenting with different numbers of hidden layers and neurons per layer does not change the performance significantly, up to a point where overfitting becomes an issue (beginning at about 5000

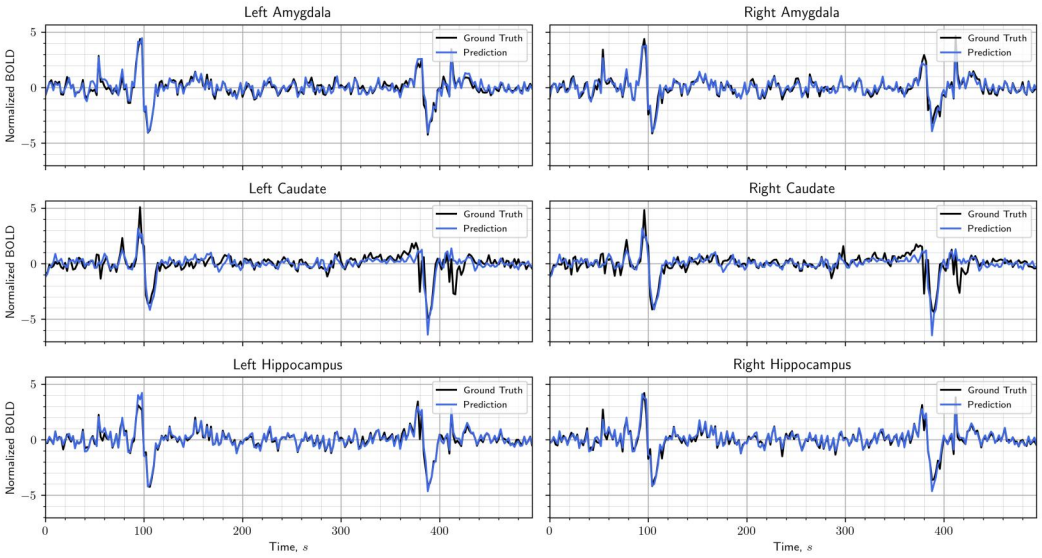

Figure 3: Predictions for MLP regression with rolling window on test dataset.

parameters). For the point estimate, two hidden layers, with 30 neurons each deliver good results and for the case with the rolling window, two layers with 48 neurons each are used. The learning rate is $\alpha = 0.001$ and the regularization parameter is $\lambda = 0.0001$. For training, minibatch gradient descent is used with $n_{MB} = 200$. For the activation function, the hyperbolic tangent was used. The performance is comparable to the linear regression with slightly better performance for the caudate. Using the rolling window improves the performance for the caudate, but the other regions' ac curacies become slightly worse (see values in Table 1). Fig. 3 visualizes the predictions for the six subcortical regions using part of the test dataset. In general, the predictions for th the caudate are the ones that are at least accurate, but still acceptable.

5 Conclusion

In this paper, it is shown how to approach the important topic of predicting deep brain activations if only cortical fMRI measurements are available. Preprocessing is an important step to clean the data from artifacts that happened during the data acquisition. Rather simple machine learning algorithms such as including linear regression, KNN, and MLPs are used, but especially linear regression and MLPs deliver very good results. For future research, it could be interesting to evaluate more brain areas and use other techniques such as independent component analysis to extract relevant and independent data to get even more insight in the functional connectivity between different brain regions. Eventually the goal is to make good predictions about deep brain activity from cortical measurements in various real-life scenarios which would be a big advancement for neuroscience and medicine.

Table 1: MSE comparison of different learning methods on test dataset

Method	Amygdala		Caudate		Hippocampus	
	Left	Right	Left	Right	Left	Right
Random $\sim \mathcal{N}(0, 1)$	1.9942	1.9919	1.9851	2.0123	2.0129	2.0225
Linear Regression	0.1691	0.1886	0.5202	0.5115	0.1230	0.1506
Lin. Reg. w/ roll. window	0.1664	0.1864	0.4819	0.4753	0.1154	0.1437
KNN ($k = 20$)	0.1992	0.2131	0.5280	0.5175	0.1595	0.1812
KNN ($k = 10$) w/ roll. window	0.2899	0.2967	0.5292	0.5211	0.2367	0.2508
MLP (30/30 hidden layers)	0.1731	0.1917	0.5198	0.5096	0.1292	0.1544
MLP (48/48 hidden layers) w/ r.w.	0.1788	0.1976	0.4724	0.4648	0.1268	0.1515

Acknowledgements

The data for this work was kindly provided through the Center for Interdisciplinary Brain Research at Stanford University (CIBSR). Even though the dataset is anonymized, the author has not the right to share the dataset with the course staff, however, if further information is required, or would be helpful, the author is happy to answer questions. The code can be found under the following link: `https://office365stanford-my.sharepoint.com/:f:/g/personal/mschl_stanford_edu/Eh2rXsfO8XNEgPsESFiqi-wB-7YUeFXZeN-1i6NO2J6dJQ?e=Psjj3c`.

References

[1] Marc D Lewis and Rebecca M Todd. The self-regulating brain: Cortical-subcortical feedback and the development of intelligent action. *Cognitive Development*, 22(4):406–430, 2007.

[2] Ning Liu, Xu Cui, Daniel M Bryant, Gary H Glover, and Allan L Reiss. Inferring deep-brain activity from cortical activity using functional near-infrared spectroscopy. *Biomedical optics express*, 6(3):1074–1089, 2015.

[3] Elia Formisano, Federico De Martino, and Giancarlo Valente. Multivariate analysis of fmri time series: classification and regression of brain responses using machine learning. *Magnetic resonance imaging*, 26(7):921–934, 2008.

[4] Alexander D Cohen, Ziyi Chen, Oiwi Parker Jones, Chen Niu, and Yang Wang. Regression-based machine-learning approaches to predict task activation using resting-state fmri. *Human brain mapping*, 41(3):815–826, 2020.

[5] Zaixu Cui and Gaolang Gong. The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features. *Neuroimage*, 178:622–637, 2018.

[6] Lara C Foland-Ross, Bruce Buckingam, Nelly Mauras, Ana Maria Arbelaez, William V Tamborlane, Eva Tsalikian, Allison Cato, Gabby Tong, Kimberly Englert, Paul K Mazaika, et al. Executive task-based brain function in children with type 1 diabetes: an observational study. *PLoS medicine*, 16(12):e1002979, 2019.

[7] Stephen M Smith, Mark Jenkinson, Mark W Woolrich, Christian F Beckmann, Timothy EJ Behrens, Heidi Johansen-Berg, Peter R Bannister, Marilena De Luca, Ivana Drobnjak, David E Flitney, et al. Advances in functional and structural mr image analysis and implementation as fsl. *Neuroimage*, 23:S208–S219, 2004.

[8] Rahul S Desikan, Florent Ségonne, Bruce Fischl, Brian T Quinn, Bradford C Dickerson, Deborah Blacker, Randy L Buckner, Anders M Dale, R Paul Maguire, Bradley T Hyman, et al. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage*, 31(3):968–980, 2006.

[9] Evelyn Fix. *Discriminatory analysis: nonparametric discrimination, consistency properties*, volume 1. USAF school of Aviation Medicine, 1985.

[10] Naomi S Altman. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185, 1992.

[11] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.