Miasya Bulger
MAIS 202
February 9th, 2019

Data Selection Proposal: EEG Grasp and Lift Detection

Dataset and reasoning

For my project, I have selected the Kaggle EEG (Electroencephalography) Grasp and Lift detection dataset (https://www.kaggle.com/c/grasp-and-lift-eeg-detection/data). This dataset is comprised of several csv files containing tagged numerical information regarding the activity of the brain of test subjects performing grasp and lift actions. This dataset appeals to me because it exists at the interface between human biological systems and engineering, which are two topics that absolutely fascinate me. The potential applications of EEG data processing, in particular, are extremely relevant to the design of contemporary brain-machine interfaces to act as assistive devices for the neurologically impaired. Finally, I hope that through this project, I can gain a foundational basis with EEG interpretation using machine learning so in the future I can record my own EEG data and have the skills to get me started on a personal project. In short, the combination of a personal interest in bioengineering and a curiosity of machine learning motivated me to select this dataset.

Methodology (overview of approach):

Data Preprocessing

Fortunately, we have quite a bit of data already labelled among several isolated files already split into training, validation, and testing sets. The data (approx. 200 csv files), comes almost ready-to-use because this dataset has been previously featured in a $10,000 competition sponsored by the WAY Consortium. For me, the state of the data was definitely a large factor when selecting this data because I understand the noisy nature of biological data, and as a beginner, I hope to work on establishing a strong overall understanding of the many steps in developing a machine learning model and consequently I need to avoid getting stuck on data preprocessing early in my stages of development. And in all honesty, if I choose to enter the computational biology industry in the future, I'll always have the opportunity to spend my time practicing data processing there.

Machine Learning Model

I believe a generalized neural net approach best suits this dataset. Often, EEG information contains a lot of noise due to the continuous activity of the brain, and the interpretation of this information tends to be not easily conceptualized. However, I predict that several (several) nets will effectively be able to best construct a mathematical basis for helping us reach our desired conclusions with considerable accuracy.

Final conceptualization

I haven't chosen between a poster project an application (more specifically, a webpage or article about it). I'm sure I would enjoy either option, so I would love to hear your advice on what you think is best!

Application specifications:

See above ☺