Project Title: Predicting Fatigue From Performance

Team Members: Nastasia Klevak
Emails: nklevak@stanford.edu

1 Motivation

This is an application project which plans to use data from two cognitive psychology datasets. These datasets both come from the same behavioral experiment (conducted two different times with slightly different parameters), which aimed to look at how task switching and performance interact to influence future performance and cognitive fatigue. In each experiment, participants were taught to play 3 games (spatial recall, digit span, and a simple rest game). Then, participants had 30 epochs of 10 trials each. Every epoch they were told whether the next epoch would be a switch or a stay (switch meaning digit span if they were currently playing spatial recall, and vice versa). Additionally, in between every epoch there was a rest period where participants could rest from 1 to 20 trials (as long as they wanted) in exchange for points they paid. In my analysis of these datasets so far, I've primarily looked at the interaction between previous performance and task switching on next-block performance using generalized linear models. This has had interesting results but has not tackled an important remaining question: how does fatigue (as measured by amount of time people choose to spend resting) fluctuate through the course of the experiment, and what factors play a role in these fluctuations?

To answer this main question, for the CS229 final project I plan to use ML techniques from this class (and some extensions) to see how well we can predict a subject's rest length at every epoch. Existing literature suggests that cognitive fatigue grows over time, and is at least partially a function of how well a person has performed on a task thus far. As a result, I'm interested in looking at how well different methods work to predict rest length, and which features are key to better predictions.

2 Methods

I'm planning to apply several different methods, as well as compare between them.

As simple baselines, I want to select features that represent an epoch of a task (avg rt, avg accuracy, accuracy sd, rt variability, game type, if the task switched, block number) and use them to predict the amount of rest in the following epoch by each subject. To do so, I plan to use ridge regression, and then create another baseline using boosted decision trees.

Next, I want to improve upon these baselines by adding in features that represent the history of the experiment (I.e. amount of rest in previous epoch, performance in last epoch, rt in last epoch) and re-running these same methods. If adding in these additional features improves the predictions (as I hypothesize it will), this would suggest that fatigue is more than just a function of the block directly preceding it (which is what the theory would predict).

As a final predictive model, I'm also interested in trying to fit a hierarchical LSTM (which I know is outside of the scope of this class) in order to better incorporate all prior blocks when making a fatigue prediction. For this, I would train on entire experimental sequences for some subjects and try to predict all 30 rest lengths for other subjects.

Finally, it would be interesting to cluster participants using k means clustering in order to understand different rest patterns, and to then compare these methods within each cluster to see if some are easier to predict than others.

3 Intended Experiments

To execute the ideas described in the methods section, I will first separate my two datasets (83 subjects, 104 subjects) into test and validation sets. I would be most interested in training on one full dataset and validating on the other, but if this split ends up being unfeasible I plan to cross the datasets to use a more common split.

Then, I plan to use the training set to train each of the methods I mentioned above, and then test it on the validation set. Since each method is meant to predict rest length, I can calculate how far each prediction is from the true value and try to minimize it. Then when using the validation set, I will calculate the MAE as a metric of success of the method. Additionally for each method, I plan to plot the distributions of predicted rest length and compare it to the human distribution (and potentially calculate the Wasserstein distance to calculate the difference between the distrbutions). I will also calculate the R squared of each method. Additionally, for each method other than the LSTM, I want to look at the weights of each feature to see what was most important in the prediction.

Finally, I plan to use k means clustering to understand different fatigue patterns in the data, and evaluate how different methods perform in the different groups. It would be interesting to find that certain features matter more in one group and less in another, for example.

4 Team Contributions

- **Nastasia Klevak**: I (Nastasia) am the only team member so I ideated and wrote up this proposal on my own.