Report

I initially planned on using a machine learning model (probably linear regression), but then realized the big number of features would not allow for this, instead I should use deep learning. I have not done something with deep learning yet, but it was a great opportunity to learn.

I first thought I should quantify my features and make them all in numbers and in one size. I checked the endpoints given and wrote some code that concluded the maximum difference between a start point and an endpoint is 190, so I set the number of features to be 190. Then I wrote some code that puts a zero for the features outside the start point and endpoint given. I then wrote a script that isolated each the 53 labels and made each label equal to a number (from 1 to 53), this basically means there are 53 possible outputs. I also wrote a script that converted the data given from feneme form to numbers, by replacing every feneme by its index in the feneme_set.txt file. By converting everything, I figured we should be ready to use the pytorch data loader.

Previously I used Scikit learn for ML models, but since this is a deep learning model, I will be using PyTorch. I created 2 hidden layers and the classic forward function using relu and softmax. After that, it's time to calculate loss, so I set the gradients to 0 and try to reshape the batch, I ran into an error when I test ran it, it had to do with the fact that x was not a tensor. I moved on and figured I'll get back to it since I had the majority of the program planned out. I then grab the loss value and perform back propagation on the network. I went back to the issue and realized it had to do with the way I inputted data.

While I did use the pytorch data loader, it returned a list of tensors and not one tensor like what was needed. I tried to convert it to a tensor afterwards, but I would end up with errors like "'int' object is not subscriptable". I tried different methods of importing the data in text format but had no luck.

I was disappointed at myself that something that had nothing to do with the neural net itself stopped my path in this challenge. I do admit I should've freed up more time for the second part of the challenge, but I was on vacation and when I got back, I got busy with other scheduled interviews. I apologize for not completing the challenge, but please consider what I have submitted, and I will prove better in the next step.

Thank you for giving me the opportunity to interview for SEASALT.AI

Bishoy B Youhana