24NOV22 Note to Avraham

Hey Avraham,

Again - sorry for the delay here. But, I have finally taken a look through what you sent and have a few comments and questions:

1. General structure of the .npz file
   1. For some reason, the file you shared (0.npz) had 200 samples of the actual spectrum data ("fi") but only 10 samples of the rest of the parameters. Not sure why that is. Obviously, when you make the full datasets, I will need all of the info for each sample
   2. My plan is not to touch (at all) the period, time\_list, or v1/v2 data unless you tell me I should
2. Spectrum data (“fi”)
   1. I probably will need a longer explanation of what has changed between the previous data (~10k pixels) and the new data (only 973)
   2. This is a pretty large reduction of input data, though honestly it may be a good thing. It certainly will make training faster and the dataset easier to manage
   3. Our previous models (and specifically the best performing model, based on convolutions) used the full ~10k input data. Because of this, it is a little difficult to predict how we will do on the Gaia data
   4. My recommendation/plan is to generate a large portion of this data, and simply retrain the model using the new dataset
   5. I also plan to retrain a model using the old data, but with some sort of 10x compression to get it down to 973 pixels. If you have a recommendation on how I can compress the old data to have it most closely resemble Gaia data, please let me know
3. Luminosity
   1. Happy to hear that the luminosities are in log scale. As long as this works for you, it works for me
4. Other data categories
   1. One of the key things that we used last time was the normalization of temp/alpha/metallicity/etc values based on the mean and std deviation values across the dataset. Is there any reason to think that the Gaia data would not closely resemble the previous data, with regards to mean, std dev, range, etc?