Semester 2 Week 6 meeting notes

|  |
| --- |
| Discussion of training of neural nets from the previous week  Harry figured out that you can get the decayed learning rate at the end training so that it can be inputted for the next leg of the training if recompiling is done so as to keep the decay schedule as consistent as possible and to minimize the effects of recompiling on the loss.  Unfortunately we haven’t found out yet how to alter the decay schedule as when you recompile, because of how the learning decay is calculated from the number of elapsed epochs when you recompile that number goes back to 1, which results in some deviation from the initial decay schedule, though this is dependent on how rigorous the decay being use is (which is another thing Harry found out could be used).  Harry wrote a script to fit a decay to a decay schedule using scipy’s Best\_fit function in order to further minimize the effects of recompiling  The best curve fit stuff can be put into the final report but we need to demonstrate that what we’re doing is not ridiculous.  Normally Guy would suggest we restrict the final report what has worked but as we’re pushing the boundaries of training neural networks and that should be reflected in the final report.  We need to be able to explain why making certain changes to the neural network setup, improved it.  Guy says it’s okay if Harry focusses more on the neural network side but still discusses the science of HBM stuff.  NGC 6819 data is in table 3 or 4 after all the text in Guy’s paper that he sent |
| Discussion of HBM stuff  5 star pseudo cluster constructed from the grid gave very good results  Guy has nodes/cores we can use on bluebear  How you’re allowed to use priors: if you set your prior and you do a run and the result looks wrong then you can set the prior up to exclude that result.  We are having an issue where the HBM is sampling outside the prior given by a beta function which causes it to crash  Guy’s suggested test: remove observational constraints, so we can see where the values HBM is sampling. If the values that are being sampled is wrong than the priors or neural network or something is wrong. |
| Guy has started to write the paper that he is going to put us down for co-authorship  If we read Guy’s paper we should be very careful about how we use it to inspire the writing of our final report.  In Guy’s paper he says that - using a trial and error of the optimization which is influenced by our knowledge of how it works.  We will need to read the paper such that we have read that acknowledged that we agree with all the statements made in the paper, of course there are statements that we may not understand but we need to make comments/corrections.  When the draft is sufficiently close all the co-authors will be sent an email with some timescale to provide comments and there will be some corrections made and then sent to a referee  And it will be maybe a month between when the it is sent to the referee, to when we get the referee report.  Then there will be another round of corrections from us and then sent back.  Guy thinks it’s a few weeks off the co-authors being sent the email for comments. |
| In our report we should talk about:  correlations between parameters shown in the corner plots from the HBMs.  The grid of stellar models we’ve used is just one set of physics, which is okay on the scale of the year 4 project. |
| Brief discussion about Hin’s Exeter PHD interview.  Guy can Hin an email address for “Sean Matt”?  For Hin talking about his project, he should talk about method quickly but mostly talk about the science.  Hin could talk about using a multilevel model for talking about exoplaents  They are looking for: enthusiasm, understands what they are doing, and are careful about what they claim. |
| Harry’s discussion with Guy  Incremental regularization  I should try and put this into the paper  Guy has found an initial batch size of 32 (with a high learning rate e.g. 0.001) works well for a solid initial drop in the loss, with only 10 epochs.  When you start by randomizing the weights the neural net has no predictive ability but because all the stars we’re training on have quite similar temperatures, the first thing it tries to do is shift everything so that it only predicts sensible temperatures (being an initial pull in), which can be done quite quickly. Then it figures out the temperature changes as a function of one of the other parameters like mass and it tries to pull in.  Once you’re trying to get it to learn the bulk properties you can afford to have larger batch sizes. You don’t want the batch size to be too small that the realization of the batch matters but you also don’t want it to be so big that it is completely overwhelmed, so you want to give it instances of the data that are sort of representative but ONLY JUST.  Guy again suggested training multiple neural networks on the different sections of the data, potentially we should do 3 different ones.  Getting small details requires large batch sizes  Although Guy is of the opinion that you need to use different batch sizes to accomplish different tasks along training. So initially if you had batches with only 1 data point the neural net would just try to pull in to that point but then when it looks at the next single point batch try to instead pull to that point with the result of the neural net just pinging about everywhere not accomplishing anything. However, if you immediately were to use full batch then the neural net wouldn’t have any idea of where to start as it can’t choose a particular feature to start learning from. The initial just generalizing batch size allows the neural net to choose a particular feature to view the trend in and as you increase the batch it can use the trend it has already learned and then see how the additional data effects that trend.  We should say in the paper why you shouldn’t put the batch size too high or too low, and to describe how we came to this conclusion. |
| Discussion of whether renting a GPU would be a good idea |
| Discussion about data acquisition  NGC 6819, NGC6791, M4 (Andre paper, although doesn’t have many stars),  Diego Bassini has a paper with 273 stars without asteroseismic data which we could look at.  Guy recommends asteroseismic data. |

Guy to do:  
send Hin an email for Sean Matt if Hin wants (redundant now)