Week 9 meeting notes

|  |
| --- |
| Discussion of about Dex  We’re focus is on Teff, Luminosity and Delnu, which we need to convert into what we actually observe i.e. apparent magnitude with some distance estimator. |
| How regularization changes with architecture:  The flexibility of the model increases as we add in more neurons and more layers (as that also increases the number of neurons).  Regularization is the opposite as the larger the regularization the less flexible the model can be.  Methodology: get close with the architecture and then tune the regularization  (Guy said that initially he had the regularization as ”0” when determining the architecture  Additionally he said that he increased the architecture until he got an overfit and then he increased the regularization to undo the overfitting but in very small increments.)  L2 kernel regularizers = squares all the values of the weights then sums them up multiplies it by the regularization coefficient we give it and adds it to the loss function. So if the weights are large the loss function is penalised.  For the mid grid I believe Guy said he used a regularization of 1e-7.  So for a larger architecture the regularization needs to be smaller.  When increasing the architecture, you quickly reach a point when you’re adding more functional forms but how they differ from the current functional forms compared to what we think the functional form of the data should look like is essentially meaningless. i.e. the functional form is allowed to be more wiggly but on the most part our data is not as wiggly as what the neural net is allowing.  Consider: figuring out how to start a set of learning with a high learning rate but then start a learning rate at a lower value and be able to do that each time you start a new leg of the training on a certain model.  If you take a different validation set on a new leg of training it invalidates the validation.  But it doesn’t seem to invalidate the validation each time you run model.fit, where the validation fraction is stated.  At a certain point the dex is limited by the finite number of data points in the data, that the neural net can learn from. You can try to improve the dex with the data by interpolating between the data points.  Total error contributing to our inference = ((σobs)2+(σmodel\_term)2)0.5.  Showing that there is no point in having a dex a couple of orders of magnitude lower than the observational uncertainty. |
| Keep reading papers. |
| Discussion on PhDs |
|  |
|  |

Students To Do:  
To make it easier for Guy, when we ask for a neural net to be run, put file we want to be run in our sub-directory as Guy will just copy it there anyway.

Guy to do:

Next week: