Semester 2 Week 1 meeting notes

|  |
| --- |
| Characterising the properties, formation and evolution of open clusters.  Looking at spread in ages and gaining some belief in that spread.  Calibrating models of stellar evolution |
| Most seminar marks are going to go to structure, present and deliver the talk.  Instead of how good our method is. |
| Backup slides for caveats that might be questioned on  e.g. additional slide for activation function stuff  or for how we handle overfitting by using L2 regularization/dropout |
| Can explain HBMs by using it as a method of approaching some population like the seminar room:  Which we then can compare to a cluster e.g. if we have a age estimate of 2.5 gyrs for 1 star and 6 gyrs for another with an error of 2.5 gyrs on both, we would think that the mean of those 2 ages is closer to the true age of the cluster. |
| give credit to person who did the work e.g. Hin’s results from HBM |
| Hierarchical Bayesian modelling is using Bayesian inference but on a population of measurements simultaneously. |
| We can deal with problems of HBM exploring unphysical parameter space by enforcing stricter priors on the important parameters e.g. age |
| Guy has moved away from using L2 regularization and is instead testing dropout  Because he feels that dropout is easier to tune.  Traditionally dropout is turned on during training and turned off during predict however you an leave it on during predict which means that each time the same prediction is made a different answer will be given within the range of the error of the trained neural net.  With L2 regularization: increasing architecture would cause more overfitting.  With dropout that seems to be less true.  Guy found that the larger architectures would train faster i.e. have a steeper loss function.  Guy also changed from MAE to MSE during the training.  We could try to do some 2-step training whereby we train using MSE and then once we feel the MSE has controlled the “bad” points, we then continue the training using MAE |
| The neural net’s accuracy doesn’t need to meet our aims but we would need to right that up in the project as a caveat that to be really useful we need it to be more accurate. |
| because relu is not a smooth activation function it makes the neural net “spiky”.  Instead potentially use elu. |
| In our talks we should discuss why we chose a particular activation function.  We want a smooth output thus relu is gone, we could try others put if we don’t have time we could just use ‘swish’ and cite the paper that says it’s the best. |
| Variational inference = approximates posterior using sophisticated minimization technique. |
| In our final report we don’t need to explain everything ourselves like NUTS sampler for instance but should reference a paper that does mention the expects the paper explains. This is because the inner workings of NUTS aren’t particularly important for our purposes.  Instead for NUTS say why we chose it over other samplers. |
| M67 grid points will be added in due time  When training on the grid just train on solar metallicity and 0.001 about that metallicity.  Same for helium  Constrain age to maybe 1gyr – 8gyr |
| Think about using eclipsing binaries. |
| Make sure to reference specific information taken from a paper, especially graphs/tables  Though we don’t need a references slide. |
| Guy is going to Sydney from the 2nd Feb to the 17th Feb, we should still have meetings on skype. |
| 2nd March the head of data science at the at some laboratory that knows a lot about machine learning and we should decide whether we would like to bring him in on one of our meetings. |
| If we felt so inclined some additional work we could do would be to look at the paper: “SMOOTH FUNCTION APPROXIMATION BY DEEP NEURAL NETWORKS WITH GENERAL ACTIVATION FUNCTIONS”, url: <https://arxiv.org/pdf/1906.06903.pdf>  Which has a mathematical formulism that predicts how the error on the function approximation changes with neural net architecture.  If we get a working neural net that works in the HBM and we go back to improve the accuracy of the neural net we could use the paper to demonstrate how much it’s worth changing the architecture.  Data augmentation = adding data to neural network training  Paper may also contain information on how the neural networks improves with increasing the number of data points.  You increase the number of data points using Gaussian processes (which is much faster than by calculating more stellar evolutionary tracks.) |
|  |

Students To Do:  
- consider doing 2-step neural net training: 1st step with MSE, 2nd step with MAE   
- potentially train with google collab  
- think about whether it would be beneficial to meet with the head of data science at some laboratory who is coming ot the university on the 2nd March.   
- keep in mind working with eclipsing binaries.

Hin: (post seminar)  
- do more HBM work (hopefully sorting out bias issues)  
- potentially try working with clusters with more stars.

Harry: (potentially post seminar)  
- debug his neural network  
- test using dropout instead of L2 regularization  
- maybe work with the formulism for how error changes with architecture.

Guy to do:  
- send slide you use to explain PGMs  
- potentially allow students to train using Blue Bear.   
- send us the paper that has a list of “locally quadratic activation functions”, and the paper that says that swish is the best of these.