Y4 Project meeting 6 record

Date and time: 04/11/19, 14:00-15:50

Attended by: Guy, Harry and Hin

Discussed:

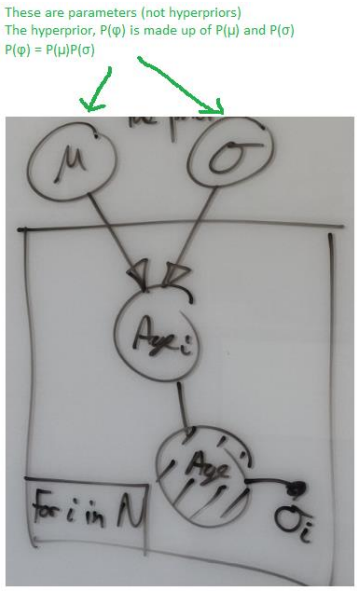
1. Project proposals marked by Guy, but more marking is required until we can get the full feedback

2. "The Prior" Terminology:

- "The Prior" = knowledge we have before we look at the data

The Prior, is split into:

- "The Hyperprior" = parameters that deal with the population level i.e. mean and standard deviation. Additionally information that encompasses our understanding of the population e.g. the age of an open cluster is limited by the age of the universe, they cannot be older than 13.8Gyrs. More specifically if we look at M67, from previous studies we know that its age cannot be less than 2 Gyrs or greater than 6Gyrs.
- "prior" = parameters that deal with the individual objects of the population
- θ = info on individual objects within population, i.e. $\theta = \{\theta_1, \theta_2, \ldots \theta_N\}$, where θ describes the information we have about each star for a population of N stars
- ϕ = info about the overall population



3. We should write a mini report from the preliminary data.

4. Discussion about setting up neural nets:

- In the past people have done 512 nodes per layer, with 10 hidden layers.
- Regularization = start with high value, make it smaller until the output is good
- Learning rate = too high is bad, too low doesn't really matter (Nadam updates learning rate anyway)
- Validation = save 10% to 20% of the data for validation. Ensure this data is distributed randomly throughout to prevent the validation being a chunk from a specific section of the data (i.e. shuffle the data)
- Validation set = want to know that the model is generalising well, if the train error is much lower than the validation error it is not generalising well (i.e. over/under fitting)
- We can try solving overfitting by changing the regularization
- Batch sizes = for data with noise, multiple batch sizes can act to average the noise so it converges

to the true value. As our data won't be noisy we should have 1 batch.
- Dropout = randomly turning off weights in each layer (potentially only useful for noisy data, see the week 6 notes for more info)
5. Discussion of a Gaussian process:
- Can be used to predict what happens between points and also what happens outside the trained region (see the week 6 notes for more info)
6. Discussion on HBM problem 3
7. Linear activation function is the logic choice for neural net output layers.

To do:
Students:
- Check we're happy with Guy's annual review comments
- Write mini report
- Mess around with NNs e.g. regularization, learning rate, beta(s), batch sizes (mainly: try 1 batch), dropout (but not necessary)
- We should set a target for next week
- If we're confident we have trained enough neural networks; tell Guy and he will share a grid with us (of some variety).

Tutor: reminders
- Check Tander Lee's grid
- Send "Model Sunlike Star" notebook

Next meeting: 11/11/19 (Wed) 14:00 in Guy's office

Recorded by: Harry