Y4 Project meeting 5 record

Date and time: 30/10/2019 (Wed) 1:30-2:45 pm

Attended by: Guy, Harry and Hin

Discussed:

1. The theories behind neural networks:
    a. Weights, biases, activation functions (ReLU, tanh, sigmoid), loss functions, learning rates, and how the NN optimizes weights from loss function (differentiate w.r.t that weight and evaluate its gradient)
2. Demonstrated the usefulness in learning and visualizing NNs through TensorFlow's playground function: A very user-friendly tool that immediately shows what the NN is doing at each step with coloured squares. Guy said it is the single best piece of software ever developed!
3. Demonstrated an example of using NN to approximate a simple MESA grid:
    a. Input = Mass, Age
    b. Output = Teff, luminosity
    c. Recreated a bunch of stellar track alongside MESA's original tracks, described the problem of smoothness and how "regularize" helps deal with that.
    d. We will want to build a simple NN that outputs spectroscopic data first for simplicity sake, then once we have a working method, move onto outputting asteroseismic values.
    e. It is better to intput logs of values instead of the values themselves as NNs like numbers with small dynamic ranges. (e.g. do this on age, mass, radius)
    f. Sometimes it is favourable to input powers/logs of input parameters to give the NN a "head start" if prior research show the correlation between that and a output value (e.g. mainsequence lifetime relates to M^(-3.2))
    g. A leaky ReLU can be a good loss function for our purpose
4. Running NNs: students write up the code and send it to Guy through Github to run on the GPU in the office
5. TensorFlow has its own HMC sampler called edward2, but it lacks a NUTS sampler, making it hard to use (requires manual tuning). For now, we are sticking with the process of exporting all weights and biases from the trained NN into pymc3 and do HBM using that.
6. Training NN and finding the appropriate architecture: test out training methods and different NN architectures on a single MESA track first, then once determined, expand sample to the whole grid (a portion of the grid)
7. Preliminary results on M67 ages: estimated age spread is many orders higher than what is observed in open clusters (1Gyr vs few Myr), showing that D. Stello 2016 have underestimated their errors on ages. The small age spread in open clusters is seen from the very clearly defined (almost iso-chromic) spread in HR diagram.

To do:

Students:

1. HBM: continue getting familiar with pymc3 despite it not having a good starting point

2. NNs: do some test training on a simple function (eg. Sine wave of increasing frequency), learn how to do the keras code stuff. Might not be a good idea to play around with adding noise as MESA tracks has no observational noise.
3. Mess around with the results on M67, that can very well be an early part of the final report.

Supervisor:

1. Upload the NN on MESA grid example to repo

Next meeting: 3/11/2019 (Mon) 2:00 pm

Recorded by: Hin