Group 16G25 Lab Project
Internal Meeting Notes

Attendance: Scott Hazelhurst, Paul Cresswell, James Allingham
Minutes noted by: James Allingham

4 October 2016
15:30 - 16:30

1 TensorFlow on the School's GPU machine

- Scott brought up a problem: the school's GPU machine runs scientific linux 6.8.
- This is an issue because it uses kernel 2 which is not compatible with TensorFlow.
- A number of options will be investigated: using the CHPC, using AWS, and installing a newer version of scientific linux.

2 Real world data set

- James enquired about the real data set mentioned by Scott previously.
- Scott said that it would be available when needed.
- It is in the PLINK data format.
- It contains 140 000 SNPs (of which about 130 000 are important).

3 Accuracy and efficiency comparisons

- Paul enquired about what sort of measurements should be made for efficiency and accuracy comparisons between the project and existing tools such as MDR and BEAM.
- Scott responded, saying that for accuracy smaller number of SNPs would be acceptable if the argument is made that performance does not degraded as the number of SNPs increases. For these data set sizes we can compare directly with MDR and BEAM. On the other hand, for efficiency we should use bigger data sets that MDR cannot handle.

4 Project deliverables

- James enquired about the requirements for test coverage and code commenting of the source code as well as ease of use of the software produced.
- Scott suggested that although these are important they are secondary concerns - the main aim is to produce and efficient and accurate method to detect epistasis.
- Paul enquired about the state of the Git repo and making frequent commits and releases.
- Scott said that these are also important but will not affect project evaluation directly. Scott reiterated that professionalism was important throughout the project. We should strive for neat, commented code and a Git repo with a good README and frequent commits.