

REPORT FOR SUMMER BURSARY 2017

|  |  |
| --- | --- |
| Name: | Donald Hobson |
| Course: | MMath |
| Current year: | 2 |
| Short title of the activity: | Machine learning on space data |
| Dates of the activity: | 17th July to 8th Sept 2017 |
| Amount of the Summer Bursary awarded: | £2,072 |

|  |
| --- |
| **Please write a short report about the activity undertaken in Summer 2017.**  Try to include a brief description of the activity, the key benefits for your personal and academic development and the way in which the Summer Bursary funding has assisted you. Please feel free to attach photos, if appropriate.  *For examples of other reports, please see the links at the bottom of the* [*http://www.joh.cam.ac.uk/travel-grants*](http://www.joh.cam.ac.uk/travel-grants) *webpage.* |
| My project was about machine learning on space data, not because there was a specific problem to solve but because exciting work was being done in that area.  This gave me the freedom to identify a problem and work on its solution. After considering several potential problems and finding the solution already existed, I decided to program neural networks for pattern detection. It was initially suggested that I should use deep learning, but making the networks deeper causes them to take longer to train and only improves the results when you have lots of data with complex patterns in it. I had toy (i.e. sample) data and wanted to demonstrate, within sensible time limits, that my algorithms did in fact work, so I used shallow networks.  All my networks assumed the data followed a certain structure. They took in two lots of data, call them A and B. Each dataset consisted of n values, so A[1], B[1] up to A[n], B[n]. It was assumed that A and B represented different properties of a set of objects or locations. So A[i] and B[i] would represent different measurements of the same thing. Note that a single A[i] or B[i] may still be a list of numbers.    My first version was the simplest. It took in a subset of A and learned to predict the respective values for B. It was then tested on the remaining values of A. If the network could guess B more accurately than always guessing the average of B, then there must be a correlation between the two. The strength of this correlation could be deduced from the accuracy.    Subsequent versions implemented more features on this model. In version 2 the network has to guess its uncertainty about the value of B. This allows you to measure how well B can be predicted from A, i.e. how well a particular object or location is correlated. It finds a stronger correlation when B can sometimes be accurately predicted. I then allowed the network to guess several possibilities for B. This would allow you to identify patterns where a certain value of A meant that the corresponding B was either this or that.    Finally I tried another approach. This network was given an A[i] and a B[j]. It had to guess whether or not i=j. This allowed it to detect patterns in which neither A nor B was a simple function of the other, but they still depend on each other in some manner. Suppose you have a single dataset C and a data point D (e.g. set of images and then another image), and you want to figure out if D would look out of place in C. Then you can divide C into 2 fragments, train the network to tell if the fragments are related, and run the network on the fragments of D. If the fragments of D are judged to fit together by the network, and both fragments themselves look like fragments from C, then D looks like its from C. Note that this algorithm breaks the problem down into several smaller problems, so must be used recursively. A histogram is used to bottom out at single values.  Over the course of this project I learned about machine learning, coming up with a series of increasingly sophisticated networks and algorithms. I got practice turning a rough overview into an exact algorithm. I also found how hard it was to find a promising research idea that no one has pursued before and can be completed by a first year student in a few weeks. |