# Deep Learning: Lab 3

Johnny Joyce (jvjj1u19@soton.ac.uk)

1 Exploring optimisation of analytic functions

The 2D Rastrigin function was implemented and a plot of this can be seen in Figure 1(a) with $A = 1$. When starting at (5, 5), the various optimisers listed arrived at the following (x, y)-coordinates after 100 iterations:

- SGD: (2.822, 2.822)
- Momentum: (-0.947, -0.947)
- Adagrad: (4.842, 4.842)
- Adam: (3.939, 3.939)

A plot of the error (value of Rastrigin function) across each iteration for all of these optimisers can be seen in Figure 1(b). We can see that under these parameters, stochastic gradient descent with momentum=0.9 was able to reduce the error the most and arrive the closest to the global minimum at (0,0). SGD was second-best, followed by Adam, then Adagrad.
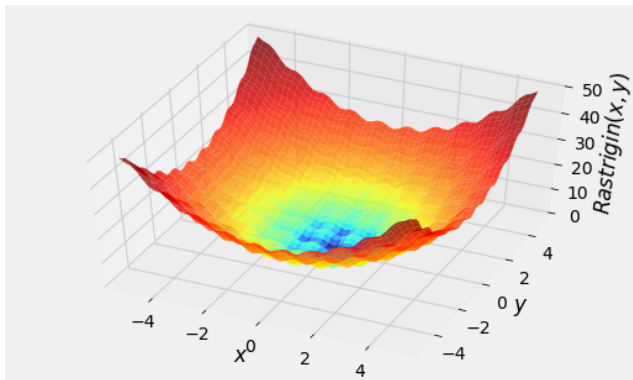
(a) Plot of 2D Rastrigin function ($A = 1$)
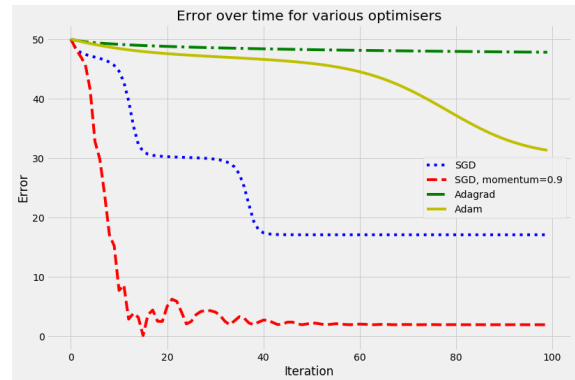
(b) Plot of error over time for various optimisers

Figure 1

SGD (with and without momentum) may have outperformed Adam since Adam's parameters had not been adjusted to suit the problem. SGD with momentum may have outperformed SGD without momentum due to the rugged nature of the Rastrigin function, causing SGD without momentum to spend much more time in the many local optima.

2 Optimisation of a SVM on real data

Using the given data and parameters, a soft-margin SVM was trained using both SGD and Adam. The accuracy values on the validation data were as follows:

- SGD: 92%
- Adam: 88%

For 1000 iterations, the training and validation data was shuffled randomly before the validation accuracy was recorded for completely random weights. This random initialisation produced a mean accuracy of 50.2%, so we can see that our optimisers produced better accuracy than the expected accuracy from using random weights.

Both SGD and Adam produced similar validation accuracy values.