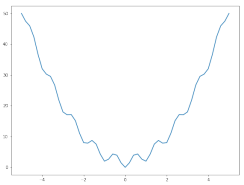
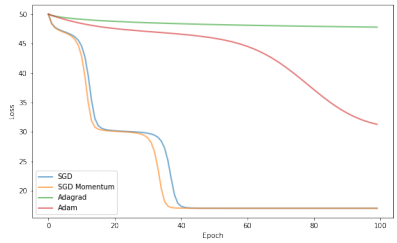
1 Exploring optimisation of analytic functions

1.1 Rastrigin

With $A = 1.0$ we obtain a curve full of 'bumps' where a gradient descent style optimiser could get stuck. Hence, the purpose of this experiment is to find the optimiser which can arrive closest to, if not, the global optimum itself. As the optimiser descends down the slope, it minimises the loss function given to it, which in this case is the Rastrigin function. From the loss plot, we find that SGD + Momentum arrives at the global minimum the quickest, followed by SGD. Adam and Adagrad both fails to reach the global minimum within 100 iterations.



(a) 2D Rastrigin showing many local minima and one global minimum.



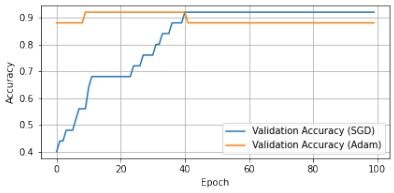
(b) Comparison of different optimisers.

Figure 1: Optimising the Rastrigin function.

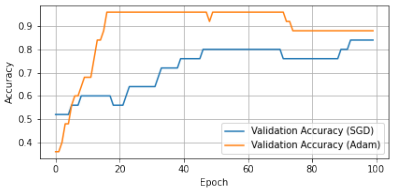
2 Optimisation of a SVM on real data

2.1 Iris SVM

Over many random initialisations, we found that Adam found the minimum quicker that SGD. This was unexpected because in the Rastrigin scenario, Adam failed to arrive at the global minimum within 100 terations whereas SGD found it within 40 iterations. Aside from faster performance, Adam seemed more likely to arrive at the global minimum as it's final validation accuracy was more stable compared to SGD.



(a) Adam = 88% SGD = 92%.



(b) Adam = 88% SGD = 84%.

Figure 2: Final validation accuracies of Adam and SGD on Iris dataset.