Lab 3-Optimise it!
By: Abhash Shrestha (35464984) as14n23@soton.ac.uk

1.1 Rastrigin

When A= 0.5 the best optimizer is SGD with momentum as seen in figure 2.

When A = 10Adam and Adagrad have similar performance, which can be seen in figure 1.
It is an adaptive learning algorithm that adapts the learning rate of each parameter based on the previous gradients.

When A = 10 we get extremely bumpy plots as mentioned in the question. Also, SGD has the highest final loss of 56.51365661621094 when A=10. The loss value indicates that it doesn't perform well in this case.

Since the learning rate is 0.01, it is unlikely that the high loss was due to fast learning rate decay as it is adapted based on the accumulated gradients. As The Rastrigin function has a highly non-convex and multimodal nature, with many local minima and steep gradients. The SGD could have got stuck, converged into one of the local minima, causing the high error.

However, this multimodal nature seem to be no problem to the SGD+momentum as the momentum helps the optimizer escape the local minima. When it gets to a local minima the momentum allows it to keep moving and explore the solution space and ultimately head in the direction of the better/true minima.

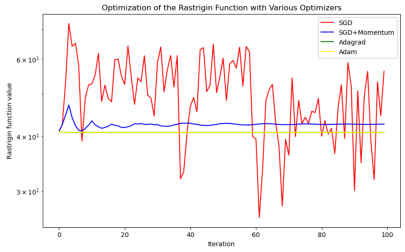

fig1: Optimisers per iteration when A=10

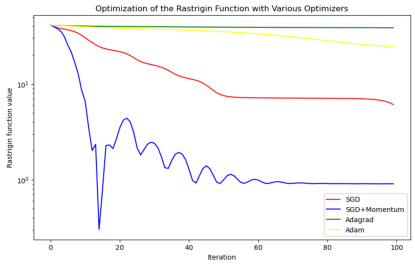

fig1: Optimisers per iteration when A=0.5

2.1 Iris SVM

The following were the results

Optimizer: SGD, Learning Rate: 0.01
Expected Validation Accuracy: 0.9200,
Variance: 0.033067

Optimizer: Adam, Learning Rate: 0.01 Expected Validation Accuracy: 0.8800, Variance: 0.025067

Optimizer: SGD, Learning Rate: 0.001
Expected Validation Accuracy: 0.6400, Variance: 0.044267

Optimizer: Adam, Learning Rate: 0.001
Expected Validation Accuracy: 0.9600, Variance: 0.044267

Optimizer: SGD, Learning Rate: 0.0001
Expected Validation Accuracy: 0.8000, Variance: 0.018667

Optimizer: Adam, Learning Rate: 0.0001
Expected Validation Accuracy: 0.6000, Variance: 0.058667

From the provided results, we observe that the expected validation accuracy for models ranges from 60-96%. This suggests that the models might not performing well on the validation data depending on the learning rates. The variance in the validation accuracy across epochs is relatively small, indicating consistency in the performance of the models during training.Possible reasons for the low performance.

SGD with lr 0.001 which is quite low has an accuracy of 64%. A smaller Lr can lead to a slower convergence and the model can be more sensitive to the initial setting and parameters.
Adam with lr 0.0001 is the lowest accuracy of 60%. Such a small learning rate maybe too small for Adam to update parameters meaningfully, resulting in a very slow convergence and the model not reaching irs optimal performance.

With such a small learning rate, the updates to the parameters during training may be too gradual, resulting in slow convergence. This slow convergence could lead to the model not reaching its optimal performance within the given number of epochs.

To verify this, I used 10000 epochs and the Adam achieved an accuracy of 88% with an lr of 0.0001. This increases the lower accuracy values of 64%, 60%, 80% etc but decreases the high accuracies of 96% that were gained previously. This can be attributed to overfitting, and perhaps a regularisation parameter would be effective.

LR	0.01	0.001	0.0001
SGD	Acc: 0.9200 Var: 0.03306	Acc: 0.6400 Var: 0.04426	Acc: 0.8000 Var: 0.01866
ADAM	Acc: 0.8800 Var: 0.02506	Acc: 0.9600 Var: 0.04426	Acc: 0.6000 Var: 0.05866