Yang Liu
3323643

Lab Exercise 3

COMP6248
yl58n21@soton.ac.uk

1 Rastrigin

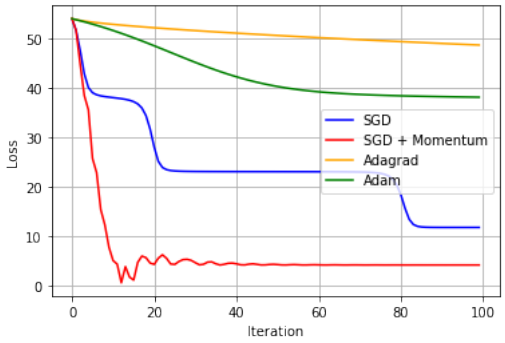

Figure 1: Loss value

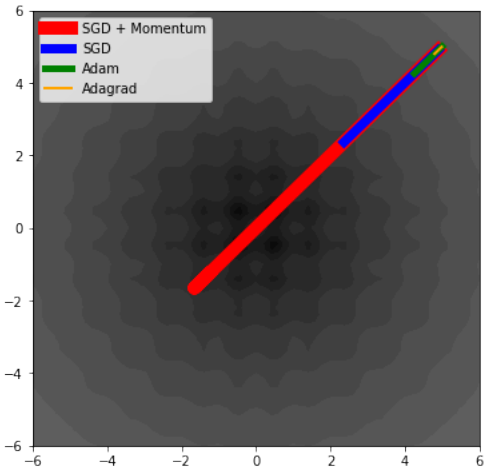

Figure 2: Path

SGD with Momentum performs best among these optimisers, because its loss value is the smallest and its path to global minimum is the shortest.

2 Iris SVM

	SGD optimiser Validation Accuracy
1	0.86
2	0.92
3	0.96
...	...
100	0.90
Mean	**0.9148**

	Adam optimiser Validation Accuracy
1	0.88
2	0.92
3	0.92
...	...
100	0.88
Mean	**0.8920**

After training each model 100 times respectively using different random initialisations, I find that SVM with SGD optimiser performs(validation accuracy) better than with Adam optimiser. This is counter-intuitive because Adam optimiser should outperform SGD since Adam is an improved version of SGD that uses first and second moments. This is probably because SGD is quite "unstable" compared with Adam. SGD updates the parameters once a data is learned, which adds the noise of that data to the learning process. This is always considered to be bad, but it also gives power to the algorithm to be able to jump out of local minimum and find a better minimum. Perhaps that's why SGD sometimes performs better than Adam.