

Lab 3 Exercise – Optimise it!

1. Exploring optimisation of analytic functions

1.1 Rastrigin

As seen below, when $A = 10$, the Rastrigin function's surface is really bumpy. The loss curves of four optimisers used to optimise this function are displayed below as well. SGD is the one whose curve is of highest fluctuation but it turns out the lowest loss—around 36. And others are not doing well with this task: they are all stuck in local minima with loss about 50. So their learning rate need to be tuned up.

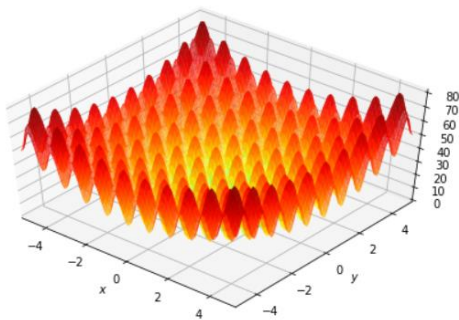


Figure 1: The Rastrigin surface

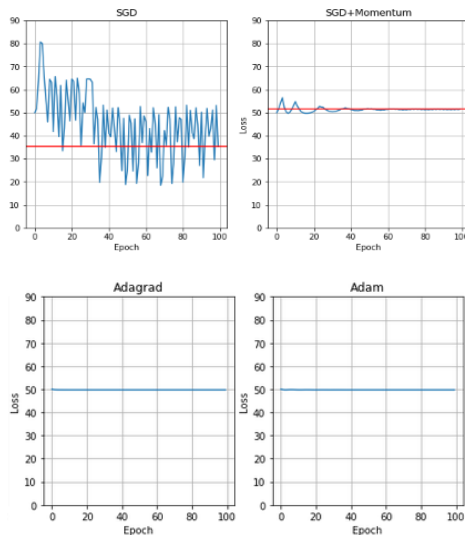


Figure 2: optimisers' loss curve

2. Optimisation of an SVM on real data

2.1 Iris SVM

I ran the Iris SVM model 50 times with SGD and Adam optimiser. First, SGD had 83.92% and 86.64% on training set and validation set respectively. And Adam yielded 92.37% and 88.16% correspondingly. The weird thing is SGD's validation accuracy is about 3% higher than its training accuracy. That may result from the imbalance between the sample numbers on the training set and the validation set.

In addition, Adam is way more robust than SGD for random initialisations: its train-and-val accuracy has a smaller variance

(0.0051, 0.0046) *against* (0.0125, 0.0100)

and a higher average value

(0.9237, 0.8816) *against* (0.8392, 0.8664)

as the lower figure shown.

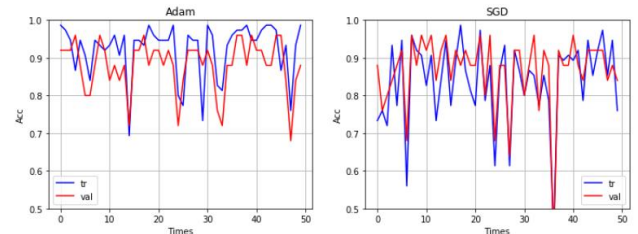


Figure 3: Robustness trial on Adam and SGD

