

# COMP6248 Lab 3 Exercise – Optimise it!

Wei Chien Teoh (Eugene)  
wct1c16@soton.ac.uk

29 April 2021

## Introduction

The results are seeded using `torch.manual_seed(0)` to provide reproducible results.

## 1 Exploring optimisation of analytic functions

### 1.1 Rastrigin

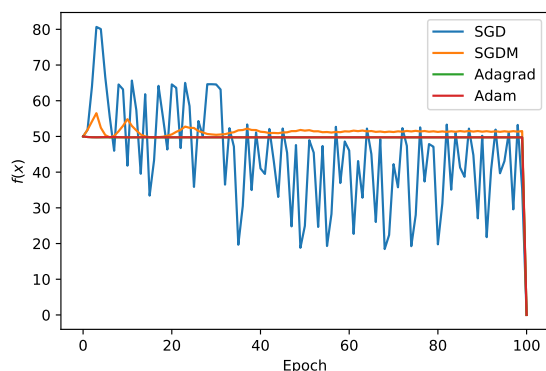


Figure 1: Learning curve for different optimisers.

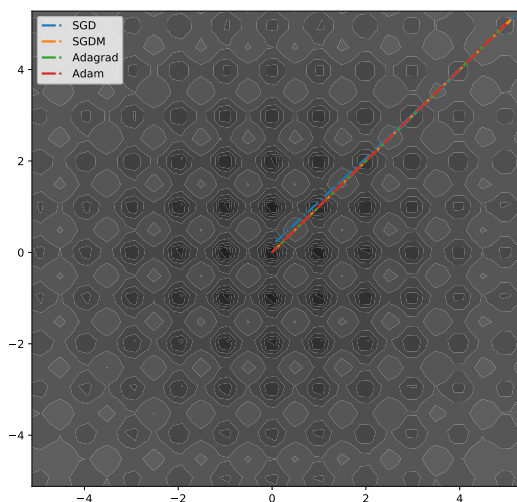


Figure 2: Contour plot and trajectory of different optimisers.

Fig. 2 shows all optimisers converge to the global minima at point  $(0, 0)$ . All optimisers converge at around the same epoch, as illustrated in Fig. 1. However, Adagrad and Adam, with identical performance, provide the most stability of between epochs. Hence, both are the best candidates for optimising the Rastrigin function.

## 2 Optimisation of a SVM on real data

### 2.1 Iris SVM

Table 1 shows the mean and variance of accuracy after the SVM is repeatedly trained for 50 times with random initialisation of weights and biases. Both optimisers provide decent and close accuracies, with SGD slightly leading for both mean and variance. This is a surprising finding since the Adam optimiser was developed to improve upon SGD on various aspects such as speed of convergence. A potential explanation for this phenomenon is that adaptive optimisers such as Adam is observed to have worse generalisation than SGD despite it's better training performance.<sup>1</sup>

Table 1: Accuracy for 50 repeated experiments.

Optimiser	Mean	Variance
SGD	0.9008	0.000399
Adam	0.8912	0.000995

<sup>1</sup>Ashia C. Wilson et al. *The Marginal Value of Adaptive Gradient Methods in Machine Learning*. May 21, 2018. arXiv: 1705.08292 [cs, stat]. URL: <http://arxiv.org/abs/1705.08292> (visited on 04/24/2021).