Report
Assignment - 1

Siddik Ayyappa
2020101089
CSE, UG2k20

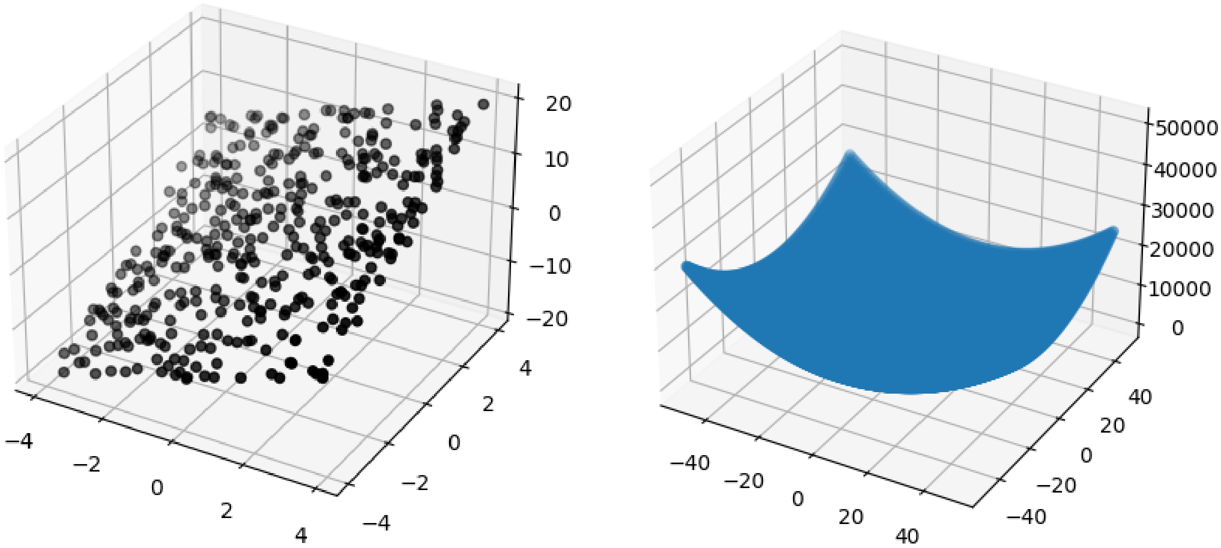
Deep Learning - Theory and Practices

INTERNATIONAL INSTITUTE OF
INFORMATION TECHNOLOGY
HYDERABAD

September 17, 2022

Question - 1: Gradient Descent Method

The problem is to demonstrate the learning of a linear regression function, with the help of gradient descent method. The error plot is on the right and the scatter plot of the points is on the left.



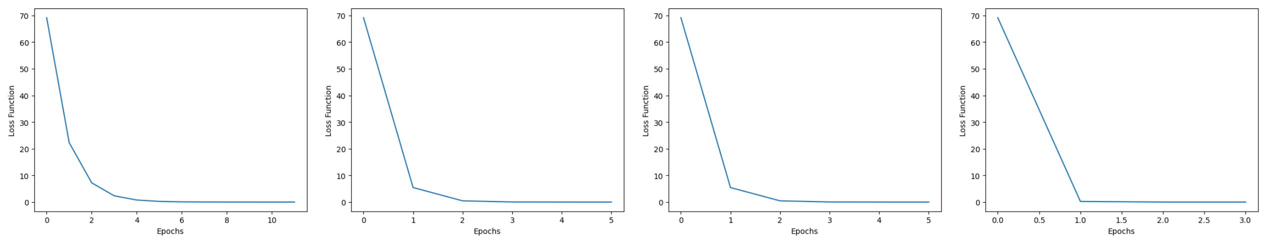
Talking about the number of epochs different learning rates took, the optimal learning rate which was around 0.0902, has the least number of epochs.

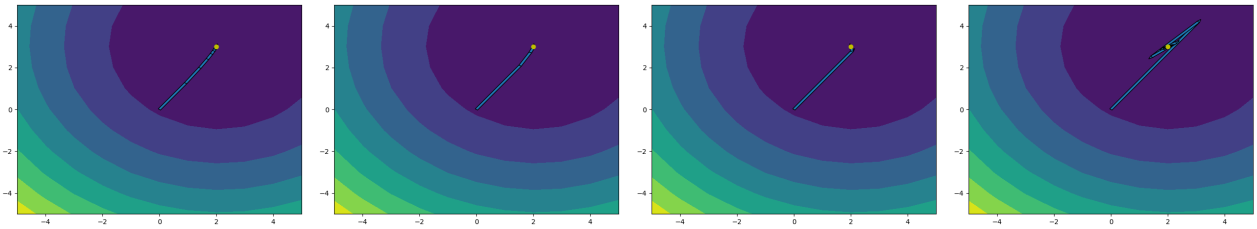
$$(0.9)\eta_{opt} \longrightarrow [\ 11 \text{ epochs}\ ]$$
$$(1.5)\eta_{opt}/2 \longrightarrow [\ 5 \text{ epochs}\ ]$$
$$\eta_{opt} \longrightarrow [\ 3 \text{ epochs}\ ]$$
$$(1.2)\eta_{opt} \longrightarrow [\ 8 \text{ epochs}\ ]$$

The method used to calculate η_{opt}: Since, the loss is quadratic the loss can be writen as ...

$$E = w^T A w + w^T B + c$$

Where the value of η_{opt} is given by $\frac{1}{\lambda_{max}}$, where λ is the eigenvalue of the matrix **A**. A can be obtained by taking the Hessian of the loss function, with respect to w. Obtaining the Eigenvalues involves solving a quadratic equation, whose coefficients depend upon the data. This can be obtained with the help of a little Algebraic Manipulation. The Error plot and Contour Movement Plot for $(0.9)\eta_{opt}, (1.5)\eta_{opt}/2, \eta_{opt}, (1.2)\eta_{opt}$ respectively (Left to right) is given below.





Question - 2: Gradient Descent and Variants - 1

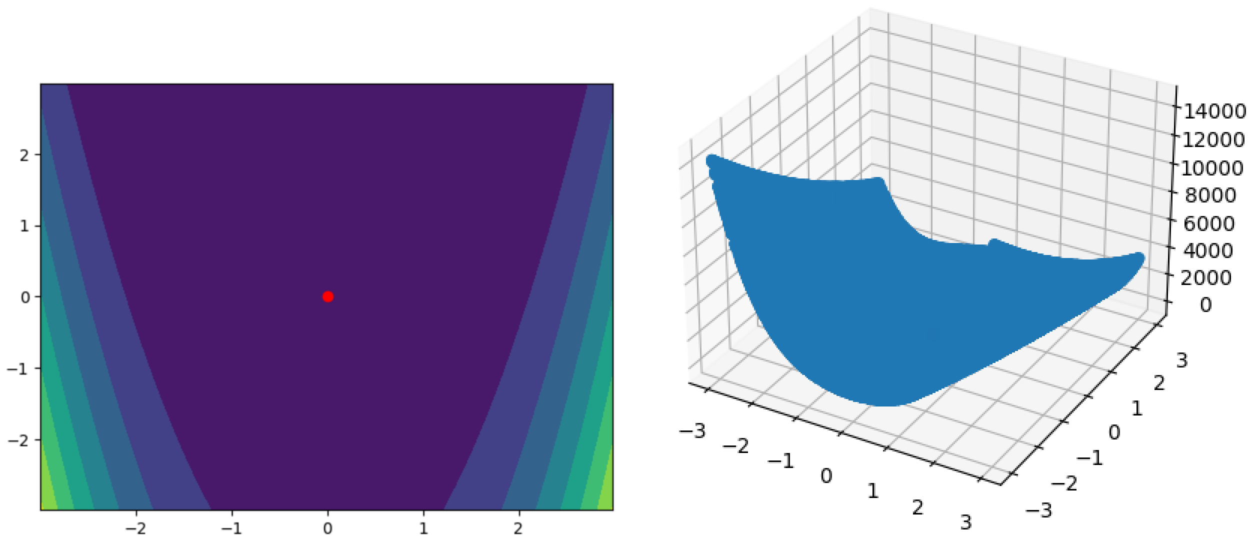
The problem is to minimise the Rosenbrock Function, using the following techniques.

a) Stochastic Gradient Descent

b) Polyak's Momentum Update

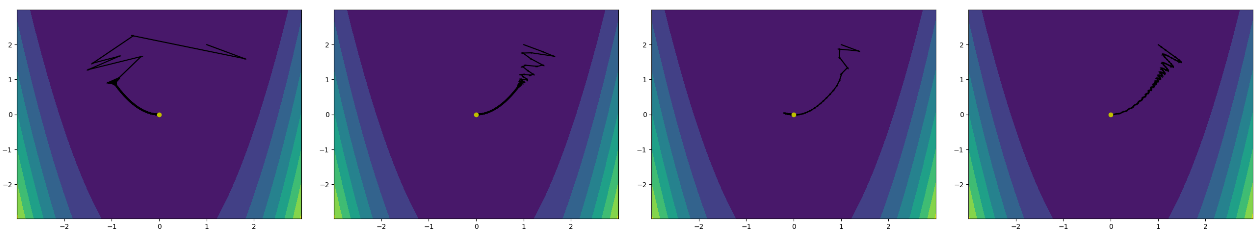
c) Nesterov's Accelearated Gradient Descent

d) Adam Optimiser



Plots of the function (Right - Scatter, Left - Contour)

The convergence movements of various methods are as follows, (left to right) for SGD, Momentum, NAG, and Adam. (1612, 534, 142, 119 epochs for convergence respectively)



Question - 3: Gradient Descent and Variants - 2

The problem is to minimise the Function given below, using the following techniques.

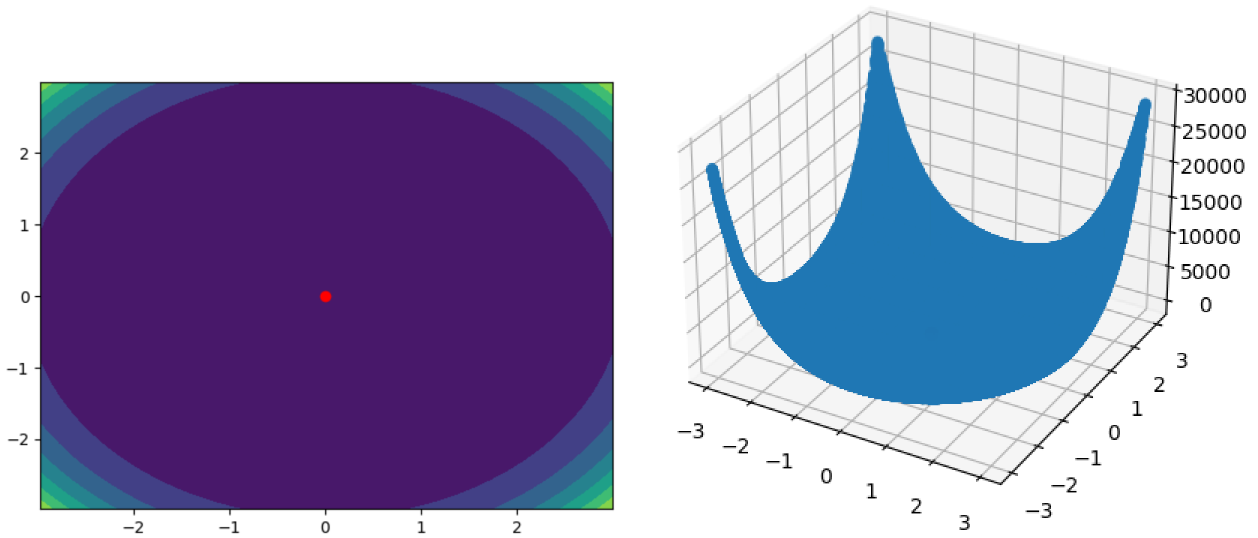
$$f(x, y) = \frac{50}{9}(x^2 + y^2)^3 - \frac{209}{18}(x^2 + y^2)^2 + \frac{59}{9}(x^2 + y^2)$$

a) Stochastic Gradient Descent

b) Polyak's Momentum Update

c) Nesterov's Accelearated Gradient Descent

d) Adam Optimiser



Plots of the function (Right - Scatter, Left - Contour)

The convergence movements of various methods are as follows, (left to right) for SGD, Momentum, NAG, and Adam. (434, 69, 71, 52 epochs for convergence respectively)