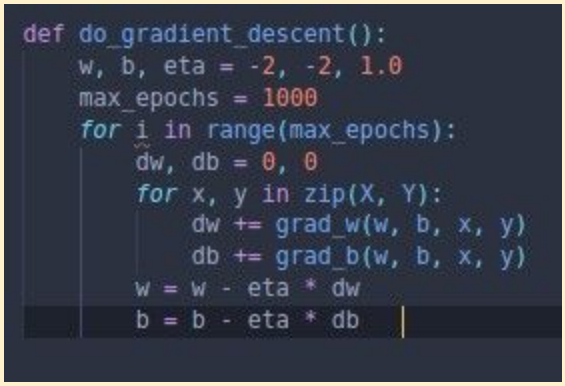
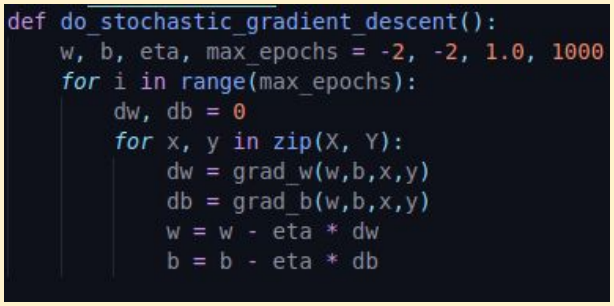
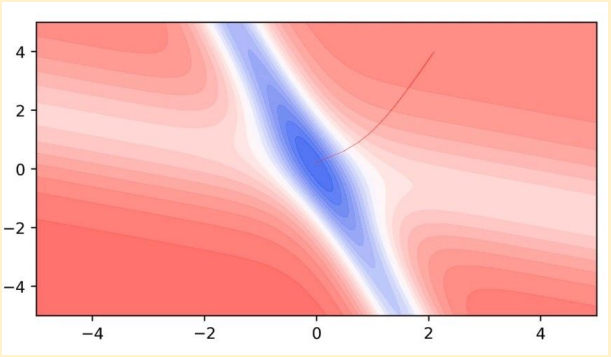
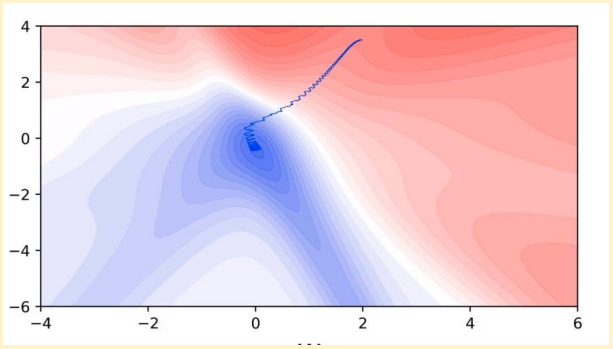
PadhAI: Variants of Gradient Descent

One Fourth Labs

Running stochastic gradient descent

Can we make stochastic updates

1. Let's do a side by side comparison of batch GD and stochastic GD

Batch GD	Stochastic GD
	
	

2. Some of the advantages of Stochastic GD are
   a. Quicker updates
   b. Many updates in one pass of the data
3. Some of the disadvantages of Stochastic GD are
   a. Approximate(stochastic) gradient
   b. Almost like tossing a coin once and computing P(heads)
4. From the Gradient descent visualization, we can see that it oscillates during movement. However, this oscillation is different from Momentum GD or NAG.
5. In stochastic GD, the oscillations are due to redirection after every point, as every point behaves as an individual greedy entity influencing w & b, thus leading to fluctuations right from the start.
6. In MGD or NAG, the oscillations appear the value approaches the minima as a result of overshooting the intended destination.