

PadhAI: Variants of Gradient Descent

One Fourth Labs

A disadvantage of momentum based gradient descent

Let us make a few observations and ask some questions.

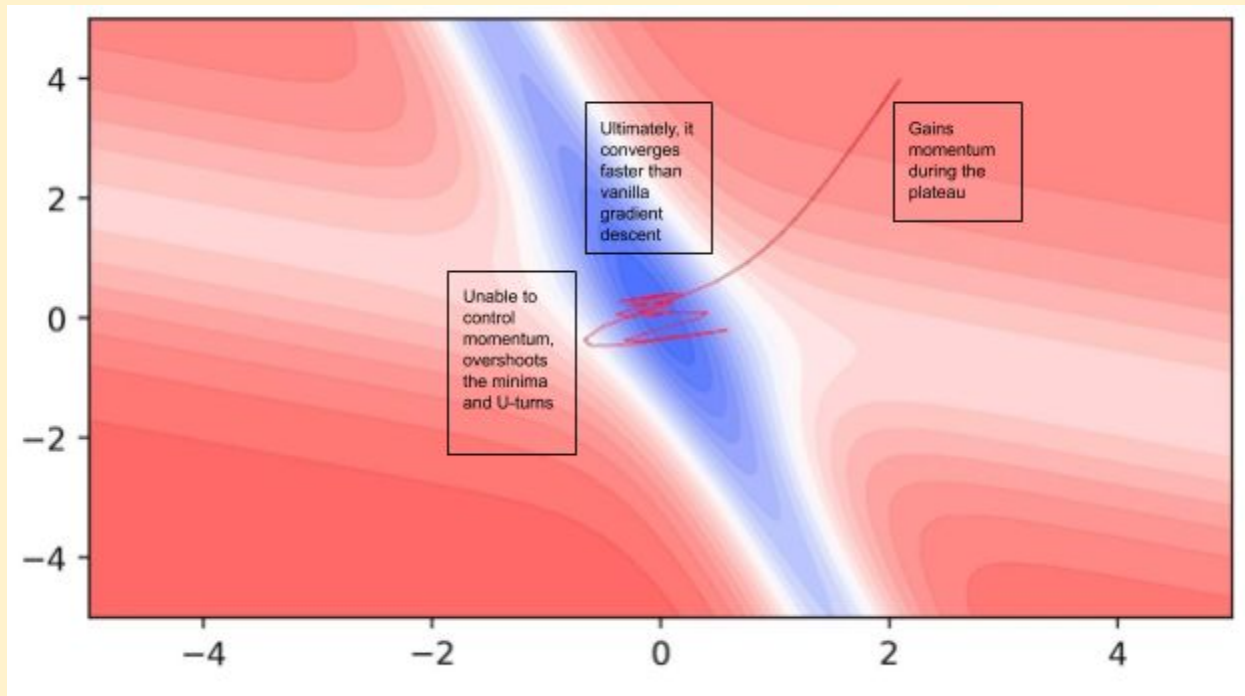
1. Observations

- Even in the regions having gentle slopes, momentum based gradient descent is able to take large steps because the momentum carries it along

2. Questions

- Is moving fast always good?
- Would there be a situation where momentum would cause us to run past our goal?

3. Let us look at an implementation of Momentum based GD



4. A few points to note

- Momentum based gradient descent oscillates in and out of the minima valley (u-turns)
- Despite these u-turns it still converges faster than vanilla gradient descent

5. Now, we will look at reducing the oscillations in Momentum based GD