[PadhAI: Variants of Gradient Descent]

One Fourth Labs

Epochs and Steps

What is an epoch and what is a step?

1. Let us go over the definitions of an epoch and a step
   a. 1 epoch = one pass over the entire data
   b. 1 step = one update of the parameters
   c. N = number of data points
   d. B = mini-batch size
2. Let's analyse the algorithms using epochs and steps

Algorithm	Number of steps in one epoch
Batch Gradient Descent	1
Stochastic Gradient Descent	N
Mini-Batch Gradient Descent	N/B

3. Let's look at stochastic version of NAG and Momentum based GD

Stochastic Momentum GD	Stochastic NAG

Stochastic Momentum GD:

```python
def do_stochastic_momentum_gradient_descent():
    w, b, eta, max_epochs = -2, -2, 1.0, 1000
    v_w, v_b = 0.0, 0.0
    gamma = 0.7
    for i in range(max_epochs):
        dw, db = 0, 0
        for x, y in zip(X, Y):
            dw += grad_w(w, b, x, y)
            db += grad_b(w, b, x, y)
            v_w = gamma*v_w + eta*dw
            v_b = gamma*v_b + eta*db

            w = w - v_w
            b = b - v_b
```

Stochastic NAG:

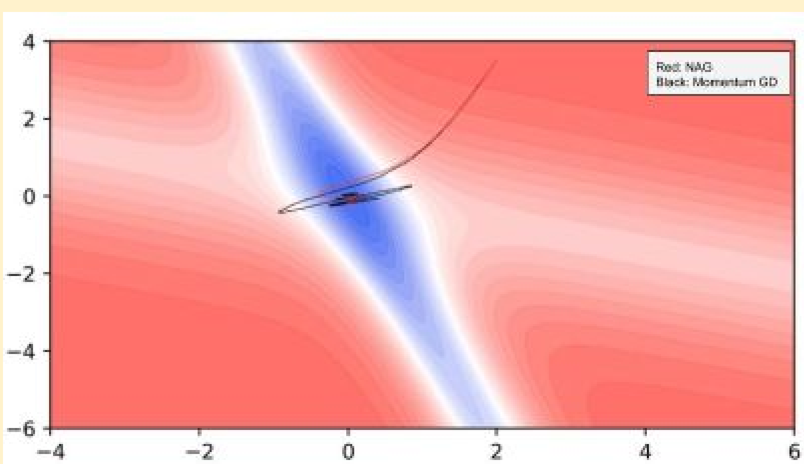
```python
def do_stochastic_nag_gradient_descent():
    w, b, eta, max_epochs = -2, -2, 1.0, 1000
    v_w, v_b = 0, 0
    gamma = 0.9
    for i in range(max_epochs):
        dw, db = 0, 0

        #Compute the lookahead value
        w = w - gamma*v_w
        b = b - gamma*v_b

        for x, y in zip(X, Y):
            #Compute the derivatves using the lookahead value
            dw += grad_w(w, b, x, y)
            db += grad_b(w, b, x, y)

        #Now move further in the direction of that gradient
            w = w - eta*dw
            b = b - eta*db

        #Now update the history
            v_w = gamma * v_w + eta * dw
            v_b = gamma * v_b + eta * db
```



4. Since there is a history component, NAG and Momentum GD have slightly smoother oscillations.