PadhAI: Variants of Gradient Descent

One Fourth Labs

The idea of stochastic and mini-batch gradient descent

How many updates are we making?

1. Let us consider vanilla gradient descent

```
X = [0.5, 2.5]
Y = [0.2, 0.9]

def do_gradient_descent():
    w, b, eta = -2, -2, 1.0
    max_epochs = 1000
    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)
        w = w - eta * dw   # Updates to w are made only after all data-points are covered
        b = b - eta * db   # Updates to b are made only after all data-points are covered
```

2. From the above image, we can see that we make one update(w,b) for one pass/epoch over the data
3. It can be exemplified as follows
    a. Consider a training set with 1 million data points
    b. With Gradient Descent, we calculate the derivatives for each of these points
    c. Once we're done, we update the parameters
    d. Thus, we pass over all 1 million points to make a single update to w & b
    e. It can also be called **batch gradient descent**, as the entire dataset is used as a single batch
4. However, we can choose to make an approximation based on looking at a smaller portion(batch) of the data points instead of analysing the whole dataset each time.
5. This is called **mini-batch gradient descent** and can be described as follows
    a. Consider a training set of 1 million data points
    b. Select a batch size of 100 data points
    c. What this means is that every batch, the algorithm calculates all of the 100 derivatives and updates the parameters
    d. Thus, passing over all 1 million data points results in 10000 updates to w & b.
6. **Stochastic gradient descent** is when the batch size is 1, i.e. an update to the parameters after each single data point
7. One key thing to note is that both stochastic and mini-batch gradient descent are approximations of the true derivative obtained by batch gradient descent.
8. However it is advantageous as it allows is to make updates faster and achieve quicker progress.