Mini-Batch Gradient Descent

Wantee Wang

2015-03-10 21:44:23 +0800

In Mini-Batch Learning, we update the parameter $\mathbf{w}$ every b examples. There are two ways to do the update.

First, using the summation of all examples in the mini-batch, i.e.,

$$\Delta\mathbf{w} = -\alpha_1 \sum_{i=l}^{l+b-1} \nabla E^{(i)} \tag{1}$$

Second, using the average of all examples in the mini-batch, i.e.,

$$\Delta\mathbf{w} = -\alpha_2 \frac{1}{b} \sum_{i=l}^{l+b-1} \nabla E^{(i)} \tag{2}$$

From (1) and (2), we can see that by simply scaling the learning rate, i.e. $\alpha_1 = \frac{1}{b}\alpha_2$, these two method can be equivalent.

http://wantee.github.io//2015/03/10/mini-batch-gradient-descent/