Bias Corrections In Adam

Pavlos Protopapas

Let us examine the equation for momentum

$$v^{(t)} = \alpha v^{(t-1)} + (1-\alpha)g^{(t)}$$

But at t=1 there is no $v^{(0)}$, since this is the “trend” from before.

Here we show how to correct for $v^{(t)}$ such that the initial guess for $v^{(0)}$=0 does not lead to biasing.

Step 1: We show that the stationary value of v and g are the same. To do so, we examine their expectations.

Starting from the equation

$$v = \alpha v + (1 - \alpha) g$$

and take the expectation

$$E[v] = E[\alpha v + (1 - \alpha) g] = \alpha E[v] + (1 - \alpha) E[g]$$

$$\Rightarrow (1 - \alpha) E[v] = (1 - \alpha) E[g]$$

$$\boxed{E[v] = E[g]}$$

Step 2:

$$v^{(1)} = \alpha v^{(0)} + (1-\alpha) g^{(1)}$$

however, if $v^{(0)} = 0$

$$v^{(1)} = (1-\alpha) g^{(1)}$$

And this is a problem because we want the stationary values to be the same

Step 3:

Assume the following is true for t

$$E[v^{(t)}] = (1 - \alpha^t)E[g^{(t)}] \qquad (1)$$

We only need to show that this is true for t=1 and t+1. The t=1 is trivial (we showed it in the previous slide). We therefore examine the t+1 case:

$$v^{(t+1)} = \alpha v^{(t)} + (1 - \alpha)g^{(t+1)}$$
$$E[v^{(t+1)}] = \alpha E[v^{(t)}] + (1 - \alpha)E[g^{(t+1)}]$$

Substituting $E[v^{(t)}]$ from Eq.(1)

$$E[v^{(t+1)}] = \alpha(1 - \alpha^t)E[g^{(t)}] + (1 - \alpha)E[g^{(t+1)}]$$

$$E[v^{(t+1)}] = \alpha(1-\alpha^t)E[g^{(t)}] + (1-\alpha)E[g^{(t+1)}]$$

But stationarity implies that $E[g^{(t)}] = E[g^{(t+1)}]$

$$\Rightarrow E\left[v^{(t+1)}\right] = [\alpha(1-\alpha^t) + (1-\alpha)]E\left[g^{(t+1)}\right]$$

$$= [1-\alpha^{t+1}]E\left[g^{(t+1)}\right]$$

So, in order to make sure that $E[v] = E[g]$ for any t we will need to adjust $v^{(t)}$ as follows:

So

$$v^{(t)} = \frac{v^{(t)}}{1-\alpha^t}$$