## **Adam Optimization Algorithm:**

Adam作为一个泛用性极强的优化算法，集Gradient Descent with Momentum 和 RMSprop的优点为一体，其基本实现如下所示：

**Adam: Adaptive momentum estimation**

On interaction , compute on current mini batch, the learning rate is :

Initial with and

Adding **bias correction**:

Then, the position will be updated using the above two equations:

Usually, , which are the default values. is computing the mean of the derivatives, called the first momentum, and is used to compute exponentially weighted average of the squares, called the second momentum.