## **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。

**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 is like averaging the gradients of the last 10 iterations. Therefore, it smooths out the steps of gradient descent.