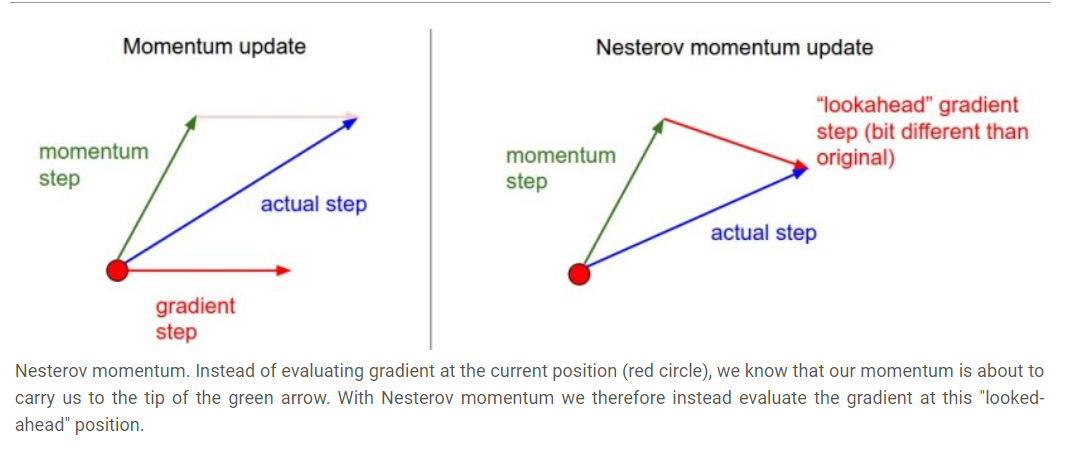
1. We learned optimizer such as SGD, ,Adagrad, RMSProp, Adadelta, and Adam. Research the other two state-of-the-art optimizers and explain their feature.

Sol:

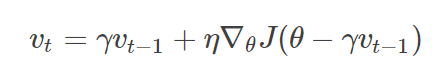
Nesterov accelerated gradient(NAG):

The following figure shows the difference between Momentum update and Nesterov momentum update.



From <https://blog.csdn.net/google19890102/article/details/69942970>

This method calculates the gradient nor based on our current parameters θ, but based on the approximate future position of our parameters θ.



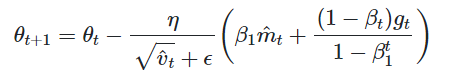


Compare to the normal momentum method, this method can prevent us from going too fast and results in increased responsiveness, which has significantly increased the performance of RNNs on a number of tasks.

From <https://arxiv.org/pdf/1609.04747.pdf>

Nesterov-accelerated Adaptive Moment Estimation (NADAM):

This optimization method combines Adam and Nesterov Momentum, and the update rule is of the form:



Adam has combined two algorithms known to work well for different reasons-momentum, which points the model in a better direction. Nadam applied Nesterov momentum on adam, and improve its performance obviously in most cases.

Based on their excellent principles, the above two methods have good application prospects.