参数向量为 θ

学习率为 η

用于数值稳定的接近0的常数 ϵ

用于计算指数平均的接近1的常数 γ、 β

1. gradient descent：  
   
2. Momentum:  
   
3. Nesterov:  
   
4. Adagrad:  
   