考虑一个简单的线性函数：

设其损失函数为：

给定个样本分别为：、、、、、、、、、

设置基本学习率为0.01，手写实现以下几种优化器：

1、SGD

2、Momentum

3、RMSprop

4、Adam

优化方法中所涉及的其他超参数自定，

初始化

以SGD方法迭代100次所达到的损失值为基准，比较其他几种优化方法达到该值所需要的迭代轮次。