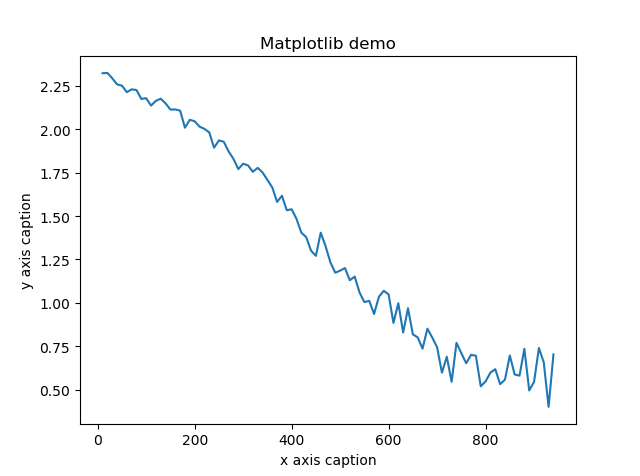
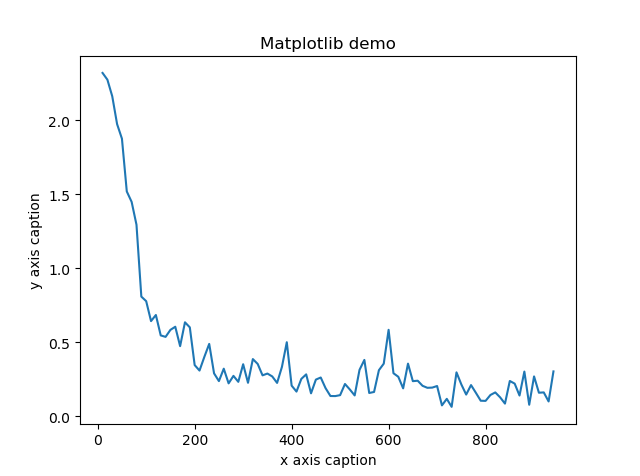
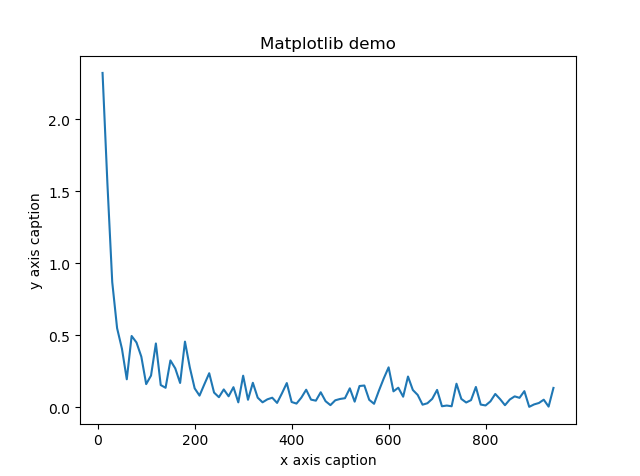
例：优化器的选择



SGD优化器收敛速度慢效果不佳，从图上看在同一训练周期下还没有收敛。



SGD-momentum0.9，给SGD添加动量之后效果变好了很多，在经过800次迭代可以看出一个平稳的趋势。



Adam优化器的效果优于加了动量的SGD，下降速度极快，后期振荡也要更小。

Reference：[关于深度学习的优化器选择](https://zhuanlan.zhihu.com/p/139041734)