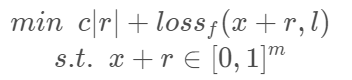
1. 为了判断之前训练的resnet50，基于gtsrb数据集的模型有效，能分辨出对抗样本，可以根据梯度反方向生成对抗样本，就是fgsm方法。如果是黑箱模型就是生成对抗样本时对真实模型不了解，要尽量模仿真实模型，根据输出对输入的负梯度来改变输入，以达到生成对抗样本的目的，并且为了提高成功率，使用了迭代版FGSM，梯度方向前有个权重参数可以调整攻击强度。优点是简单有效，缺点是权重参数需要预先指定，没有足够的经验无法确定合适的权重参数，从而达到更高的攻击成功率。
2. 而jsma算法就是同样生成负样本，但是根据对分类结果等影响最大的特征来生成对抗样本，优点是只要微小扰动便可以提高攻击率，求解最大值的过程中很难找到一个合适的点，从而改为找一对点，缺点是只能进行有目标攻击，且需要方向，在样本像素过多时寻找两个位置极为困难，某些情况下无法使用。

3、使用L-BFGS以生成GTSRB的数据集中测试集的对抗样本，以展示之前训练的resnet50能够分辨出对抗样本和正常样本，显示其性能好。搭建了一层flatten层和两层全连接层来拟合正常样本，以达到起码90%的准确率，通过box-constrained优化问题找到对抗样本，以线搜索的方式找到了最小r值满足f(x+r)=l最小值，其中c>0，从而找到最小扰动，。

优点是对抗样本能说明网络的泛化能力弱，缺点是代码实现中，是逐张图片搜索的，耗时长。

参考博客：<https://blog.csdn.net/StardustYu/article/details/104574051>

参考论文：https://arxiv.org/abs/1312.6199