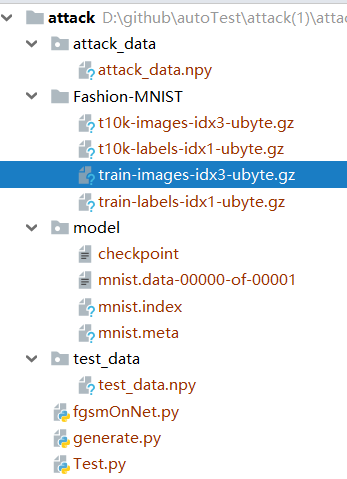
1. 项目介绍

针对深度学习图像识别模型的自动化测试框架。使用Fashion-MNIST数据集，使用fgsm算法生成对抗样本。

1. 代码结构
2. test\_data文件夹下面保存**test\_data.npy**文件，为fashion-mnist数据集的测试数据文件
3. attack\_data文件夹下面保存**attack\_data.npy**文件，为test\_data.npy中的测试数据一一对应的基于本项目实现算法生成的对抗样本文件。

3、Fashion-MNIST存放下载的训练数据集和测试数据集

4、Model为训练得出的模型

5、FgsmOnNet.py主要功能用于生成对抗样本。

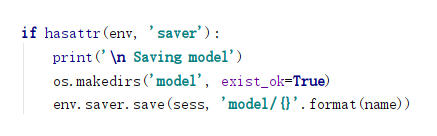
6、Generate.py用于生成对抗样本方法调用

7、Test.py自己测试用

1. 如何运行以及如何进行预测

运行：

此处将模型存储为上述代码结构中model的格式



其中：

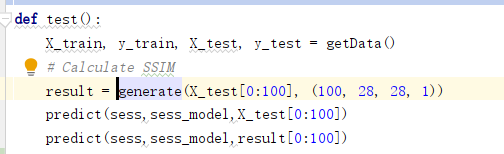
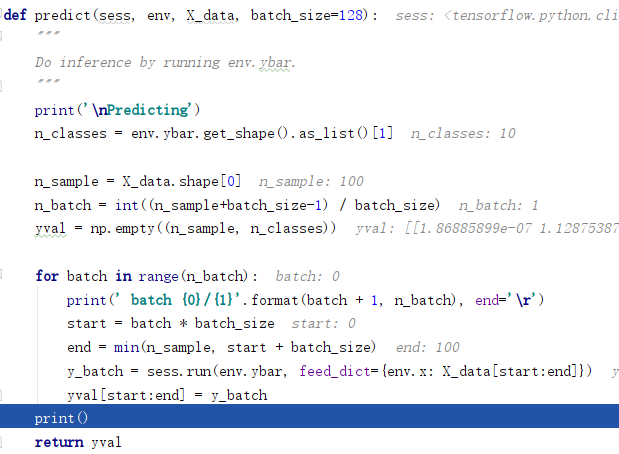
.meta文件保存了当前图结构

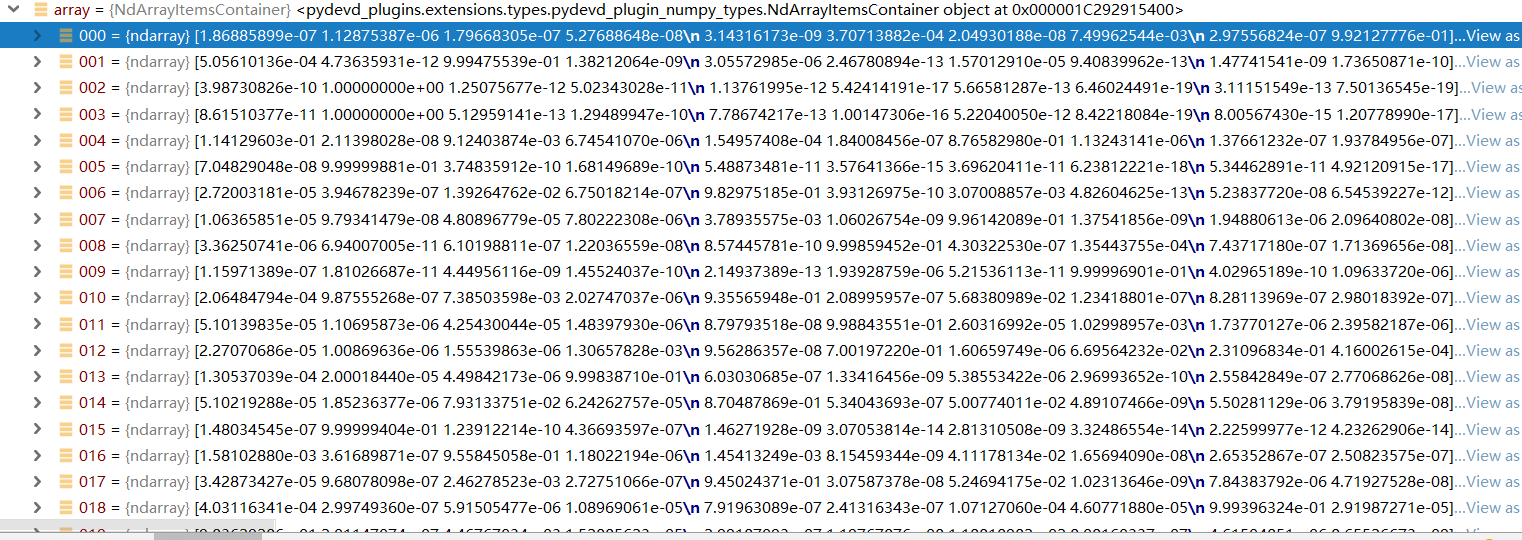
.data文件保存了当前参数名和值

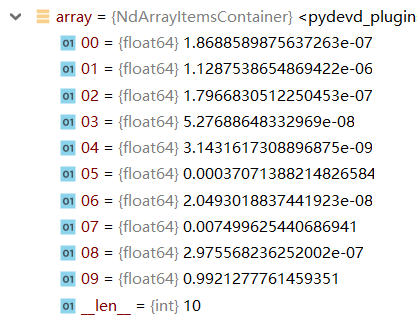
.index文件保存了辅助索引信息

通过下面的语句将模型重新加载进来

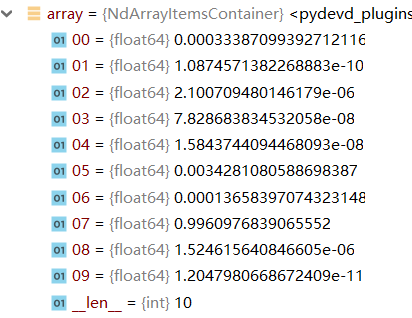
预测：

 用测试集中的前一百条数据做predict示例

运行到此处时，可以看到这一百条数据的预测类型

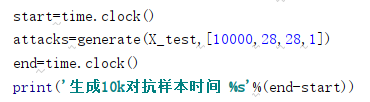
以其中一条为例：  


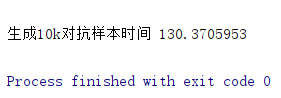
09类型最接近1（上图），所以预测该类型为09号

生成对抗样本后，该数据的07类型最接近1（上图），所以预测为07类型，预测错误

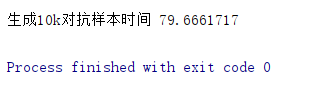
1. 生成attack\_data.npy的时间

计算时间代码：

结果：



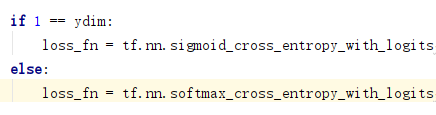
但是有时会70秒：



1. 算法详解

我们采用的是Fast Gradient Sign Method，即通过在梯度方向上进行添加增量来诱导网络对生成的图片X进行误分类，针对指定预期分类的对抗样本生成来说，我们需要更改它的loss损失函数，交叉熵损失函数变为：

loss = tf.nn.sparse\_softmax\_cross\_entropy\_with\_logits(logits=y\_logits, labels=[pl\_cls\_target])

通过计算网络对于参数的梯度进行噪声的生成即对抗样本的生成，在代码中体现在：

直观来看就是在输入的基础上沿损失函数的梯度方向加入了一定的噪声，使目标模型产生误判。

我的fgsm主要代码结构：

其中，model()构建图结构，evaluate()用来评估相似度和准确性，train训练模型时调用，predict用来预测结果，Make\_fgsm（）用来生成对抗样本

1. 个人感受

深度学习基础没啥，是真的难受。理解题目理解半天，好不容易知道需要做什么，TensorFlow配起来还老是出错，最后玄学配好。其次就是找模型，由于不知道助教那边使用的是什么模型最后测试，所以最好的方法因该是多模型assemble，然后尽可能覆盖，让黑盒尽可能白……但是时间和能力感觉有些吃力，就只选择了一个模型生成方法，fgsm生成对抗样本，也是要看很久，里面语法和思路蛮难理解，还有就是对TensorFlow很陌生，不知道怎么写，看他们的博客都需要查没一句是什么意思。应该感谢身边大佬们的不懈讲解和帮助。作业的收获是：如果我TensorFlow很熟练，应该可以效率很高，基本知道怎么尝试去找模型，找算法；配置环境应该题目出来之后就立即配好；如果自己陷入困境不会就问，别人的指点会让你茅塞顿开。

1. 参考
2. <https://undefinedf.github.io/2018/11/02/%E5%AF%B9%E6%8A%97%E6%94%BB%E5%87%BB%E4%B9%8BFGSM/>
3. <https://www.jianshu.com/p/0eb1e16a3850>
4. <http://baijiahao.baidu.com/s?id=1602583596173323823&wfr=spider&for=pc>
5. <https://blog.csdn.net/liuxiaodong400/article/details/83421164>
6. <https://www.jb51.net/article/136036.htm>
7. <https://www.cnblogs.com/my-love-is-python/p/10541030.html>
8. <https://blog.csdn.net/sinat_36618660/article/details/98665482>