**基于机器学习和深度学习Iris数据集的分类**

**2017Z8009061079 刘林虎**

1. **数据集**

Iris数据集以鸢尾花的特征作为数据来源，包含150个数据，分为3类，每类50个数据，每个数据包含4个属性，三类分别为:setosa, versicolor, virginica，数据包含4个独立的属性,这些属性变量测量植物的花朵,比如萼片长度, 萼片宽度,花瓣长度, 花瓣宽度.

1. **实验目标和工具**

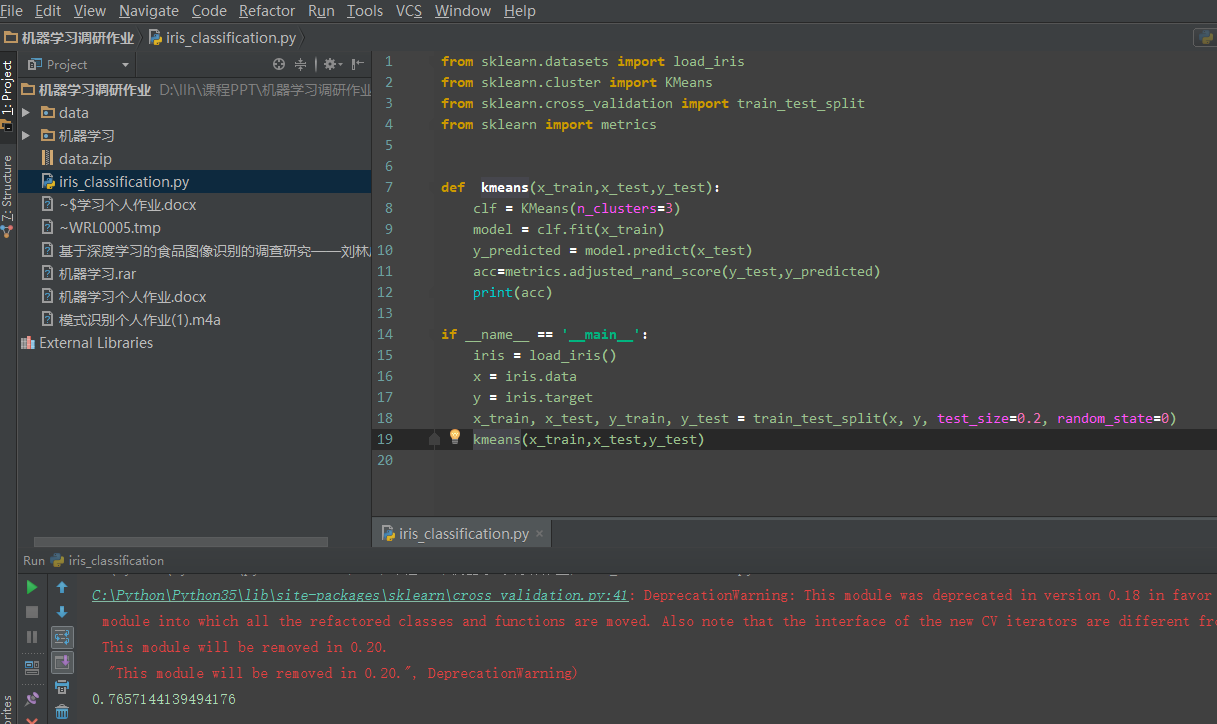
实验目标：1）熟悉机器学习一般算法。2）学会用深度学习框架搭建深层网络。

实验工具：1）python语言，使用scikit-learn库。2） 使用Tensorflow深度学习框架搭建简单神经网络。

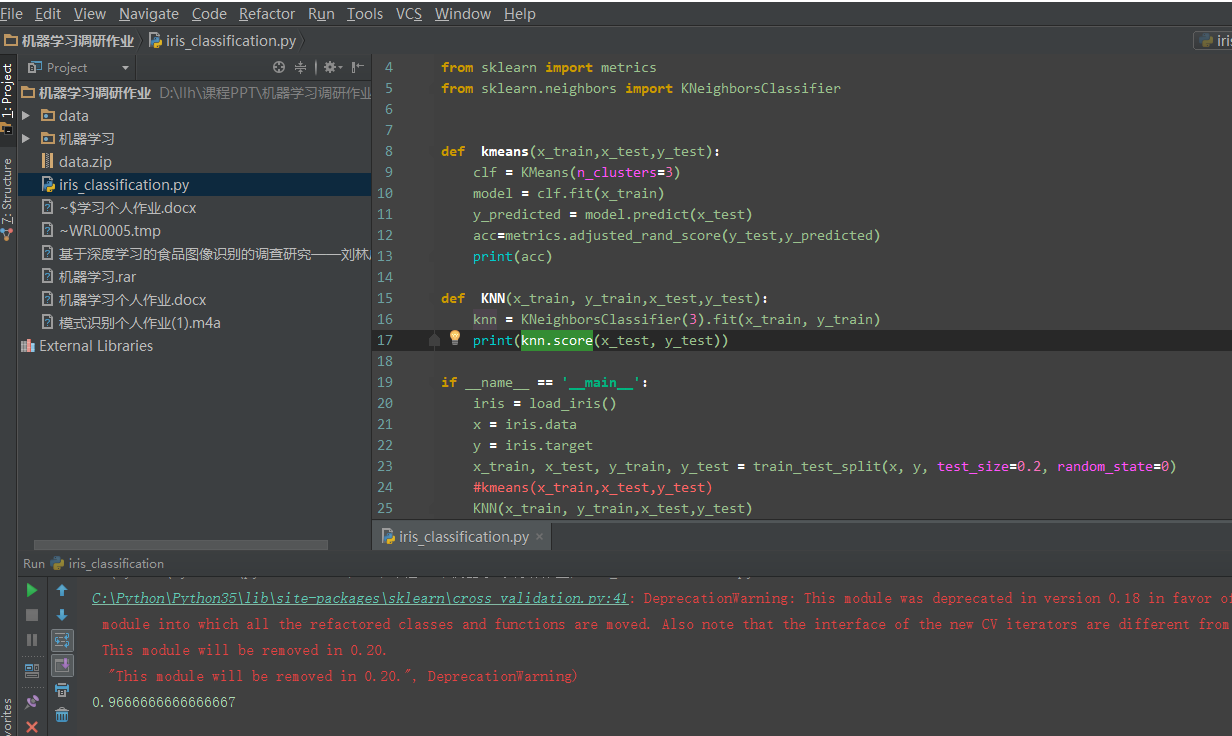
1. **机器学习方法**

Scikit-Learn是用Python开发的机器学习库，其中包含大量机器学习算法、数据集，是数据挖掘方便的工具。本实验所有机器学习算法都是基于Scikit-Learn库进行的。代码简单，运行速度快。

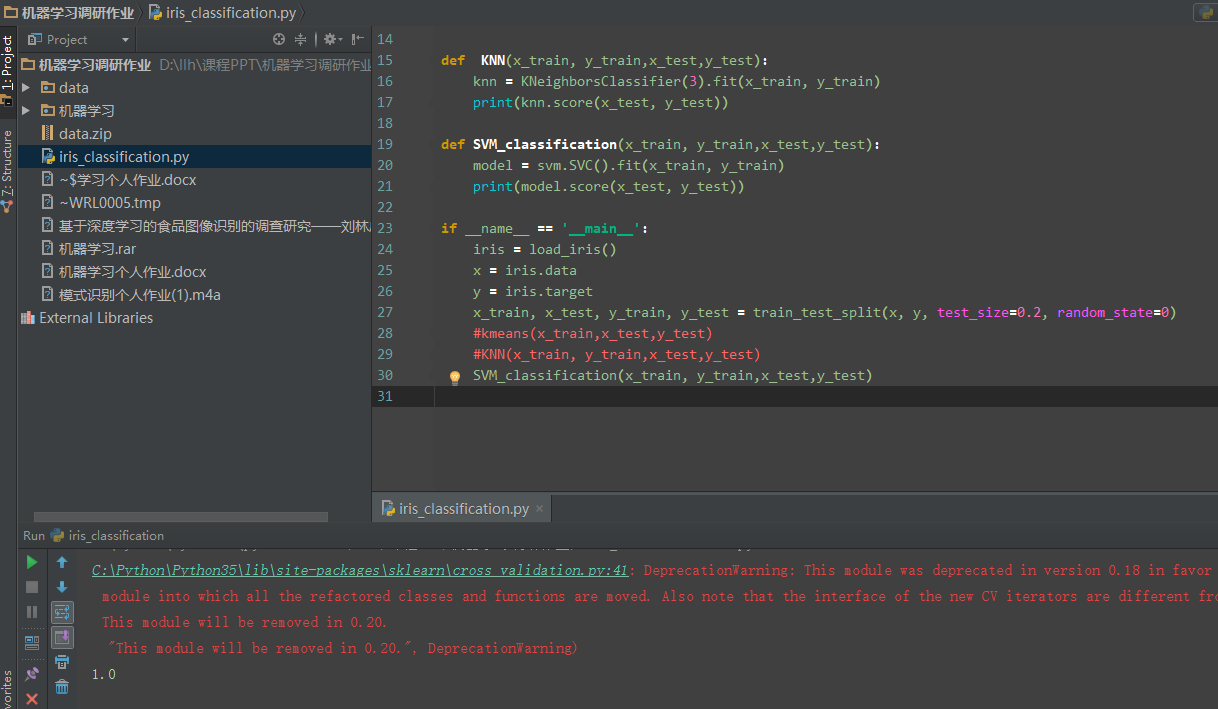
1. 用K-means分类。将数据集20%的划分为测试集，用来测试K-means模型。聚类中心K=3,具体代码请见下图，最终分类准确率为0.7657



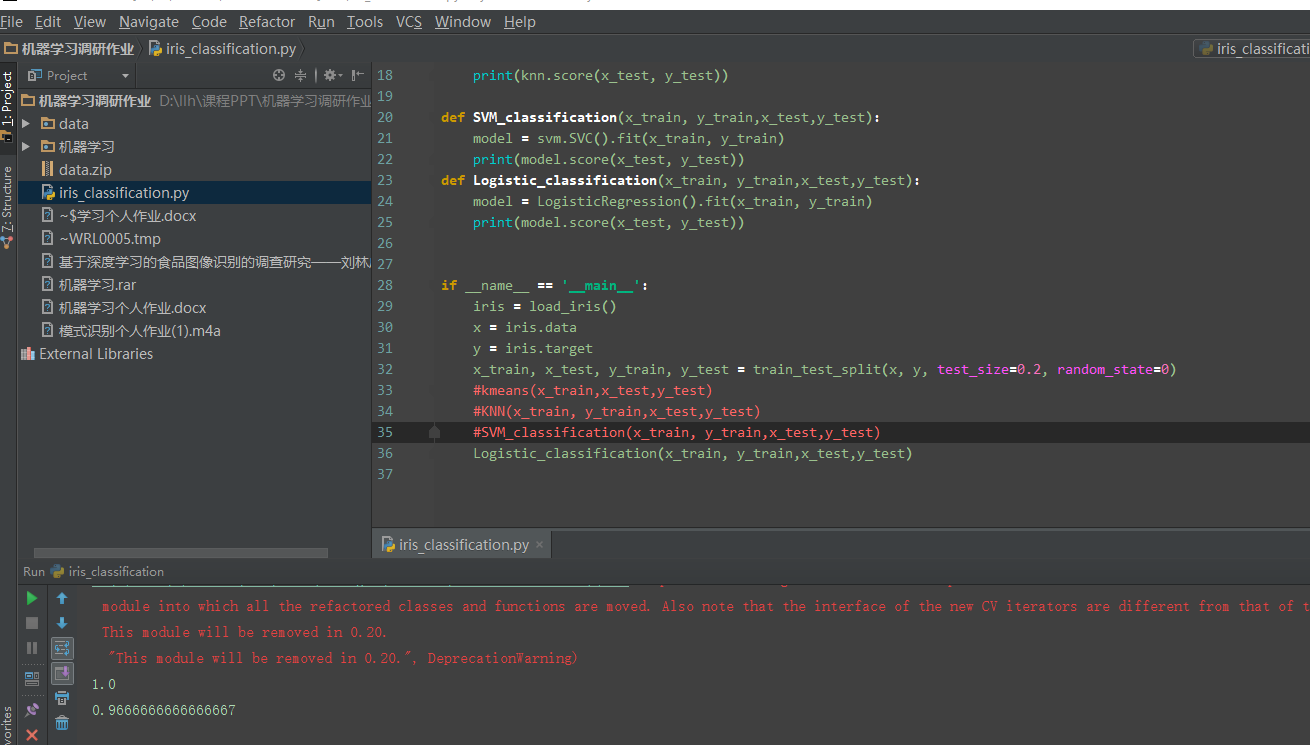
2) 用KNN分类。选取数据集20%作为测试集。最终分类结果为0.966



3)用SVM分类。。选取数据集20%作为测试集。最终分类结果为1



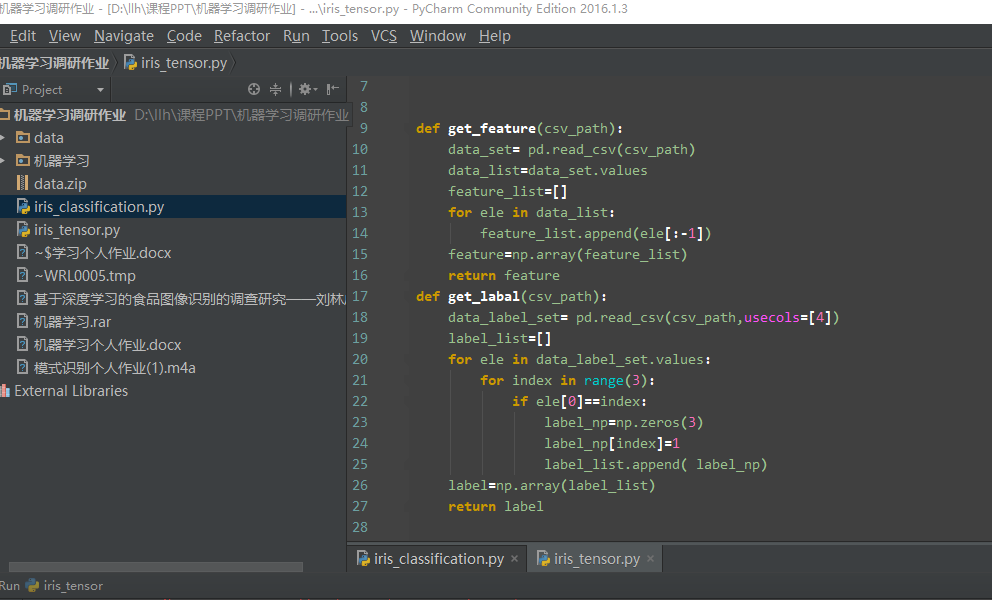
4)用逻辑回归的方法。选取数据集20%作为测试集。最终分类结果为0.966

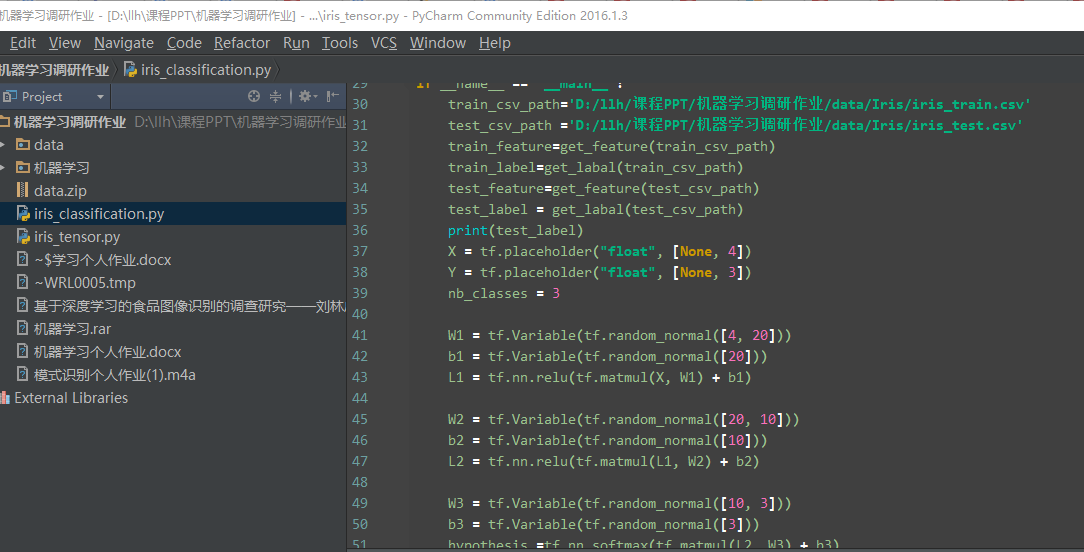


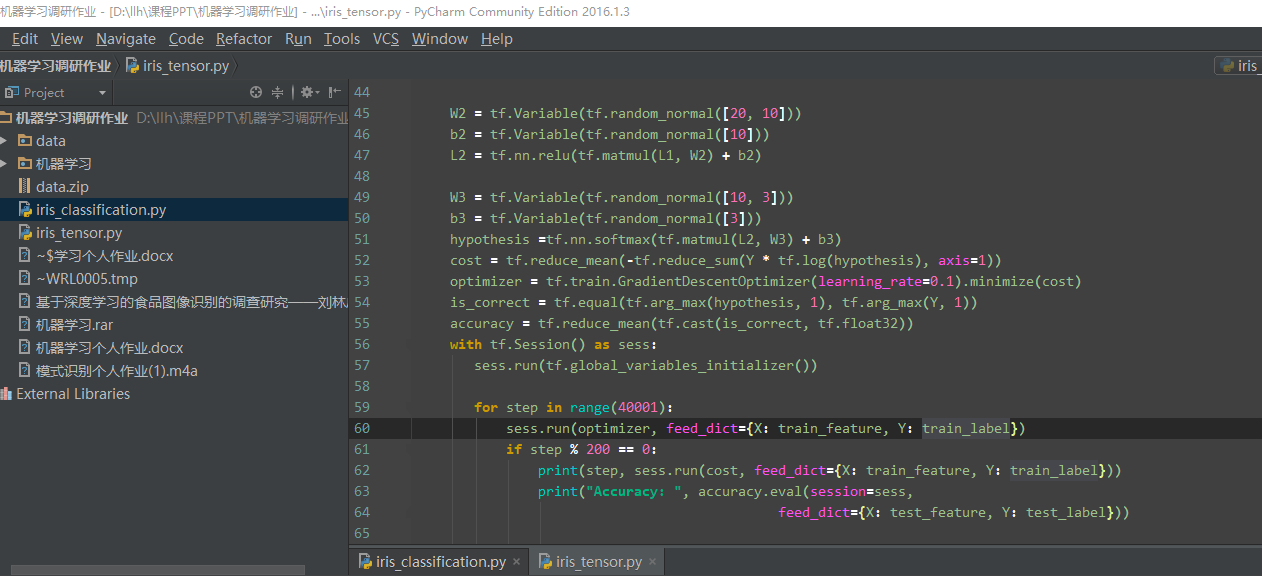
1. **深度学习方法**

现在比较流行的深度学习框架为caffe、Tensorflow等，但caffe封装太过于紧密，而且Iris数据集没有提供图片，是人为的手工特征，所以caffe不适用，还要借用于python接口。刚好我正在学习Tensorflow，所以本次实验使用Tensorflow深度学习框架，由于数据比较均衡和少量，所以本次试验建立2层网络全连接，最后用softmax分类，使用梯度下降法优化，Tensorflow官网有具体iris数据集的案例。为了自己更好的学习Tensorflow深度学习框架，自己重新编写框架。150数据我划分为120的训练，30的测试。

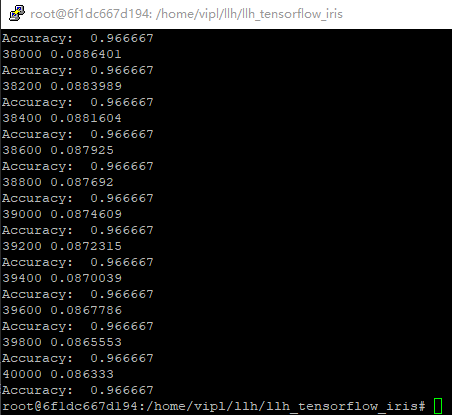
具体代码如下所示：







由于台式电脑没有装Tensorflow,所以在服务器上跑的程序。学习率为0.0001,经过40000次迭代，最终准确率为0.966



1. **总结**

经过这次机器学习算法和深度学习在iris数据集上分类任务，熟悉了普遍机器学习算法，也学会了python Scikit-Learn库的使用。更重要的是自己亲手使用Tensorflow 搭建了一个深度模型，使自己入门Tensorflow，也对Tensorflow运作方式有了进一步的了解。让自己从caffe转入Tensorflow打下了基础。以上所有代码都是自己亲手敲上，对代码能力也得到了加强。感谢王老师8周课时的辛苦讲授，使我受益匪浅，也对人工智能领域有了更深的认识。