实验名称：线性回归

一、实验环境：

系统版本：CentOS 7.5
scikit-learn 版本： 0.19.2
pandas 版本：0.22.4
numpy 版本：1.15.1
python 版本：3.6.2

二、实验目的：

掌握 Python 编程
掌握 Pandas 编程编程
掌握 Sklearn 的使用
掌握线性回归模型的基础知识
掌握线性回归的使用方法
掌握 matplotlib 绘图方法

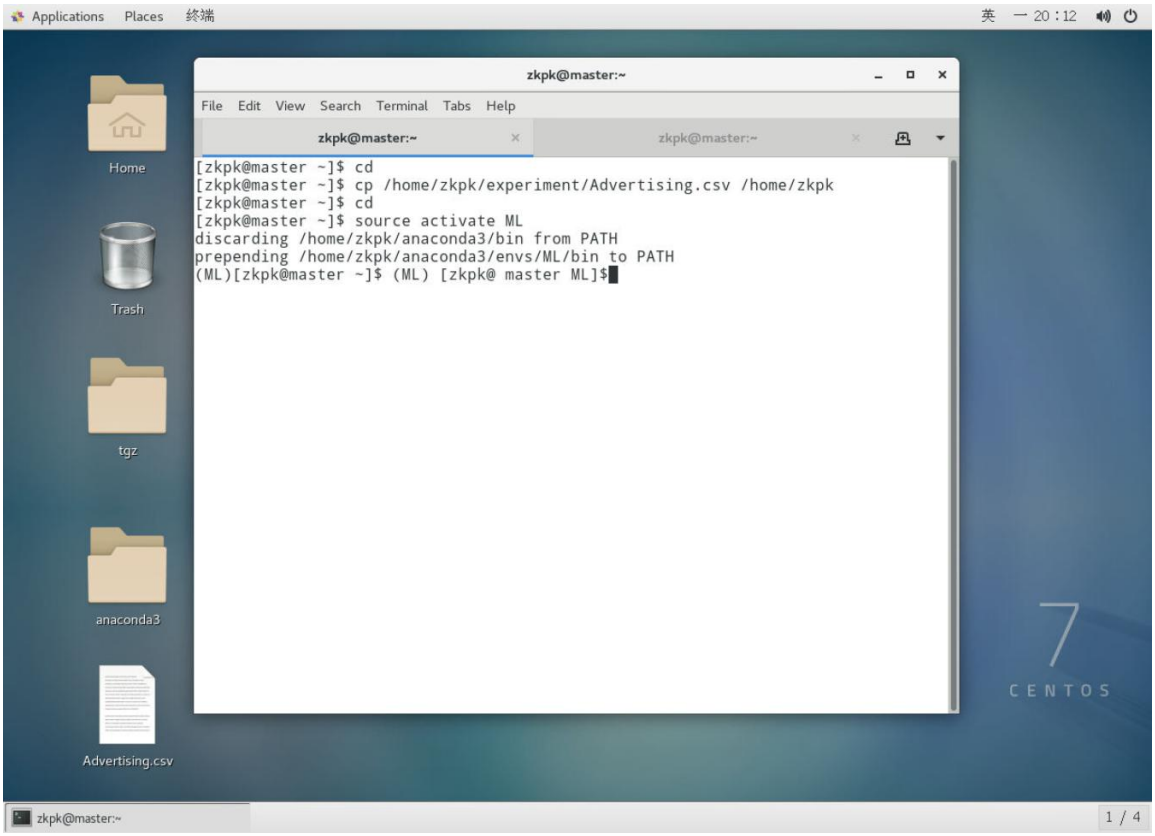
三、实验要求：

四、实验内容：

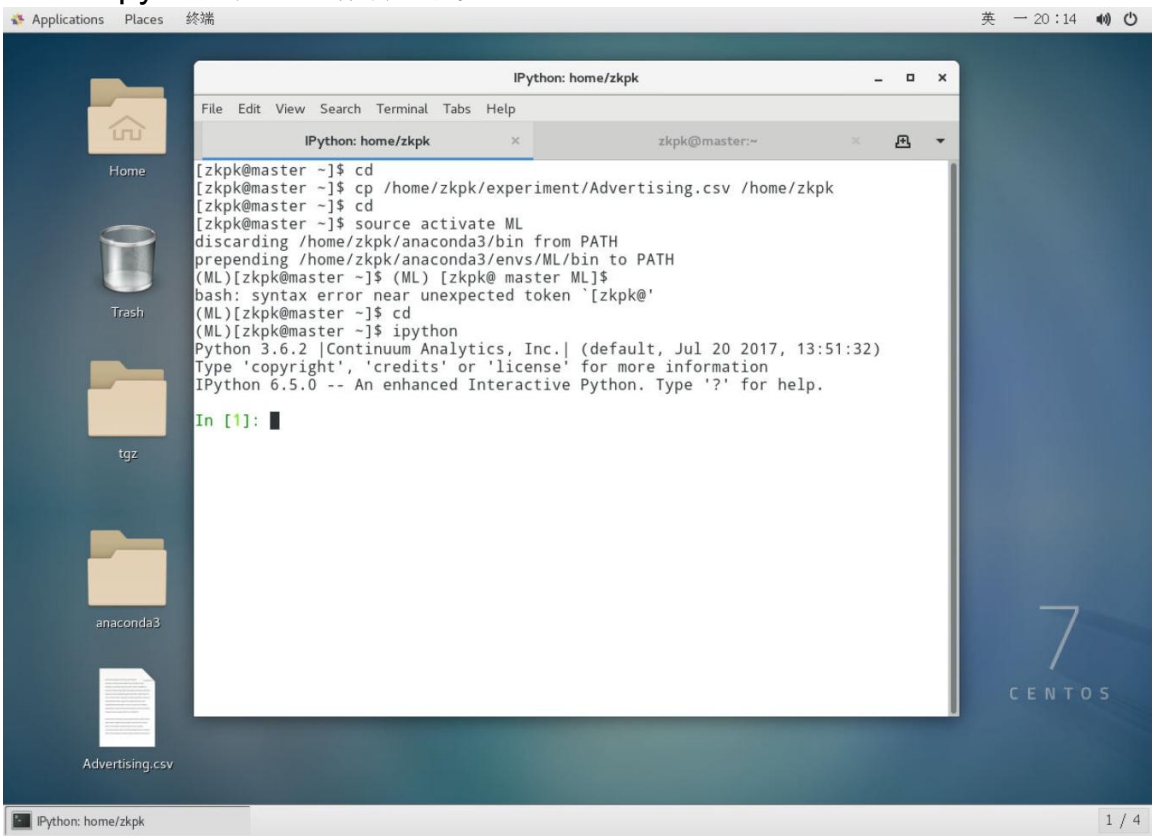
本实验中提供一份关于产品广告费用与对应产品销量的数据文件 Advertising.csv 文件，利用此文件建立线性模型、训练模型、用模型做预测分析。

五、实验步骤：

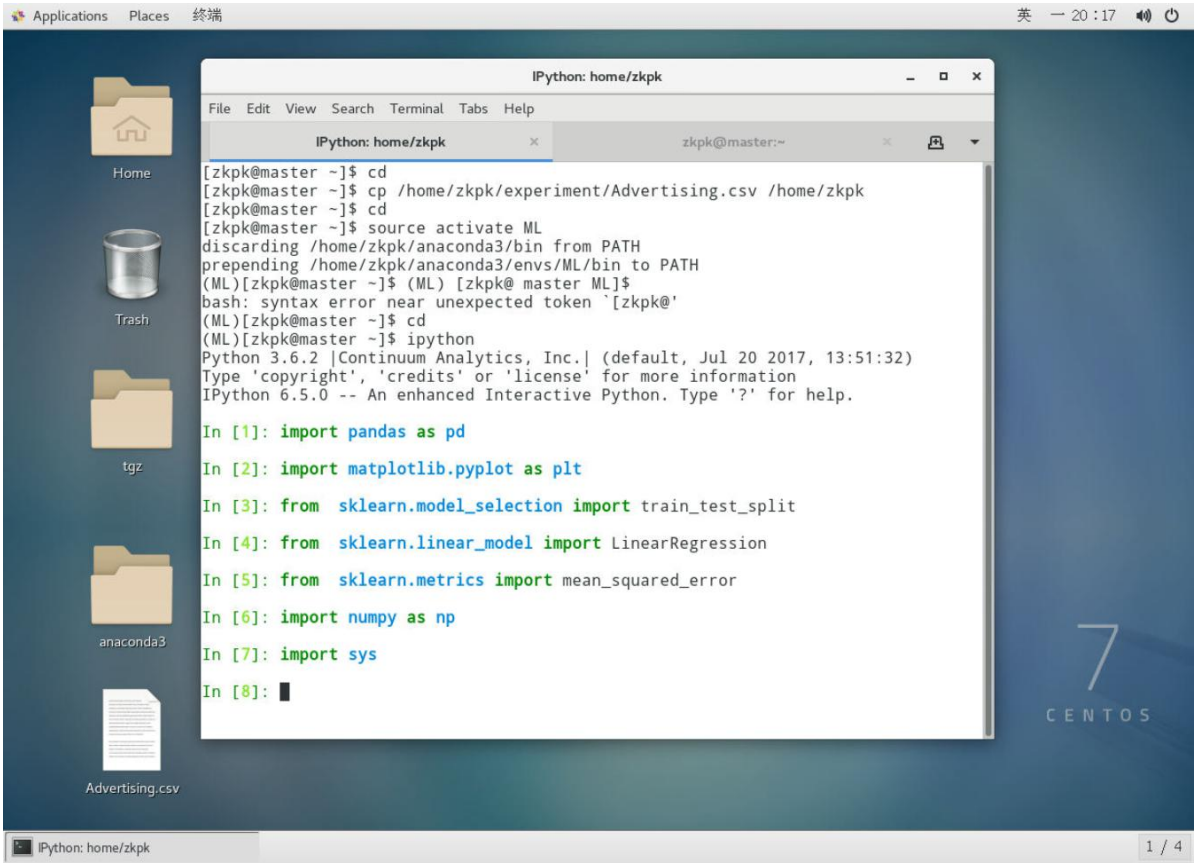
1.从 zkpk 的公共目录下拷贝实验所需的数据文 Advertising.csv 到 zkpk 的家目录下
2.进入 ML 虚拟环境



3.进入 ipython 交互式编程环境



4.导包



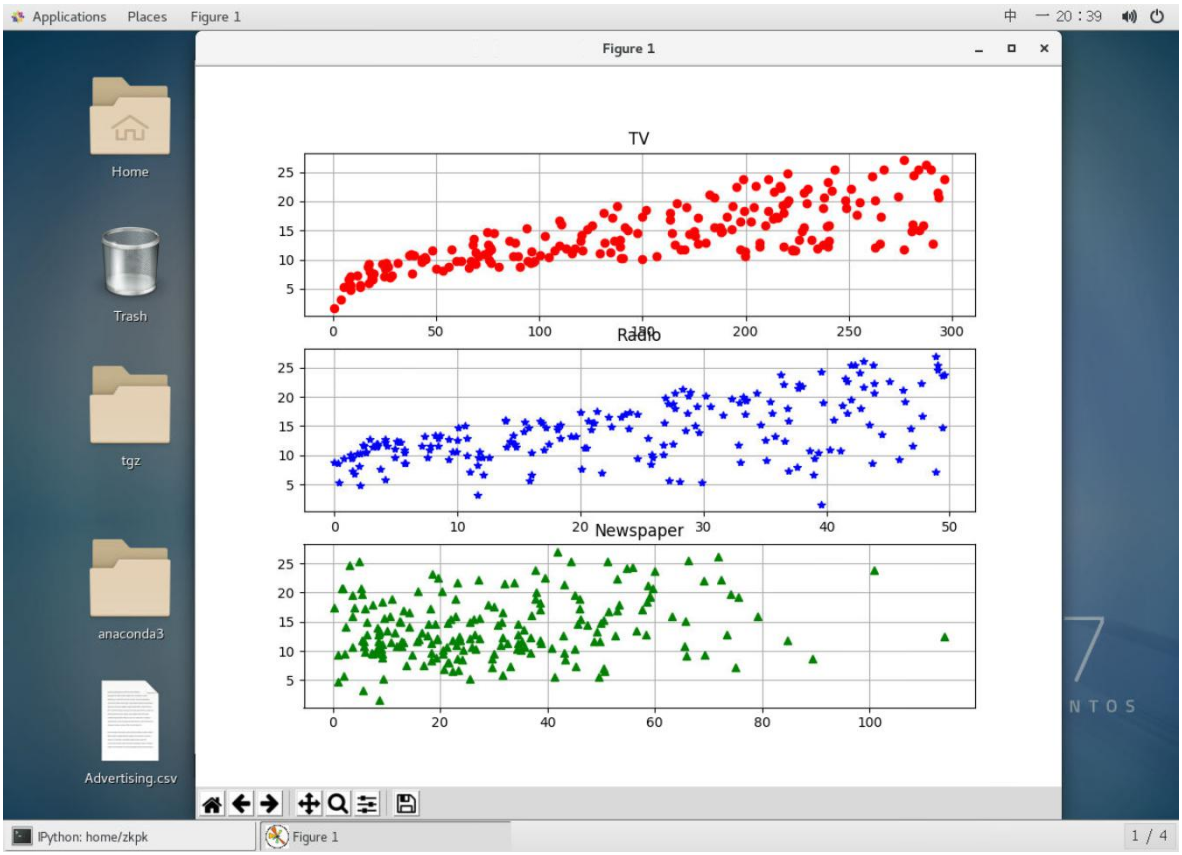
5.

读取数据文件

打印文件的前几行

显示文件的 shape

使用 pandas 读取相应的维度分别作为特征值 X， 和标签值 Y 绘制不同特征和标签的关系



6.

分析上边结果图，在报纸“Newspaper”上所花广告费用与商品的销量不成线性相关的，所以后面建模时，可以尝试删掉该特征

使用 sklearn 自带的数据预处理模块对数据集进行切分，构建训练集和测试集，比例为 7 比 3

使用 sklearn 的线性回归类建模，参考 normalize=True 表示指定对训练数据进行归一化操作；n_jobs=-1 表示使用所有的 cpu 进行训练。

打印模型的相关参数

使用训练好的模型进行预测

使用 **RMSE**（标准误差）对模型进行评估

将标签的实际值和预测值用图展示出来，直观的观察拟合程度。

六，实验结果与分析：


sales real with pred
predict
test
value of sales
the num of sales
0
10
20
30
40
50
60
5
10
15
20
25