**第二次个实验报告**

1. 平均值与方差计算

环境：python3.6 + jupyter notebook

依赖库：pandas, numpy

首先，从csv文件中读取原数据，使用到pandas中read\_csv()读取，读取出（0, 5）两列数据，分别是点号与“钳形表100A电流”。数据类型Pandas.DataFrame

通过阀值（“钳形表100A电流” > 1.0）比较，找到有效数据，即从触电点（开始波动）到无波动之间的数据。数据类型Numpy.Array

计算每个csv文件中有效数据的平均值与方差。使用numpy.mean()算取均值，使用numpy.var()算取方差。并将每一个csv所算取结果形成列表，并给每个csv的种类赋值，“兔子”为1，“小羊”为2。数据类型：mean列表，var列表，type列表。

写出csv，将三个列表，转化为Pandas.DataFrame格式，使用dataframe.to\_csv()写出。

结果：详见mean\_and\_var.csv文件，111行，三列数据

代码：见“experiment2.ipynb”

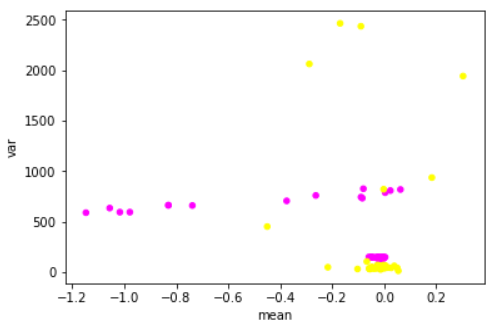
二．分类

环境：python3.6 + jupyter notebook

依赖库：sklearn，SVC，RandomForestClassfier，train\_test\_split

数据集为上面所得数据，输入X为平均值与方差，输出Y为种类。数据集使用train\_test\_split()划分为训练集与测试集，测试集数据数量占总数0.3，随机划分。

数据散点图：



分类器SVC，高斯核函数rbf

结果：

train score = 0.961038961039

test score = 0.735294117647

实际: [1 1 2 2 2 1 2 1 2 2 1 2 2 1 1 1 2 2 1 1 1 1 1 2 1 2 1 2 1 2 1 2 1 1]

预测: [1 1 1 2 1 1 1 1 1 2 2 2 1 2 1 1 1 2 1 1 1 1 1 2 1 2 1 2 1 2 1 2 2 1]

分类器RandomForestClassifier

train score = 1.0

test score = 0.764705882353

实际: [1 1 2 2 2 1 2 1 2 2 1 2 2 1 1 1 2 2 1 1 1 1 1 2 1 2 1 2 1 2 1 2 1 1]

预测: [1 1 1 2 1 1 1 1 1 2 2 2 1 2 1 1 1 2 1 1 1 1 1 2 1 2 1 2 1 2 1 2 2 1]

代码：详见“experiment2.ipynb”

二．画折线图

环境：python3.6+Spyder

库：matplotlib

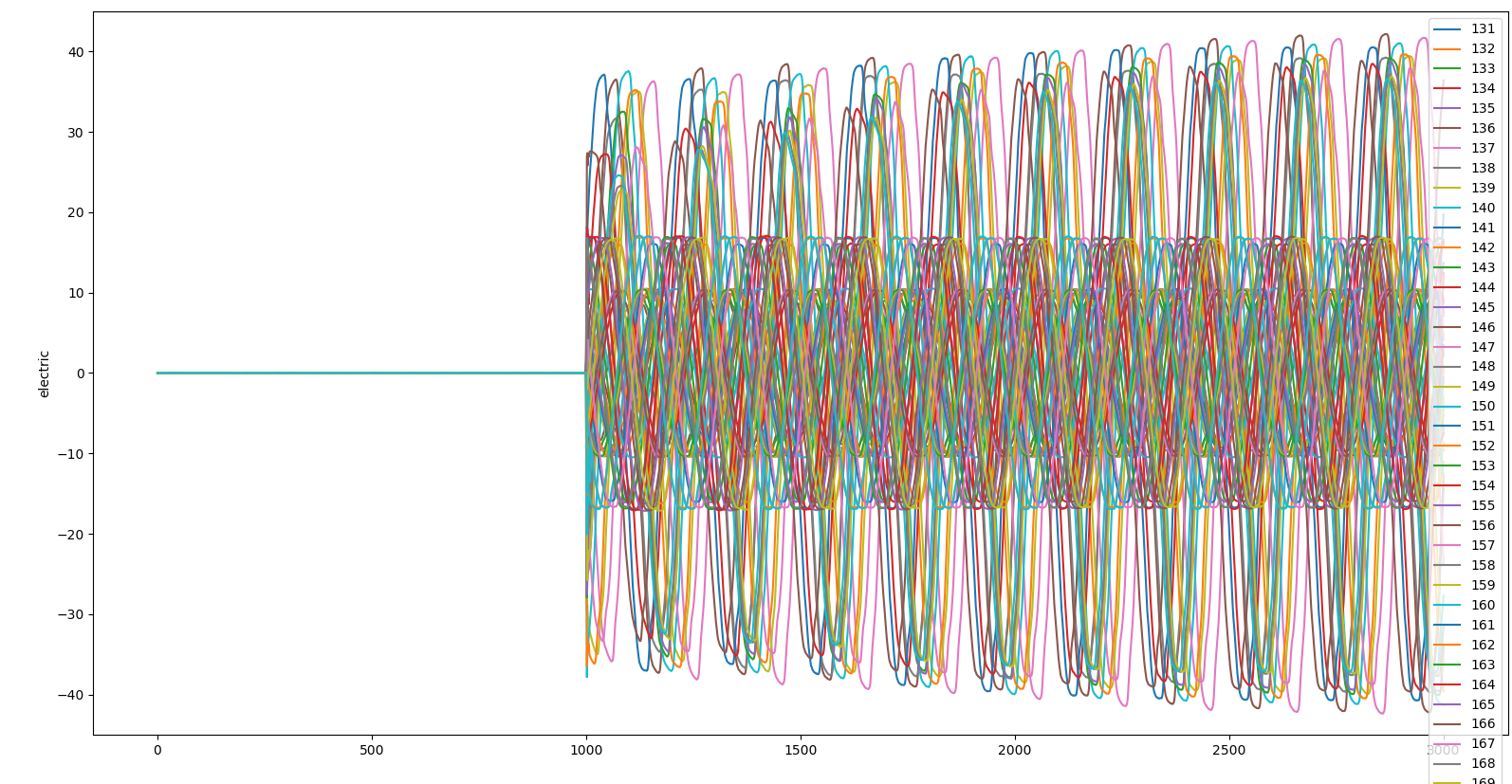
横轴为点号，纵轴为“钳形表100A电流”

将60条折线图画到一张图中：（em，三种理解。。。）

第一种，60条折线图全部重叠在一起：

只需一个for循环，直接plot，即可获得重叠效果。十分丑陋

代码：详见“subfigure2.py” ，运行程序即可获得图



第二种，60个子图：画布分为15\*4

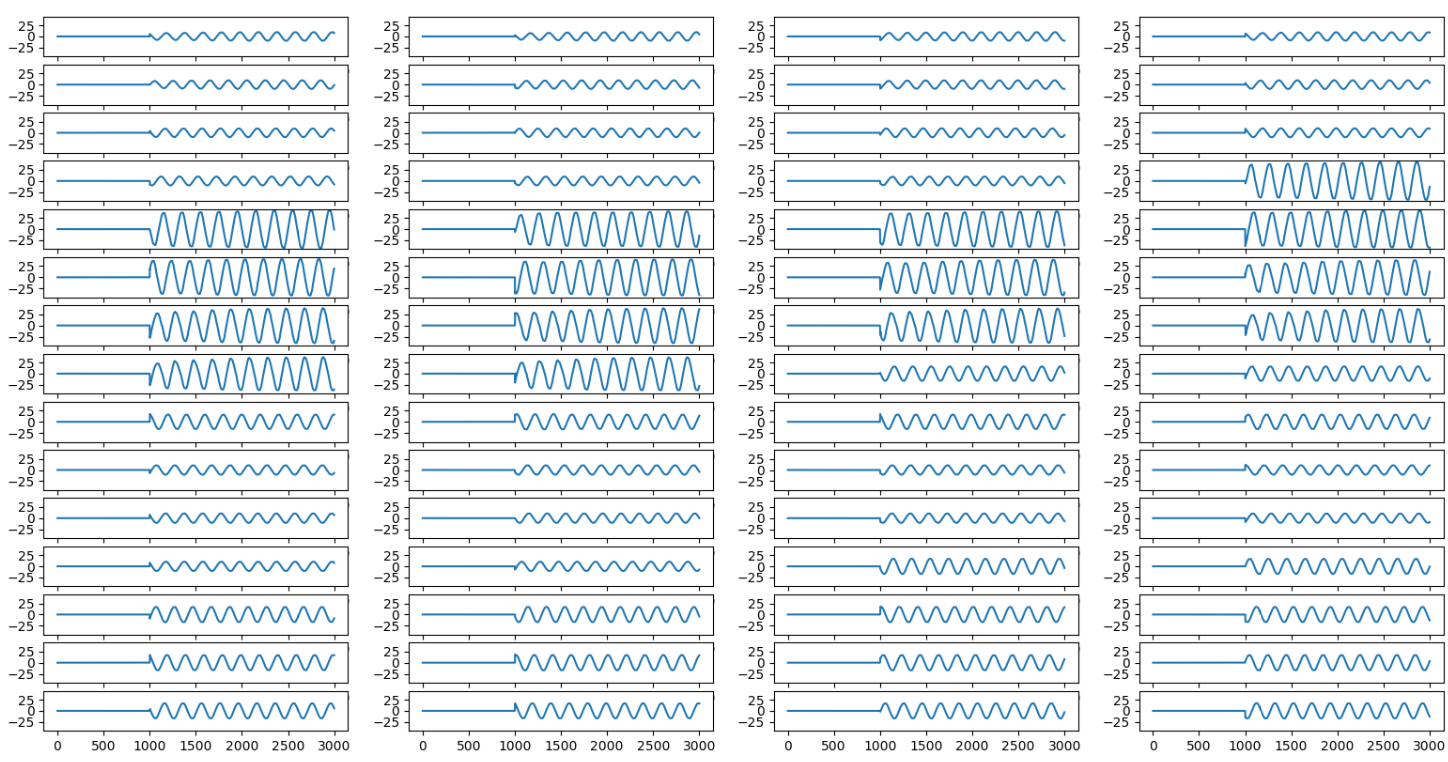
使用subplot，一张图中包含60个子图

for i in range(131,191):

Data = get\_one\_csv(i, '兔子数据')

plt.subplot(15,4,i-130)

代码：详见“subfigure.py”，运行程序即可获得图



第三种，动态图，一副图为一帧，每200毫秒自动播放下一帧。

使用matplotlib包中FuncAnimation，能够产生动态效果

anim = FuncAnimation(fig, update, frames=np.arange(0, 61), interval=200)

产生60帧动画，每帧之间间隔200ms

Update为自定义函数，每次通过传入帧号，更新折线。

代码：详见“experiment2.py”，运行程序将能看到动态效果图，横坐标标签显示当前是哪个csv文件中数据。

