KNN，（K均值分类）：

核心思想：距离公式

d==

Ps：KNN聚类后续再做，这个只是单纯的分类，我已经认为的给GDP增长率划分了4个等级，以10%,13%,15%三个点作为分类界限，所以本次不需要随机假设聚类中心点

KNN部分流程

将国民生产总值转换成增长率人为划分增长率等级标准化数值计算距离并根据距离来划分类别计算正确率画图寻找超参数

1. gdp\_to\_rate(gdp)

gdp-----数据库中的国内生产总值

a--------gdp增长速率

由于数据库里存放的数据是生产总值，为了探索GDP增长速度与进出口量的关系，所以需要先将生产总值转成GDP增长速度

1. groups(gdp\_rate)

gdp\_rate--------gdp增长速率

gdp\_grade------gdp增长速率的等级（手动划分为ABCD四个等级）

1. normalize(data)

data------------从数据库中导出来的数据

将data里的特征值进行标准化，标准化的作用有效防止了某个特征值偏大的问题，也有效的降低了0值的作用，（出现0的时候既可能是真是数据就会产生，也有可能是数据丢失造成的，需要考虑0值）

此部分可百度数据的归一化和标准化

1. create\_set(data)

data------------从数据库中导出来的数据

dvl\_set---------验证集，抽出来剩下的数据集

test\_set--------测试集，从18行数据中抽出来的4行数据集

随机从18行数据中抽取4行数据作为测试集，测试KNN算法的准确率以及K值的取值问题

1. create\_result(test\_set,dvl\_set,k,data)

dvl\_set---------验证集，抽出来剩下的数据集

test\_set--------测试集，从18行数据中抽出来的4行数据集

k--------------K值，表示从距离最小的前K项中挑选分到的类别最多的作为测试数据的分类

data------------从数据库中导出来的数据

1. cor\_rate(result\_set)

result\_set保存两列数据，pre列表示预测的分类，cor为实际的分类

计算正确率

1. create\_plt(data)

画图，可视化K=3/4/5各情况下的正确率；数据量大小了，正确率有点没眼看