

- 前情回顾
 - 预测整体的交通模型
 - 使用天气因素进行预测
 - 小结

前情回顾

上一周的任务是实现几何分类（K-mean)与T聚类的迭代回归，目前尚未完成，因此，这周要继续完成。

下一阶段的任务是

预测整体的交通模型

本次实验中预测交通信息的时候，分为两个阶段：

1. 考虑天气因素、时间因素与温度因素，使用以上因素分别进行预测；
2. 对各自的实验结果进行加权平均；

这两周尽量完成的任务是使用时间段因素进行预测与根据天气因素进行预测：

使用天气因素进行预测

使用天气因素进行预测需要考虑以下前提：

- 前提一：天气因素有四种
- 前提二：不同时段的天气是有关系的
- 前提三：不同天气出现的概率可以由频率预测

因此在根据过去数据进行预测的时候需要遵循以下规则：

	雪	雨	雾	晴
雪	1	a1	a2	a3
雨		1	a4	a5
雾			1	a6
晴				1

可以凭借此矩阵根据天气情况确定权值，该权值将会用于预测当前的车辆流量。

1. 允许没有记录的天气信息。

因为实际上有一些天气信息是没有记录的，那么对于聚类中的计算而言，该部分信息是可以被其他时段的信息所代替表示的。

也就是说，即使没有早6点的数据，使用晚6点的数据代替亦可。

2. 一天可能有多种天气。

此时权值就根据一天中的多种天气综合考虑，返回矩阵中的值a1 a2 a3 a4 a5 或者 a6。

3. 计算权值信息

填充矩阵中的值，即计算两种天气同时出现的频率。需要对两个数据集DC和NYC分别进行计算，计算的频率一定处于 $[0, 1]$ 中，将计算得到的数据归一化，填入表格即可。

4. 检验结果与调整

该矩阵中的变量遵循以下原则，请务必使其符合以下原则，允许适当的修改与数学模型的套用：

$$a_1 > a_2 > a_3;$$

$$a_4 > a_5;$$

$$a_6 > a_5 > a_3;$$

$$a_4 > a_2;$$

小结

综上，这份任务比较简单，只需要统计数据集中天气出现的频率即可，请将计算得到的矩阵写成一个函数：

- 参数：使用0、1、2、3表示四种天气；
- 输出：介于0到1之间的权值double值；