**问题3.** 使用附件1、2中的数据，建立一个同时适用于A、B、C三个监测点（监测点两两间直线距离>100km，忽略相互影响）的二次预报数学模型，用来预测未来三天6种常规污染物单日浓度值，要求二次预报模型预测结果中AQI预报值的最大相对误差应尽量小，且首要污染物预测准确度尽量高。并使用该模型预测监测点A、B、C在2021年7月13日至7月15日6种常规污染物的单日浓度值，计算相应的AQI和首要污染物，将结果依照附录“污染物浓度及AQI预测结果表”的格式放在论文中。

技术路线：这一问需要使用“监测点A逐小时污染物浓度与气象实测数据”结合“监测点A逐小时污染物浓度与气象一次预报数据”共同完成，重点是建立“监测点A逐小时污染物浓度与气象一次预报数据”和“监测点A逐小时污染物浓度与气象实测数据”之间的联系。即使用实测数据来修正一次预报数据，因为一次预报数据的预测效果不好，如果单独对其进行建模的话得到的模型偏差将会很大。

首先，对数据进行处理。这里因为题目中提到一次预报数据对临近日期的预测较为准确，而一次预报数据中每日会对当天、第二天、第三天进行预测，故考虑到其对后两天预测准确率低，将只保留第一天的预测数据。

其次，将“监测点A逐小时污染物浓度与气象实测数据”中的六种污染物值和一次预测数据拼接到一起，这里因为数据存在丢失的情况，所以需要对缺失的数据进行修补，本文使用“监测点A逐日污染物浓度实测数据”来对“监测点A逐小时污染物浓度与气象实测数据”的缺失数据进行补充，在将两表之间的对应关系处理好了以后需要对缺失值和异常值进行处理，缺失值使用其前一条数据和后一条数据的均值来填补，（对于一条数据缺失数据较多的则直接删除），对于异常值（如负数的浓度）的处理方式和缺失值相同。

最后，使用python导入数据，将“监测点A逐小时污染物浓度与气象实测数据”中导入的污染物浓度减去一次预测数据中的污染物浓度，保留其差值和环境因素作为新的数据集，这样我们的模型预测出的结果就是实测值减一次预测值，当我们对未来三天的污染物浓度进行预测后，将模型结果加上一次预测值就能得到实测值，以此，建立了一次预测数据和实测数据中的关系，达到了实测数据修正一次预测数据的目的。到此处就结束了数据的预处理部分。

关于模型的构建采用神经网络的方式完成，但由于气候因素较多，为了将那些关联性较低的数据剔除，使用了灰色关联度分析（该算法具体见群里链接）。

在完成气候因素的选择之后，观察数据发现各类数据的量级不同，为了消除量级不同造成的影响对数据再次进行处理，主要为对神经网络输入数据进行零均值处理和归一化处理（公式可以写）。

针对构建的神经网络因为一次预测数据偏差较大，最终效果不佳的情况，使用CS算法对其进行优化，这里对CS算法进行了优化，主要为使用混沌反向初始化的方式取代其原本随机初始化的方式，（随机初始化的方式生成的数据在可行域中表达能力并不强，不能充分均匀布满整个可行域区间），混沌初始化选择Tent混沌初始化，其初始值经测试alpha选择0.77，xn初值选择0.33.Tent初始化公式为：

其中。

完成Tent混沌初始化之后，在Tent算法产生的初始解的基础上取其反向解，并计算这两组解的适应度，依据贪婪法则留下其中的最优，以此构成初始解。这样将能在解空间中生成更加均匀、更具遍历性、表达能力更强的初始解。

在CS算法得到神经网络模型的一组较优参数之后，赋值给神经网络，优化其效果