**设备故障与开机时间的相关分析模型使用说明**

1. 模型用途

本相关分析模型通过考虑设备不同时间段的开机时长、设备开机和关机时间节点以及设备不同阶段的状态数据等因素，分析设备发生故障是否是由开机时间不当引起的。

1. 数据说明

输入数据X具有4个维度，具体如下：

第一维特征是开机时长，用数字表示，数值越大表示时间越长；

第二维特征是开关机节点，用数字表示；

第三维特征是设备运行状态数据，用数字表示；

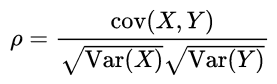
第四维特征是设备是否故障，用0表示没有故障，1表示故障。

比如第一组数据25\_28\_8\_0，表示开机时长为25h，开关机节点在第28h，设备运行状态为8，未发生故障。

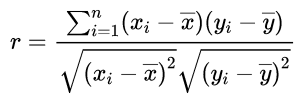
1. 算法原理

相关分析是研究两个或两个以上处于同等地位的随机变量间的相关关系的统计分析方法。相关系数可以用来描述定量变量之间的关系。相关系数的符号（±）表明关系的方向（正相关或负相关），其值的大小表示关系的强弱程度（完全不相关时为0，完全相关时为1）。

两个变量之间的相关程度通过相关系数r来表示。相关系数r的值在-1和1之间，但可以是此范围内的任何值。正相关时，r值在0和1之间，散点图是斜向上的，这时一个变量增加，另一个变量也增加；负相关时，r值在-1和0之间，散点图是斜向下的，此时一个变量增加，另一个变量将减少。r的绝对值越接近1，两变量的关联程度越强，r的绝对值越接近0，两变量的关联程度越弱。总体相关系数计算如下：



而样本相关系数计算如下：



本模型是规范化的相关分析模型，通过对设备故障与开机时间进行分析，对两组数据进行线性相关分析，计算典型相关系数，判断哪些因素与设备故障有关。

1. 使用方法

将输入数据用下划线连接，比如一组数据25\_28\_8\_0。每组数据用空格分隔。

示例中使用的是input.txt文件，真正使用时将新文件名写入到下列指令最后一项，即

**python model.py --train\_dir 新文件路径**

在此目录下打开命令行，输入：

**python model.py --train\_dir ./input.txt**

输出内容保存在output.txt文件中。