# 实验8\_任务1\_多元回归

# 问题：

任务描述：在给定的数据⽂件中，每⼀⾏代表⼀个开盘⽇中的股指交易涨跌值，第⼀列记录具体⽇期，其后每⼀列代表⼀项股指数据，共九列，依次为：ISE(TL-based), ISE(usd), SP, DAX, FTSE, NIKKEI, BOVESPA, EU, EM。回归任务是通过后⼋项股指来对第⼀项股指（ISE(TL-based)）的数值进⾏预测。请完成下列⼯作：

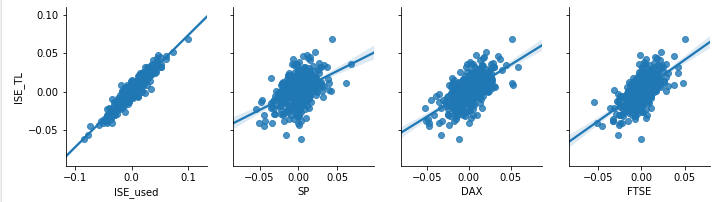
1. 使⽤可视化的⽅法观察数据之间的关联，推测该数据是否适合进⾏回归分析/线性回归分析。

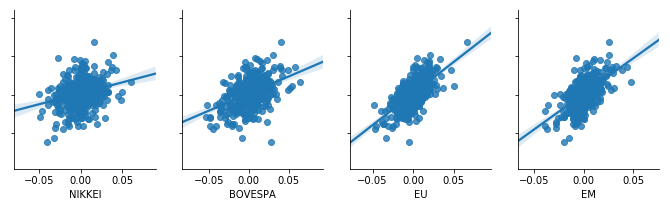
2. 使⽤回归分析的⽅法（如线性回归）进⾏回归分析，并与你的推测结果进⾏对⽐和思考。（实验过程中请注意评价指标、训练误差、泛化误差、测试数据划分等内容，并记录在实验报告中。）

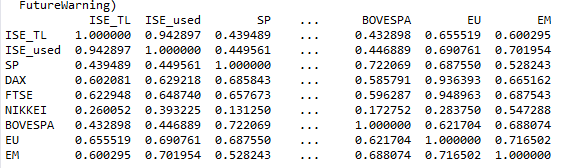
3. （附加题）尝试使⽤降维前后的数据表示分别进⾏回归，并⽐较回归的结果，思考降维该回归任务的影响。

# 解答及实验过程记录：

1. 可视化数据分析：





皮尔相关系数：

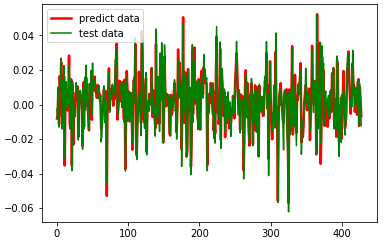
**问题1：**

通过对进行处理，获得了后7列数据分别与第一列数据的关系，对数据进行可视化观察和相关性分析。（利用seaborn库进行作图，并且绘制了拟合曲线和95%置信区间），发现数据量较大，数据点对于除了NIKKEI和BOVESPA两列数据外的数据，其他数据都可以拟合出一条直线，并且数据在直线的投影分布离散型较大，而且BOVESPA和NIKKEI与ISE\_TL的相关性较小，因此，我认为该数据集适合进行多元线性回归。

**问题2：**

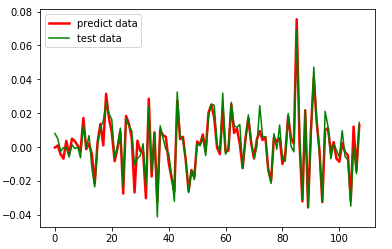
利用了skilearn工具对该数据进行了多元线性回归分析。

1. 划分数据集：80%作为训练集，20%作为测试集。
2. 评估方法：MSE方法，将预测结果与训练集比较，利用如下公式（MSE）进行评估.
3. 训练误差：



MSE: 2.489414504745773e-05

1. 泛化误差：



MSE: 2.3850799709279053e-05