

# 线性回归实战

这次实战是基于最近为实验 Pentaho PMI 制作的一个简单机器学习 demo。目的是掌握 sklearn 包中线性回归模型的应用。附带的 ZIP 文件中包括原始数据文件 (currency\_exchange\_rates\_output.csv)。

任务：

1. 准备训练数据 a：利用前 5 天所有所有汇率的收市价预测明日人民币对美元收市价格 (CLOSE.USDCNY.FX)， $h(x)$ 。实际训练数据中需要加入 Y。用 Python 实现。
2. 准备训练数据 b：利用前 5 个 10 日均值（相邻，无重叠）作为输入 X，加入 Y，用 Python 实现。
3. 用 cross\_validation 以 test\_size=0.2 分割成训练集和测试集。
4. 利用 sklearn 中 Ridge 线性模型对两个训练数据集分别进行训练。从 0 开始，试验 3 个 alpha 值。
5. 利用 sklearn 中 LassoLars 线性模型对两个训练数据集分别进行训练。从 0 开始，试验 3 个 alpha 值。
6. 对于 3，4 训练出的共 12 个模型，plot 实际测试集 Y 和预测  $Y'=h(X)$ ，给出模型的 score()值，最大和最小的 5 个 coef\_ 和 对应的 column name (如 LOW.USDCNY.FX)
7. 选做：利用 sklearn.preprocessing.PolynomialFeatures 生成数据 b（步骤二）包含 2 次项的新输入数据，并用 LassoLars 建模。给出 6.中要求的结果。