

1. 划分训练集、验证集和测试集

对已有数据进行训练集、验证集和测试集的划分目的在于交叉验证调参并做早停处理提高模型泛化能力，同时通过测试集评估模型鲁棒性。

划分规则：将末尾100期作为测试集，前1857期依据10折交叉验证划分训练集和测试集，或者依据ARMA模型进行整体训练集处理。

2. 确立baseline思路

由于时间序列数据滞后性性质，模型在训练学习过程中倾向于做序列平移直接预测，虽然同样是优化MSE，但这并不是我们所希望看到的，因此我们将baseline定为直接用 X_{t-1} 预测 X_t ，如果模型不能有效的超过baseline，说明模型并没有提取到足够的信息超过随机预测。

$$\hat{x}_t = x_{t-1}$$

3. ARIMA (1, 1, 4) 模型尝试

对训练集1857期建立ARIMA模型后预测测试集100期与baseline对比（MSE）

预测期数	ARIMA (MSE)	baseline (MSE)
第1期	0.01006	0.01038
第2期	0.02515	0.02510
第3期	0.04214	0.04156
第4期	0.06323	0.06179
第5期	0.08707	0.084390

- 对于ARIMA模型在第一、二期时有超过baseline，但三、四、五期时ARIMA模型表现不如随机预测。

将ARIMA模型预测每个五期作为一个样本，测试这100个样本对比baseline的胜率。

ARIMA 胜率 0.48958

- 综上所述ARIMA模型思路比不过baseline，无论是误差对比还是胜率对比，都不如随机预测。
- 原因分析：
 - ARIMA模型本身是一个多输入（单输入）单输出结构，并不适合进行多输入多输出结构预测，在同时预测五期后，会在后面产生较大误差。
 - ARMA模型本质是前面信息的线性组合进行下一期预测，当预测多期时本质依然是前面线性的线性组合，也就是可以认为当预测第二期开始，预测值是信息的线性组合的线性组合，我们这里称之为复杂线性组合。

我们在模型学习时

$$x_{t-p}, x_{t-p+1}, \dots, x_{t-1} \rightarrow \hat{x}_t$$

$$\Rightarrow \hat{\beta}_{t-p}, \hat{\beta}_{t-p+1}, \dots, \hat{\beta}_{t-1}$$

但模型在预测时 $x_{t-p}, x_{t-p+1}, \dots, \hat{x}_{t-1} \rightarrow \hat{x}_t$ 那么 $\hat{\beta}_{t-p}, \hat{\beta}_{t-p+1}, \dots, \hat{\beta}_{t-1}$ 是否还是最优的估计那就能难说了。

类似于学习器通过狗的毛色分辨狗的品种，训练时一直如此，可直接预测时狗的毛色却是猜测的并不，是准确的，这样显然会降低学习器精确度。

4. 基模型思路

基于上诉思路，在此我们提出对五期预测进行分开建模，重新组织数据结构，采用时间滑窗整理数表。

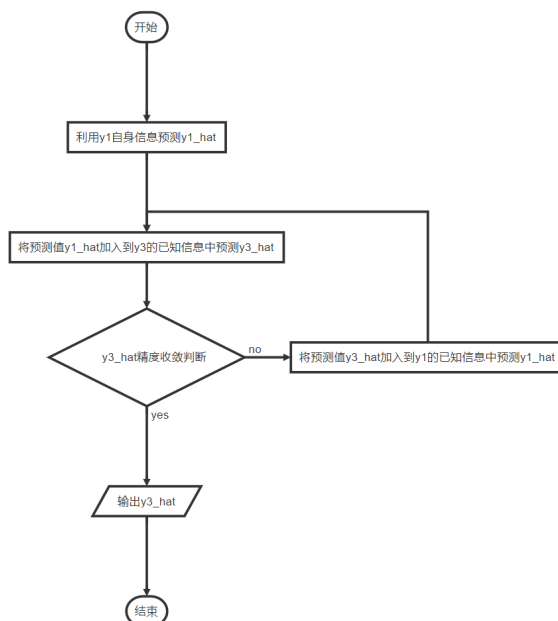
第一期	第二期	第三期	第四期	第五期	第一期
X_1	X_2	X_3	X_4	X_5	X_6
X_2	X_3	X_4	X_5	X_6	X_7
X_3	X_4	X_5	X_6	X_7	X_8
...

利用相关信息问题

通过对 Y_3 和 Y_1 的分析，二者时存在非常强烈的相关关系，这是除了自生信息外，我们可以利用的有效信息。

1. 迭代思路尝试

- 流程



- 思路说明：

Y_3 和 Y_1 的相关性特征可以辅助两个学习器相互提升，举个简单的例子： Y_3 和 Y_1 本身具有强相关性，而 \hat{Y}_3 值如果与 \hat{Y}_1 没有呈现足够的相关性，那么二者中较为优异的模型，会辅助另一个模型进行预测，比如用自身信息 \hat{Y}_3 精确度较高，而 \hat{Y}_1 精确度较低，那么那么由于在同一期时相关性的关系，较高精度的 \hat{Y}_3 会提供估计值从而辅助 \hat{Y}_1 进行估计，从而达到模型优化。

那为什么不将 Y_1 的信息直接加入到 Y_3 中而是采用将估计值传入呢？因为传入估计值是对信息的有效降维。

- 实验结果：

训练结果并不如人意，原因在于：两个学习器的“成绩”都非常差，两个“差生”相互促进，只会越来越差。真实结果在第二轮之后不再收敛。

2. 预测第一期时加入相关组的滞后信息

对于利用相关信息的尝试，退而求其次，采用相关性最为强的一期加入特征进行训练。

- 思路说明：

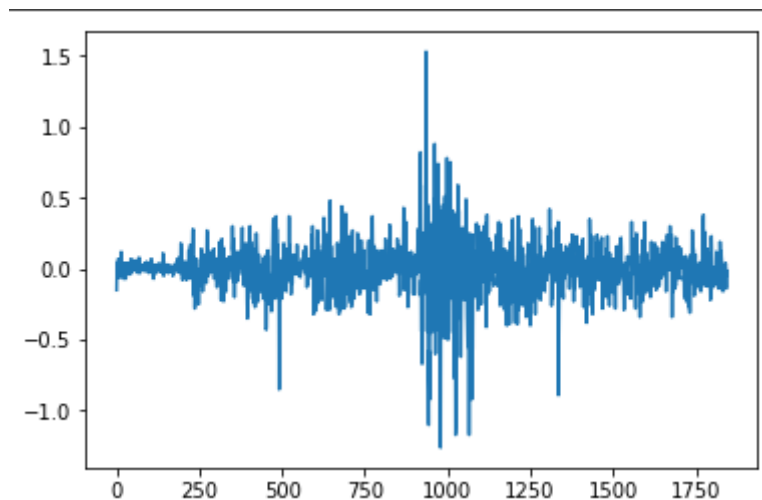
降维传入信息的方式不好，原因在于数据的噪声十分严重，因此在这种情况下，退而求其次加入一定的信息是可取的。

- 实验结果：

效果有一定提升。

去波动性

所有的学习器都有一个假设，那就是训练集和测试集数据分布的一致性，有时序图我们可以明显的观测到在index = 0~200 和 800 ~ 1150左右的行为逻辑与大部分情况不一致，因此，建模时因将其去除。



- 实验结果：

去除后有显著提升

对不同模型的测试

- 使用不同基模型进行训练LSTM、SVM、linear_model 筛选。

- 实验结果：

几种模型的结果上限无显著差异，无论是非线性还是线性的，有无时间序列结构的都是如此。数据决定上限、模型只能逼近上限。

对不同模型进行调参

- 对基模型进行调参选择滑窗周期。
- 实验结果：
- Y_3

预测期数	时间窗口期数	SVM(MSE)	baseline
1	23	0.00913	0.01038
2	22	0.02293	0.02510
3	27	0.03761	0.04156
4	22	0.05698	0.06179
5	23	0.07734	0.08439

- Y_1

预测期数	时间窗口期数	SVM(MSE)	baseline
1	22	0.006701	0.007337
2	18	0.017447	0.018509
3	20	0.028650	0.030685
4	19	0.042084	0.044638
5	23	0.056430	0.058889

模型对比

- Y_1

五期总MSE	SVM 0.0302627	ARIMA 0.032907
五期胜率	0.52083	0.479166

- Y_3

五期总MSE	SVM 0.0403434	ARIMA 0.043955
五期胜率	0.5625	0.4375

- 实验结论：
对于金融序列评价标准不应该使用五期进行评价，而是应该有足够的样本进行测试，并动态输出进行评价。

5. 预测结果

- Y_1

期数	一阶差分预测	预测值
1	[0.01112]	5.12112
2	[0.01430638]	5.12430638
3	[0.03514156]	5.14514156
4	[0.02856328]	5.13856328
5	[0.05645678]	5.166456780

- Y_3

期数	一阶差分预测	预测值
1	[0.01103134]	5.69103134
2	[0.02153156]	5.70153156
3	[0.02252713]	5.70252713
4	[0.03437449]	5.71437449
5	[0.05550684]	5.73550684

6. [github仓库链接](#)

<https://github.com/feng5shang/TimeSeries>