

机器学习和人工智能领域中一些在应用上较为成熟的算法，很多都是以分类为目标的，同时一些算法也可以用来进行非线性回归之类的拟合和预测，因此本质上就和量化交易策略的框架具有相容性。但是即便如此，作者仍然不太认同行业中存在的拿来主义，一些量化交易策略的研究人员在没有搞清楚机器学习算法内在逻辑的情况下，简单的将现成的工具包直接套用到交易数据之上，其实是不可取的。例如使用支持向量机做分类，至少也应该先弄清楚使用支持向量来判断类别所具有的优点和缺点，是否适用于自己的策略和实际情况。况且一部分机器学习算法在工业界的应用是有大规模的样本数据作为训练基础的，直接用来构造量化交易策略时，会有多大的可能性造成过度拟合，都是应该提前考虑的问题。

以这一节的均线趋势策略为例，也就是在固定了均线趋势策略的基本框架之后，对于短期均线长度和长期均线长度两个参数，在全部模拟交易样本上考虑不同参数组合所带来的收益情况，选取收益最好的一个参数组合为优化后的参数组合。随后考察这一个参数组合的均线趋势策略在相同样本上的模拟交易表现，进而判断均线趋势策略的优劣。

通过收益最优原则选择出来的短期均线为 3 日均线、长期均线为 21 日均线的均线趋势策略，所处的区域应该算是比较稳健的

最后通过收益最优原则选择出来的短期均线为 3 日均线、长期均线为 5 日均线的均线反转策略，至少在样本上具有一定的收益表现。过计算，整个交易过程中平均的年度最大回撤为 15.19%，与 14.64% 的年化收益率接近一致。仅仅和上一节中的结果进行对比而言，均线反转策略在收益和风险两个方面都逊于均线趋势策略。

$$r_t = \sum_{i=1}^N b_i r_{t-i}$$

自回归策略

过计算上面的回归式就可以得到对 t 时刻收益率的预测值。当 t 时刻的预期收益率为正时，则看多 t 时刻的交易资产；当 t 时刻的预期收益率为负时，则看空 t 时刻的交易资产。

多因子选股策略方面，本书也选用一个较为成熟的策略框架来进行介绍。首先假定已经选出了 N 个有效因子，分别用 f_1 到 f_N 标记，然后针对每一支股票，用这 N 个因子的值来预测下一期或者下一段时间的收益率。这里选用简单的线性回归来完成预测工作，如下式所示：

$$r_t = a + \sum_{n=1}^N b_n f_{n,t-1}$$

其中 r_t 是时刻 t 的股票收益率， $f_{n,t-1}$ 即为 $t-1$ 时刻下第 n 个因子的大小， a 和 b_n 是回归式中的系数。进行交易决策的时间点为 t 时刻初、 $t-1$ 时刻末，因此回归式左边为预测值，回归式右边的所有成分则都是决策点下的已知信息。在预测出每一支股票在时刻 t 的收益率之后，按照收益预期值从大到小进行排序，然后选取排在前列的股票作为当前可以建仓的股票。

需要特别说明的是，在某些量化交易策略的相关资料当中，会把对于不同股票而言取值一致的**回归系数 b** 称之为**风险因子**，而将具体的**股票特征值 f** 称之为各支股票在因子上的**溢价**。这主要是因为学术界在套利定价理论等研究的基础上，形成了一种约定俗成的叫法，其中风险因子对于所有资产应该保持一致，而因子溢价则各有不同。不过在量化选股策略中，对比本节所使用的称谓，这种叫法以及其他一些叫法并不是非常直观，因此不予以使用。如果读者在阅读其他资料时碰到不一样的名称，只需对号入座弄清准确含义即可。

本书中所述的多因子策略方法基于**线性回归模型**，因此预期收益和因子大小之间符合线性关系是最优状态，当两者的关系呈现出如图 5.5 中所示的非线性关系时，在线性回归步骤之前对因子或收益数据做合适的预处理。在进行线性回归之前，需要对流通市值进行对数化处理，其目的是为了更好的反映出月度收益率和流通市值之间的线性关系

反转因子是另外一个比较重要的选股因子，在实际的金融市场上，股票往往对当前的信息存在过度反应的现象，因此在短期内具有回复到正常状态的动力，从股价上看就是一种反转的行为。简单来说，就是上个月收益率越低的股票，下个月的收益情况往往越好。

但是值得注意的是，其在样本期的末尾也经历了一段失效时间，并且与市值因子的失效时间几乎重合。因此在那几个月中，使用因子选股策略的基金普遍产生亏损并不简单是市值因子失效的问题，市场特征发生变化导致若干个有效因子同时出现问题才是深层次的原因。

有效因子越多、因子配置越均衡，多因子选股策略受到单一因子失效影响的程度也就会越小，这其实和投资组合配置的情况是类似的，实际交易中研究人员还是应该尽可能的多发掘出一些相关性较小的选股因子，从而增强多因子选股策略的稳健性。

样本外检验来进行过度拟合的判断的

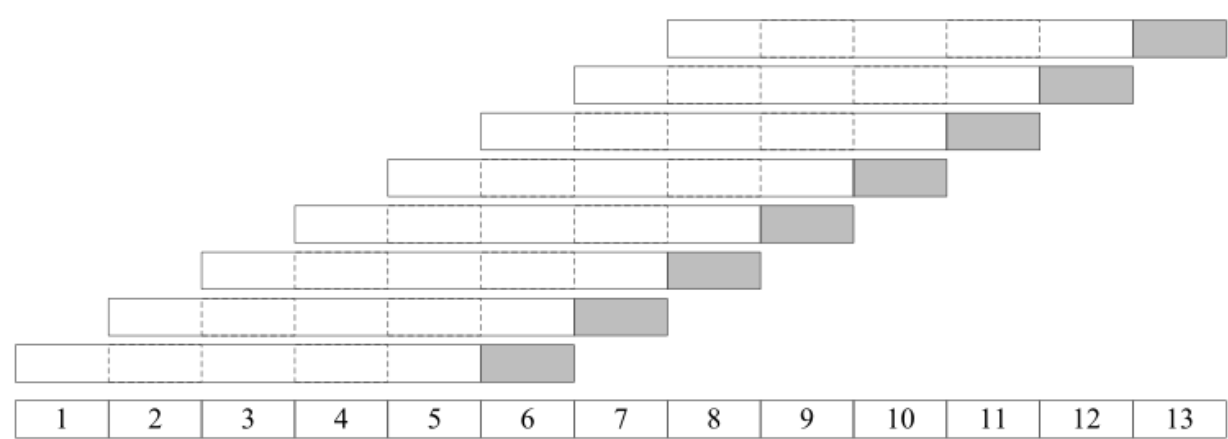


图 6.2 推进分析示例

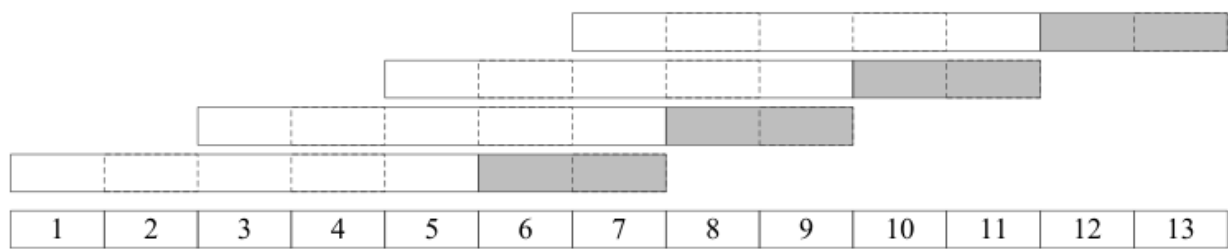


图 6.3 计算紧张时改动的推进分析示例

使用周期 1 到周期 5 的数据优化得到的量化交易策略，会由于交易资产和市场数据的变化而对周期 7 的数

据适应能力变差，周期 9、11、13 下的检验情况也类似。对于相隔较远的交易决策而言，策略的适应能力同样会降低，进而弱化盈利能力

同期对比来看，均线趋势策略的简单优化下，2011 年 12 月 30 日收盘到 2015 年 6 月 30 日收盘时的收益率为 104.69%，是这里推进分析（减小过拟合）的 19.96% 的 5 倍多。

均线 混合策略的推进 分析预测：

如果最优均线趋势策略的收益高于最优均线反转策略的收益，那么选取最优参数下的均线趋势策略作为最终的优化结果，反之则选取最优参数下的均线反转策略作为最终的优化结果。

原因在于优化过程综合选择了均线趋势策略和均线反转策略这两个策略中更优异的一个，因此必然会比两个策略分开讨论的情况都要好。而检验期下的收益率水平有所提升，就是一个对于量化交易策略而言具有实际意义的结果了，其说明至少在回溯测试的样本下，策略的混合改进确实带来了一定的盈利能力。

净值在相当长的一段时间内低于初始净值且没有明显变动。通过这些研究结果来判断，这三种均线策略均不适合应用于实际交易当中。

自回归策略不仅通过了推进分析的初步筛查，而且在一定程度上展现出了逐行优化更贴近实际交易的优势。

鉴于简单优化中存在的未来信息偏误，推进分析的结果已经在一定程度上说明了多因子选股策略的有效性，同时也体现出了逐行优化更贴近交易点的优势