

风险中性的深度学习选股策略

文巧钧 S0260517070001

邮箱: wenqiaojun@gf.com.cn

安宁宁 S0260512020003

邮箱: anningning@gf.com.cn

广发证券金融工程

2018年6月3日

01

I

组合表现

>

02

II

风险中性的机器学习模型

>

03

III

策略与实证

>

04

IV

总结

>



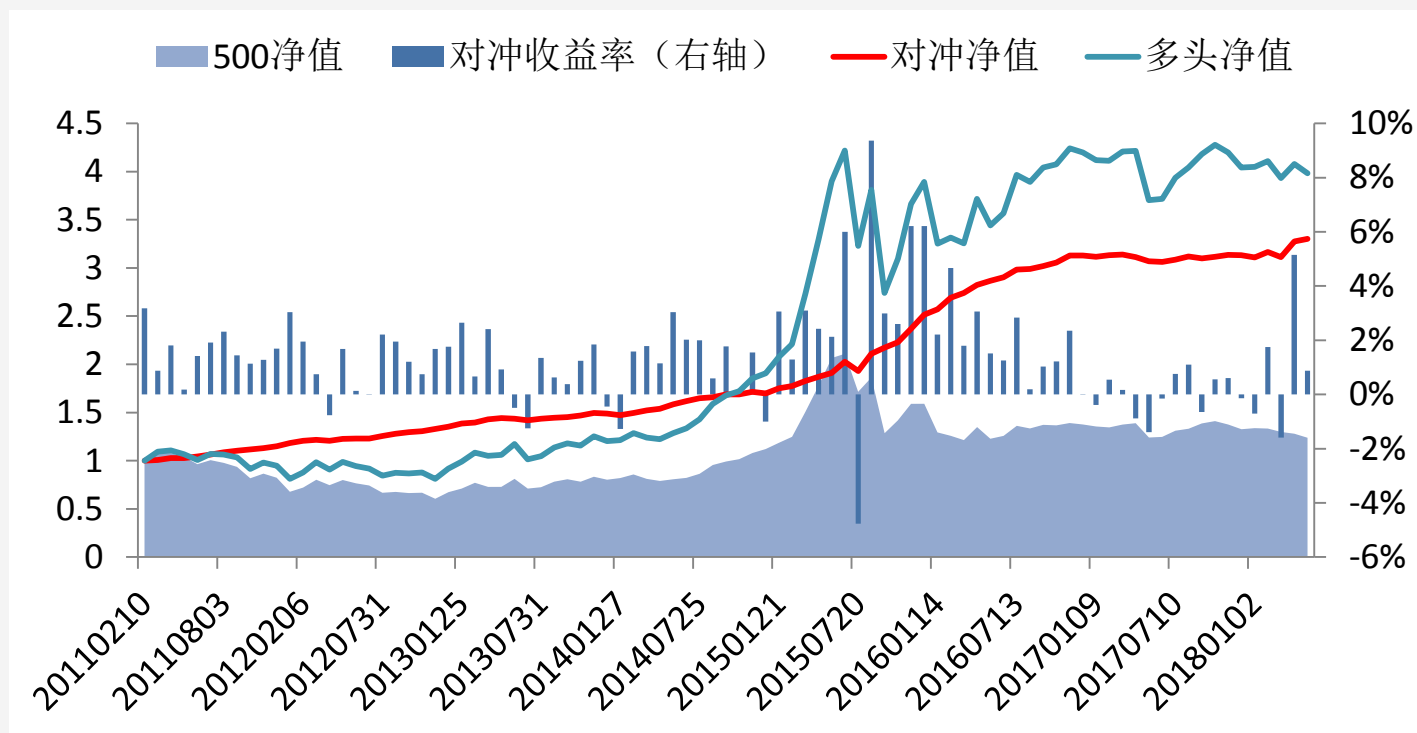
01

| 问题背景 |



深度学习选股策略表现：全市场选股，中证500指数对冲

2011年以来，年化收益率18.5%，最大回撤 -4.77%，月度胜率为 81.8%



新的思考

1

如何减小组合的风险暴露？

组合优化？

2

如果当前市场与历史数据差别较大时，如何使得机器学习模型有效？

模型更新？

3

能否构建在不同风格市场上表现稳定的机器学习选股模型？

？

4

如何减小机器学习交易策略的同质化问题？

？



01

02

03

04

02

| 风险中性的机器学习选股模型 |

>

深度学习预测模型

模型训练：通过训练样本，确定模型结构，优化模型参数

预测输出 Y 的维度：3

输入特征 X 的维度：156（128个因子+28个行业）

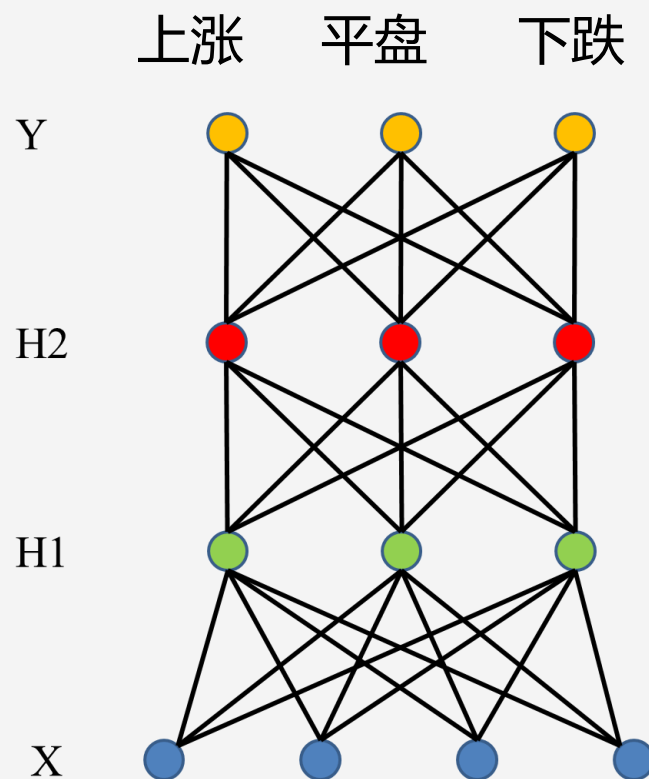
通过网格搜索获取最优的模型结构

选取模型结构为：

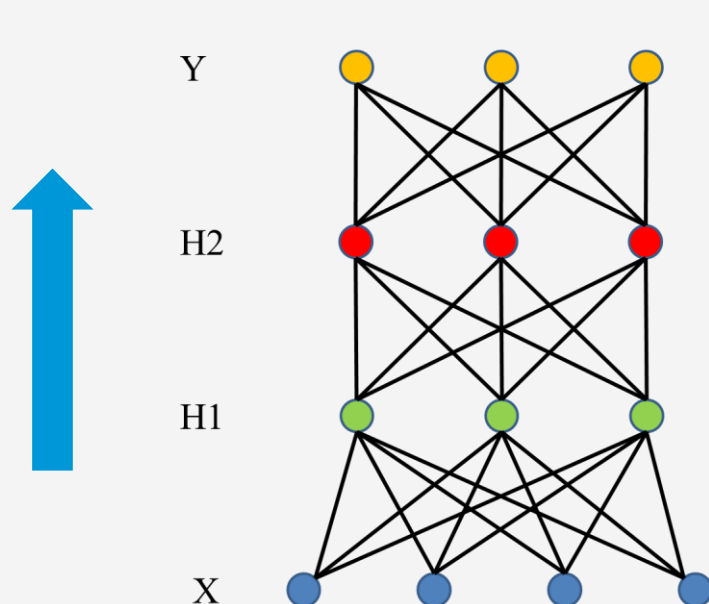
156（输入层）-512-200-200-200-128-3（输出层）

即一共包含5个隐层

隐层节点数依次为：512（隐层1）、200（隐层2）、
200（隐层3）、200（隐层4）、128（隐层5）



深度学习预测模型（以2个隐层的网络为例）



$$y_k = \sigma_o \left\{ \sum_{j=1}^{N_2} (w_{kj}^{(2)} h_j^{(2)} + w_{k0}^{(2)}) \right\}$$

$$h_j^{(2)} = \sigma_h \left\{ \sum_{i=1}^{N_1} (w_{ji}^{(1)} h_i^{(1)} + w_{j0}^{(1)}) \right\}$$

$$h_j^{(1)} = \sigma_h \left\{ \sum_{i=1}^{N_x} (w_{ji}^{(0)} x_i + w_{j0}^{(0)}) \right\}$$

隐层激活函数： σ_h

输出层激活函数： σ_o ，对于分类问题，一般用Sigmoid函数或者Softmax函数

多分类问题
$$\sigma(\theta_i^T x) = \frac{e^{\theta_i^T x}}{\sum e^{\theta_i^T x}}$$

深度学习预测模型

待优化网络参数： \mathbf{w}

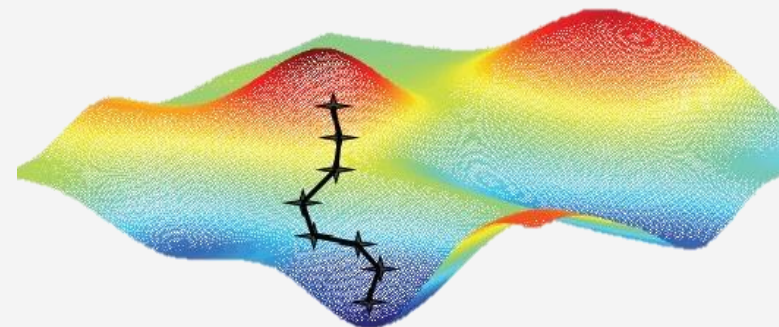
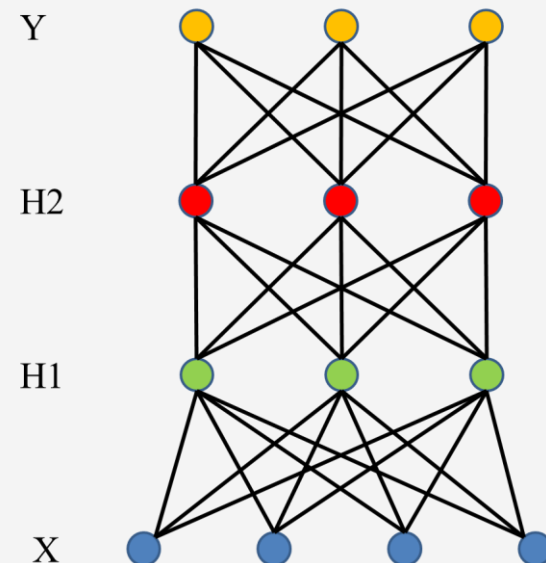
优化目标：最小化均方误差（MSE）

$$E(\mathbf{w}) = \frac{1}{NK} \sum_{n=1}^N \sum_{k=1}^K \left(\underset{\substack{\text{预测输出} \\ /}}{y_{nk}} - \underset{\substack{\text{实际标签} \\ \backslash}}{t_{nk}} \right)^2$$

参数优化方法：BP算法，迷你批量梯度下降算法

➤ 集成了梯度下降法和随机梯度下降法的特点

$$w_{ij}^{(n)} := w_{ij}^{(n-1)} - \alpha'' \frac{\partial}{\partial w_{ij}} \sum_{n_k \in \text{Batch}(n)} E_{n_k}(\mathbf{w}^{(n-1)})$$



提高深层神经网络选股性能的主要方法：

- 1、采用relu等激活函数
- 2、将优化目标函数MSE改成交叉熵
- 3、Batch normalization技术
- 4、Dropout技术

深度学习预测模型

采用Keras作为机器学习平台：

- 可以选择Tensorflow、CNTK、Theano（目前已经停止更新）作为后端
- 目前已经支持多GPU
- 显卡选择：Nvidia GTX Titan XP, GTX 1080Ti, GTX Titan, GTX 1080, GTX 1070, GTX 1060

风险中性的选股策略

机器学习多因子选股模型：从股票特征中提取信息，对股票未来的收益进行预测，选出能够产生超额收益的股票组合。

多因子选股模型：
挑选出未来一期能够产生超额收益的股票



风险中性的多因子选股模型：
挑选出未来一期，**在剥离风险因子影响后**，能够产生的超额收益的股票

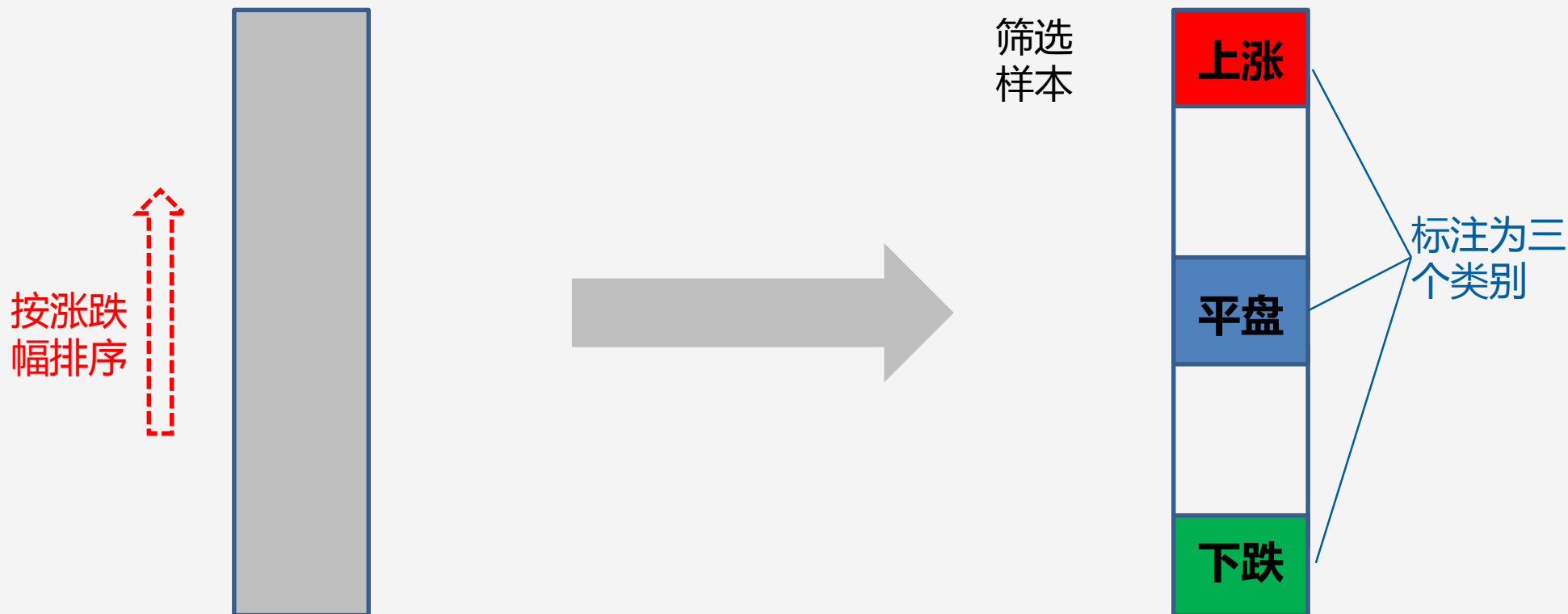
如何实现上述目标？

——针对具体的应用场景，构建适当的机器学习预测模型。



风险中性的选股策略

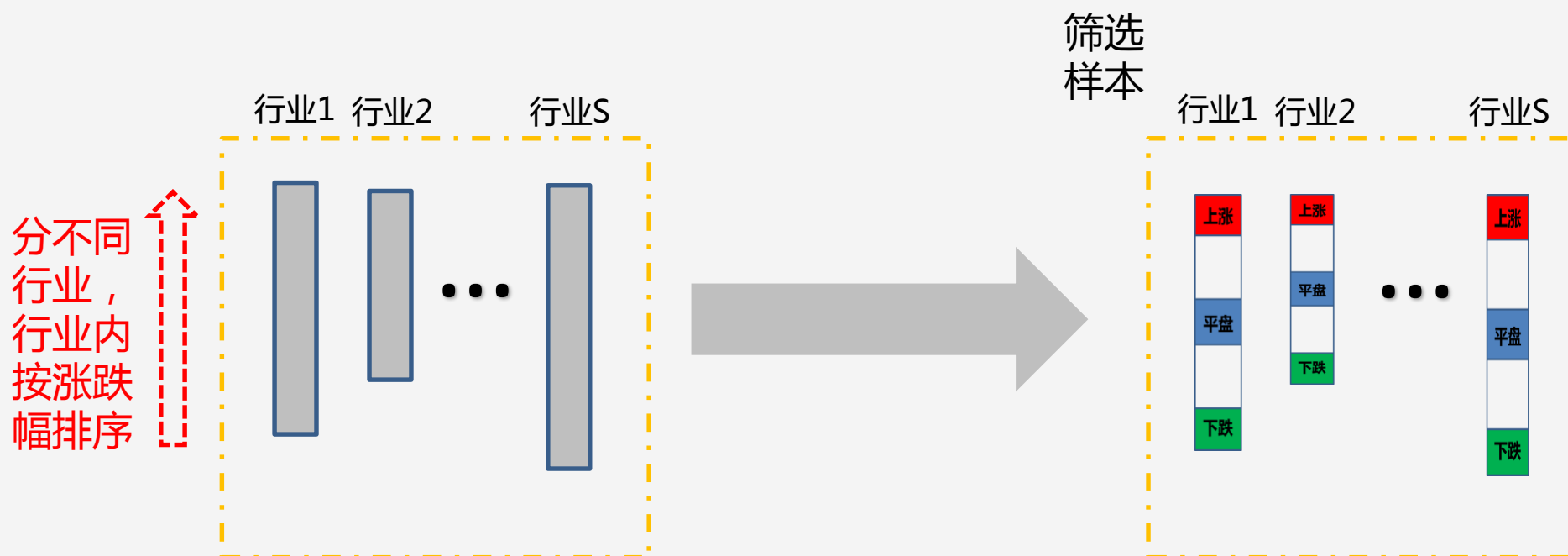
普通的样本标注方法



训练时，根据未来一期的股票涨跌幅来给样本贴“标签”：上涨、下跌、平盘
同一时刻，按照涨跌幅进行排序，确定样本的输出标签

风险中性的选股策略

行业中性的样本标注：寻找不同行业内能够跑出超额收益的股票

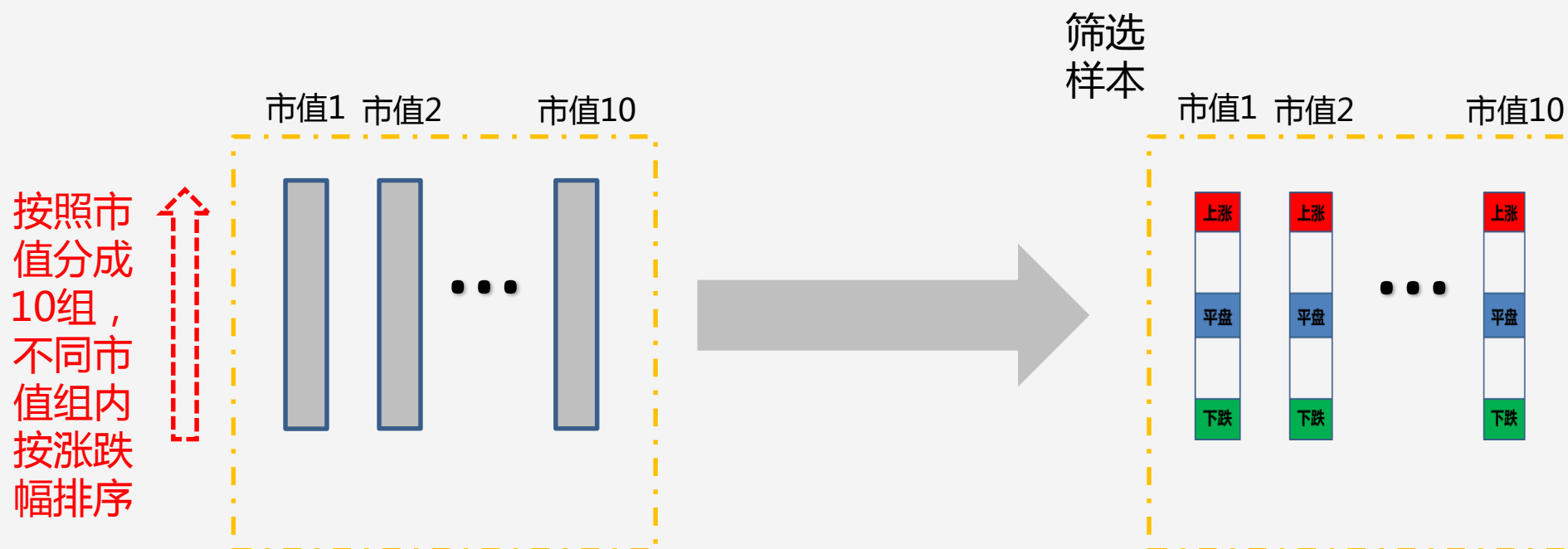


训练时，根据未来一期的股票涨跌幅来给样本贴“标签”：上涨、下跌、平盘

同一时刻，按照涨跌幅进行排序，确定样本的输出标签

风险中性的选股策略

市值中性的样本标注：寻找不同市值区间内能够跑出超额收益的股票



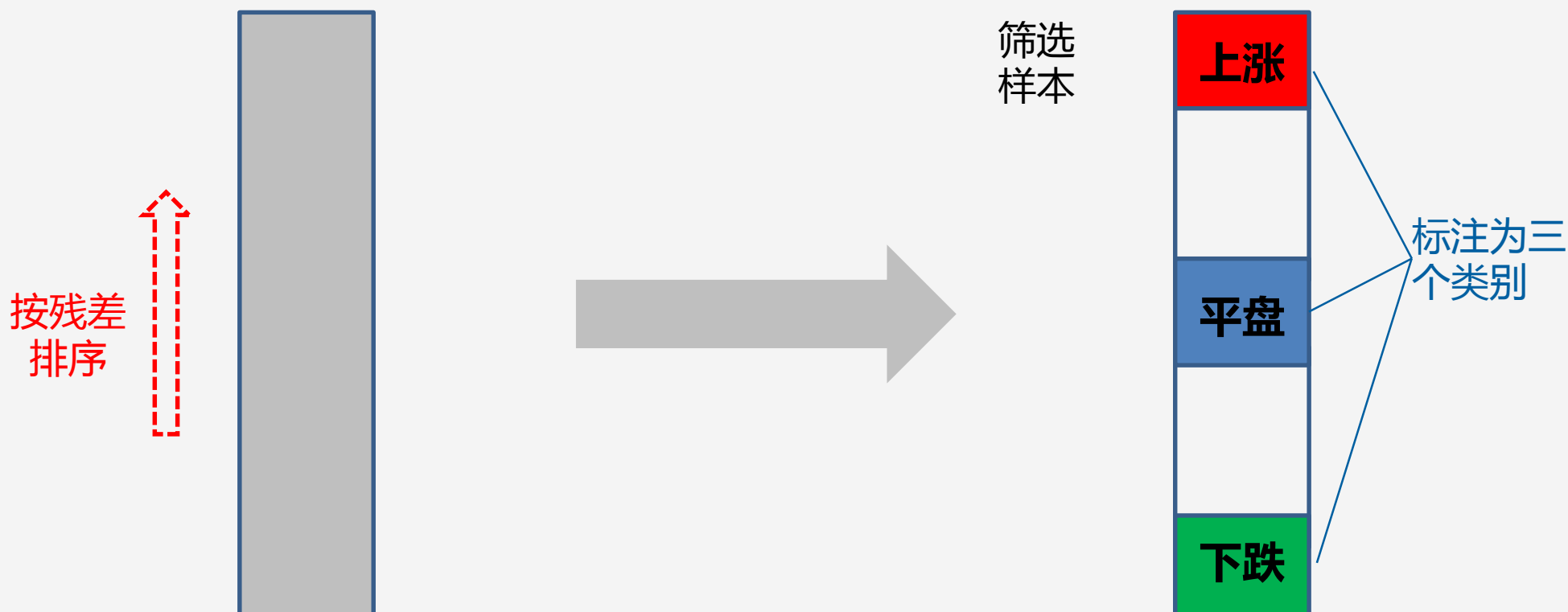
训练时，根据未来一期的股票涨跌幅来给样本贴“标签”：上涨、下跌、平盘
同一时刻，按照涨跌幅进行排序，确定样本的输出标签

风险中性的选股策略

风险中性的样本标注：更一般的方法

未来一期股票收益率对风险因子做线性回归，获得残差，按残差进行样本标注

$$r_i^t = X_{i1}^t f_1^t + X_{i2}^t f_2^t + \cdots + X_{iK}^t f_K^t + \boxed{\epsilon_i^t}$$





03

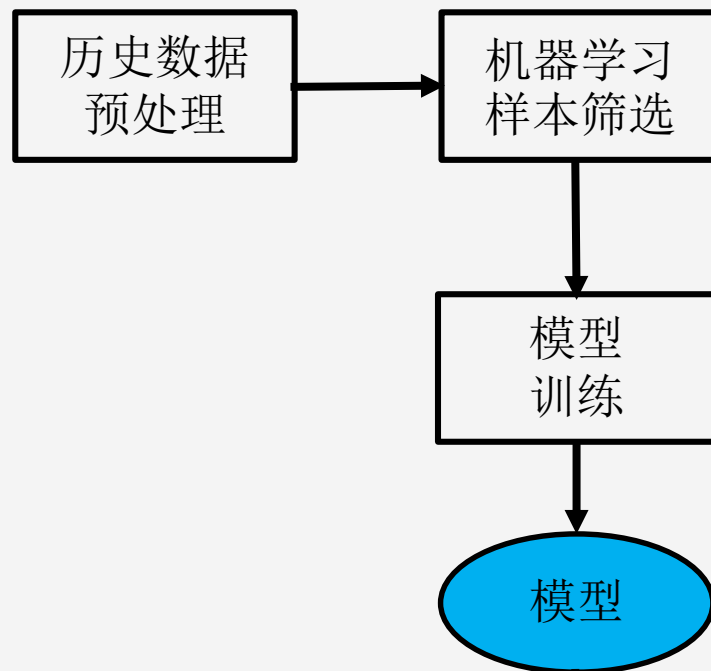
| 策略与实证分析 |



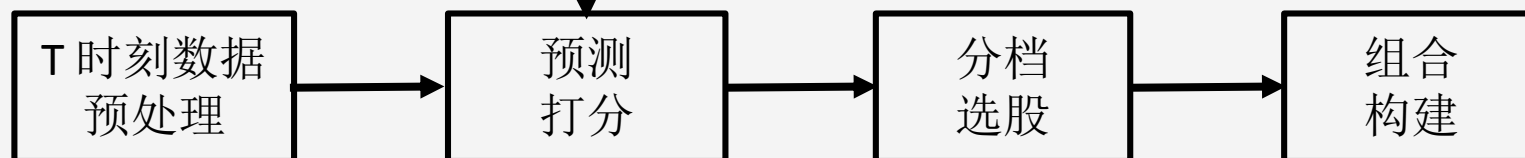
策略与实证

策略流程

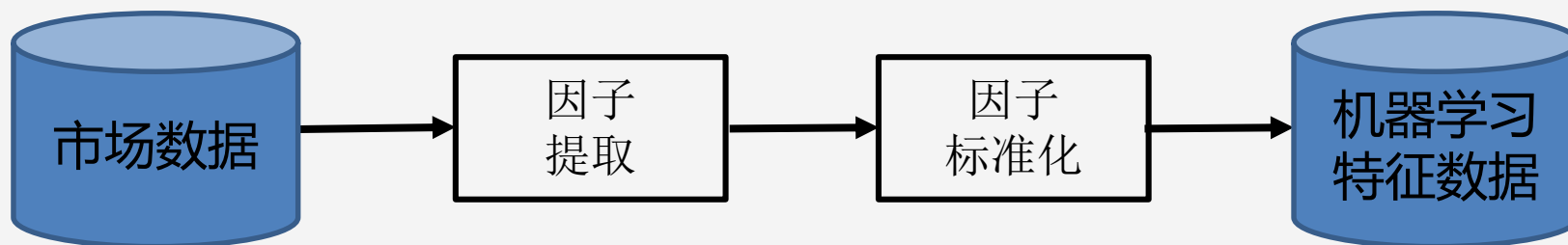
模型训练



选股交易



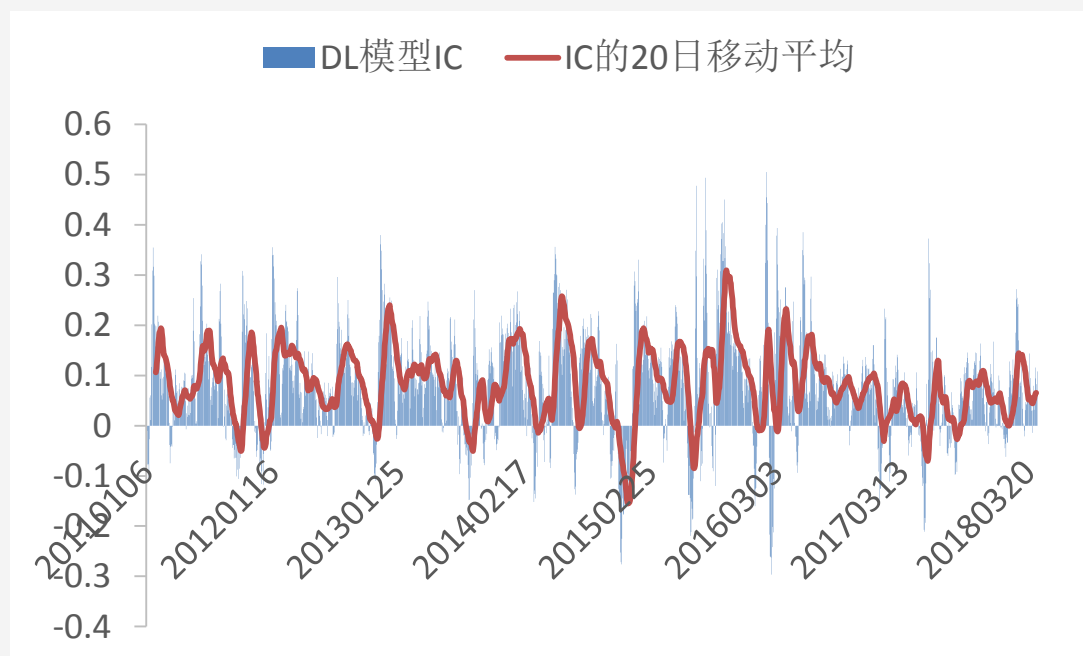
数据预处理



因子标准化：

- 1、异常值、缺失值处理
- 2、极值压边界处理
- 3、沿时间方向的因子标准化
- 4、沿截面的因子标准化
- 5、按照机器学习模型来调整因子分布

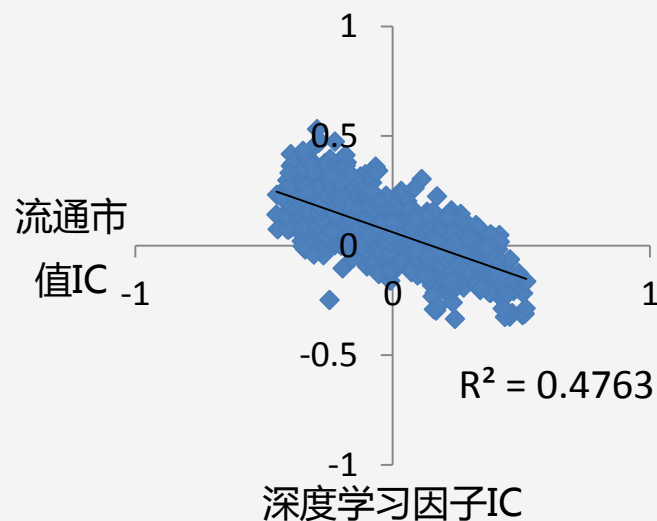
风险中性的深度学习选股模型的IC



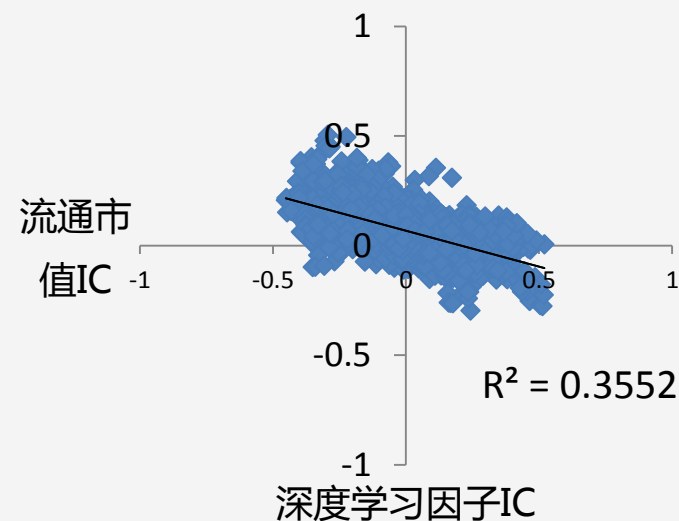
IC的平均值为0.082，标准差为0.108.

深度学习因子与流通市值的相关性：IC序列的相关性

普通的深度学习选股策略



风险中性的深度学习选股策略



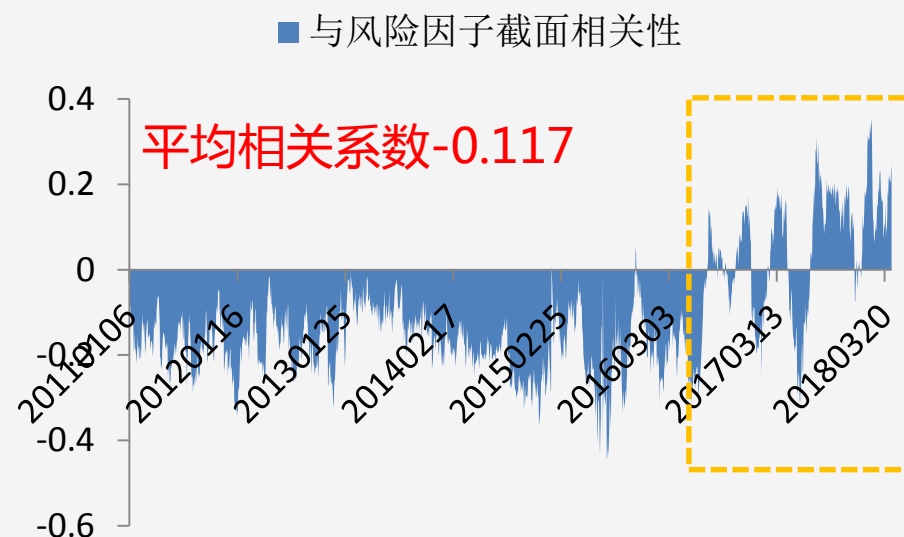
经过行业和市值中性化之后，深度学习因子更少受到市值因子的影响。

深度学习因子与流通市值的相关性：截面相关性

普通的深度学习选股策略



风险中性的深度学习选股策略



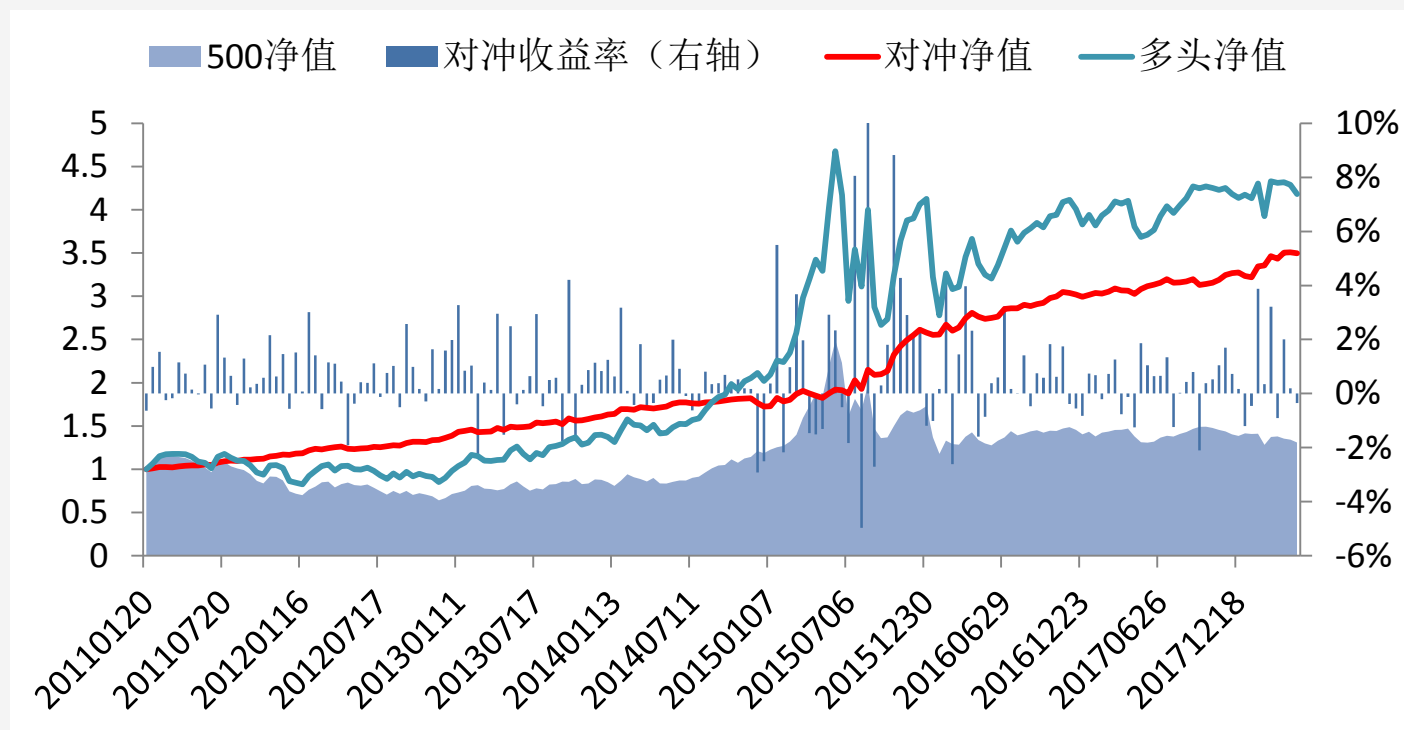
经过行业和市值中性化之后，深度学习因子更少受到流通市值因子的影响。

回测参数设置

- **调仓周期**：10个交易日
- **股票池**：中证500成份股，剔除ST股票，剔除交易日停牌和涨停、跌停的股票
- **超配组合**：调仓时分10档，等权买入深度学习模型打分最高的一档
- **对冲基准**：中证500指数
- **原始因子数据**：估值因子、规模因子、反转因子、流动性因子、波动性因子、技术指标，共计128个因子，以及28个行业0-1变量
- **风险因子**：行业、流通市值
- **深度学习模型训练**：每半年滚动更新模型，采用最近4年的样本作为训练集
- **策略回测**：2011年1月-2018年4月30日
- **交易成本**：千分之三

普通的深度学习选股策略

2011年以来，年化收益率19.71%，最大回撤 -5.35%，胜率为 69.5%，信息比2.47



普通的深度学习选股策略

2011年以来，年化收益率19.71%，最大回撤 -5.35%，胜率为 69.5%，信息比2.47

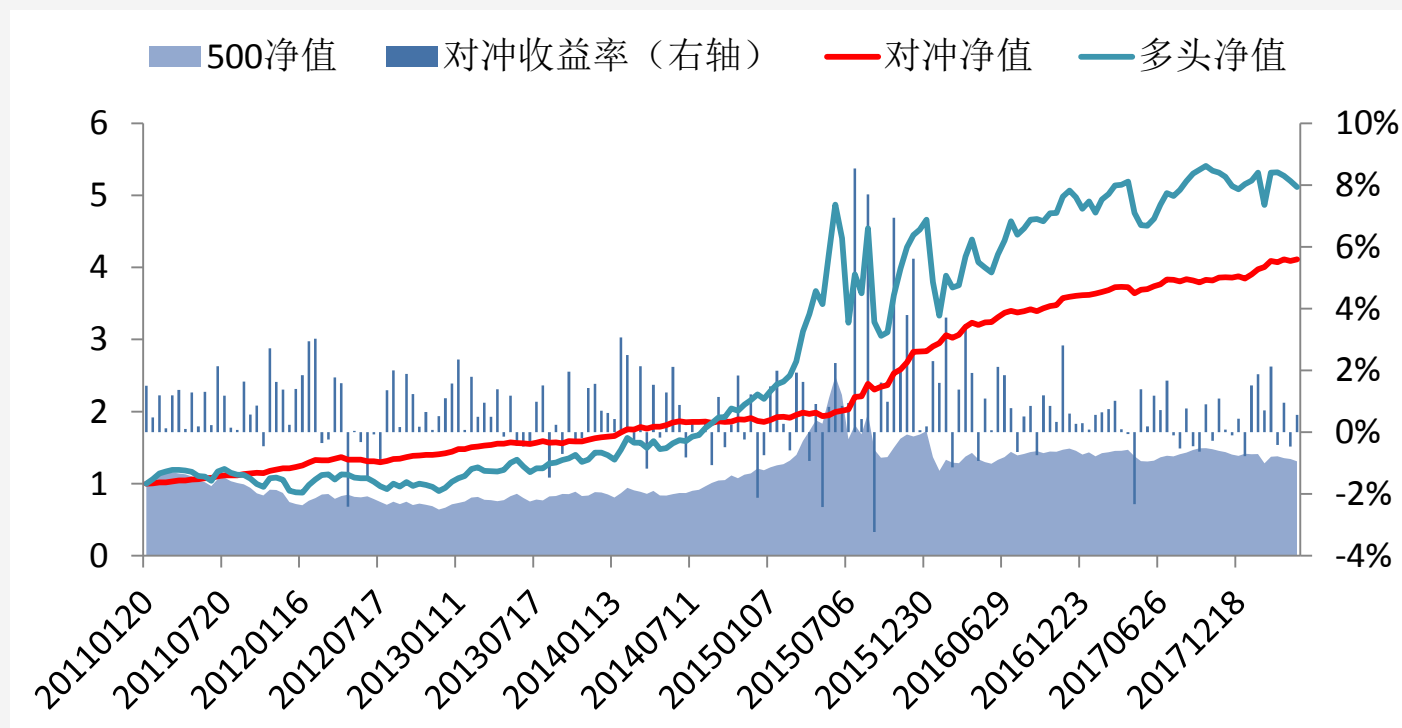
年份	累积对冲收益	对冲收益（年化）	最大回撤	多头累积收益	基准累积收益	换手率	信息比
2011*	18.32%	19.12%	-0.56%	-14.06%	-27.04%	17.79	3.64
2012	18.77%	18.77%	-2.28%	18.63%	0.28%	17.77	2.94
2013	16.73%	16.73%	-2.25%	36.29%	16.89%	17.58	2.33
2014	5.82%	5.82%	-5.35%	46.89%	39.01%	17.70	1.07
2015	48.08%	48.08%	-4.97%	98.30%	43.12%	19.92	2.96
2016	16.46%	16.46%	-2.62%	-5.66%	-17.78%	18.05	1.99
2017	8.31%	8.31%	-2.13%	8.13%	-0.20%	17.63	1.42
2018**	8.13%	24.39%	-1.66%	1.26%	-6.24%	5.09	3.44

*：2011年前10个交易日数据含有样本内信息，不进行选股

**：截至2018年4月30日

风险中性的深度学习选股策略

2011年以来，年化收益率21.95%，最大回撤 -5.03%，胜率为 74.6%，信息比2.92



风险中性的深度学习选股策略

2011年以来，年化收益率21.95%，最大回撤 -5.03%，胜率为 74.6%，信息比2.92

年份	累积对冲收益	对冲收益（年化）	最大回撤	多头累积收益	基准累积收益	换手率	信息比
2011*	23.75%	24.78%	-0.46%	-9.85%	-27.04%	17.73	4.59
2012	17.54%	17.54%	-5.03%	17.46%	0.28%	17.77	2.89
2013	13.65%	13.65%	-1.94%	32.63%	16.89%	17.43	2.04
2014	13.68%	13.68%	-2.85%	57.41%	39.01%	16.83	2.38
2015	52.02%	52.02%	-3.23%	108.99%	43.12%	19.16	3.59
2016	26.28%	26.28%	-1.14%	3.66%	-17.78%	17.63	3.17
2017	6.93%	6.93%	-2.39%	6.58%	-0.20%	17.41	1.25
2018**	6.90%	20.69%	-0.78%	0.21%	-6.24%	5.19	3.42

*：2011年前10个交易日数据含有样本内信息，不进行选股

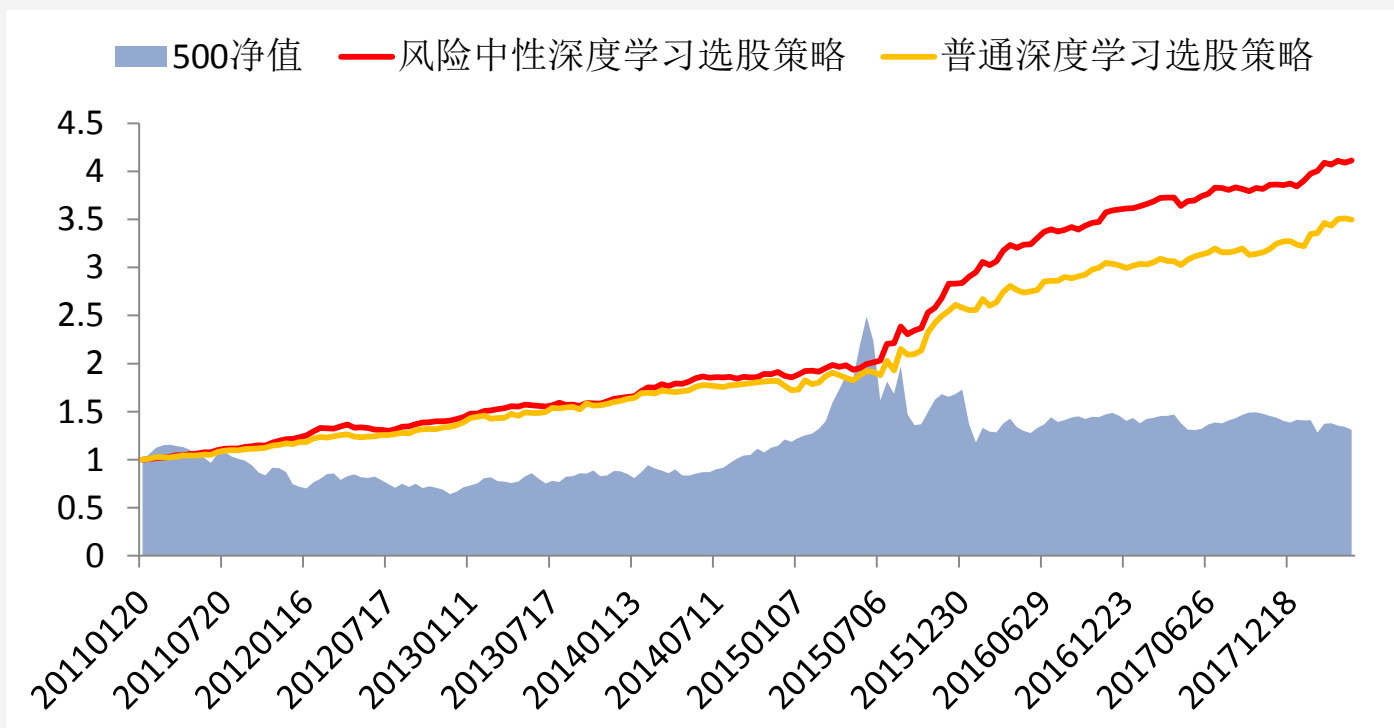
**：截至2018年4月30日

风险中性深度学习选股策略

年化收益率21.95%，最大回撤 -5.03%，胜率为 74.6%，信息比2.92

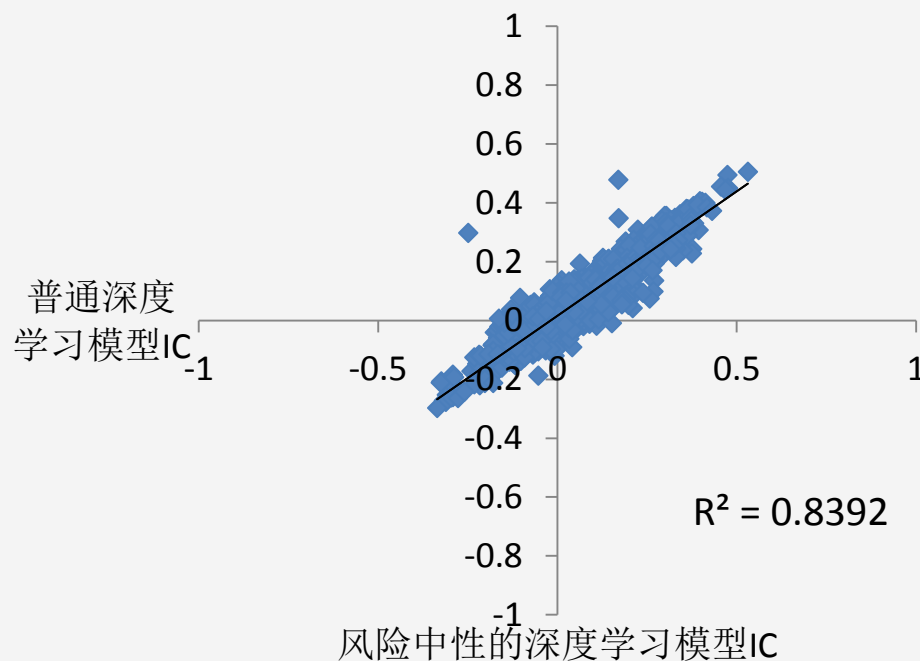
普通深度学习选股策略

年化收益率19.71%，最大回撤 -5.35%，胜率为 69.5%，信息比2.47



机器学习策略表现的同质性：不同策略的表现有一定的相关性

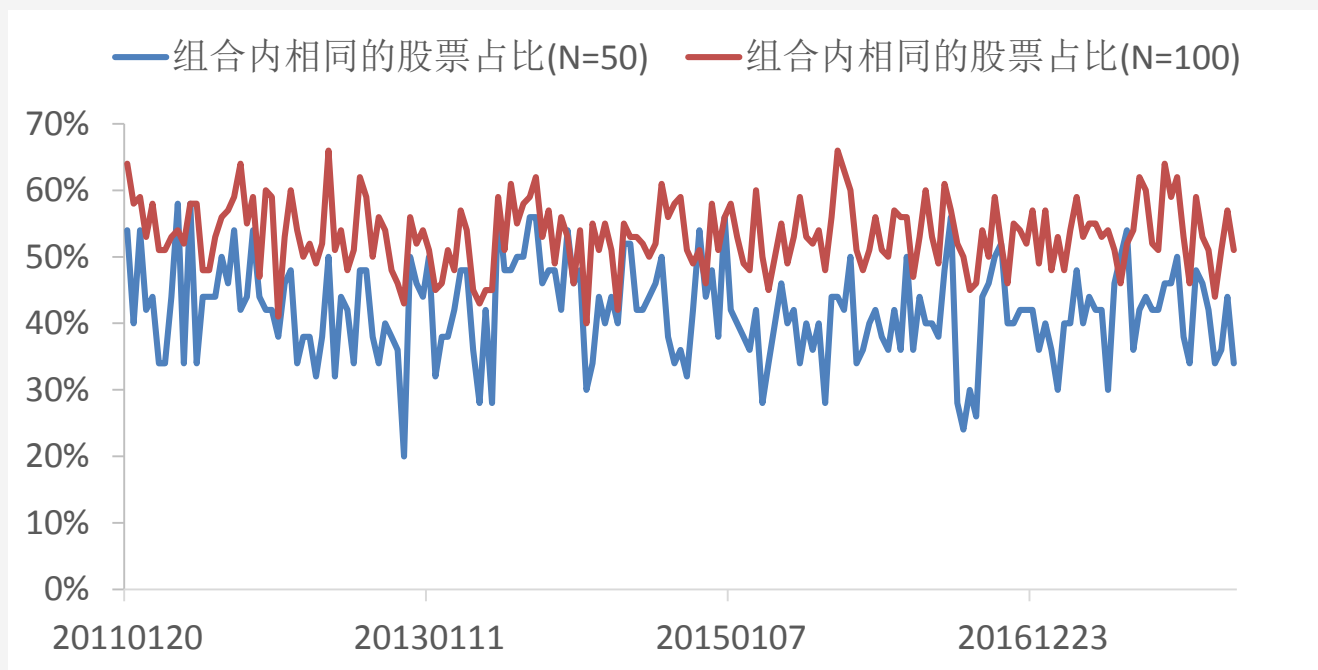
用普通深度学习模型的IC对风险中性深度学习模型的IC进行回归，R方为0.84。
说明两种策略的表现有较强的相关性。



机器学习策略每期组合的同质性：选取的组合有较大的差异

当组合规模N为50时，两种深度学习选股策略平均每期选到的股票有41.9%重合。（参考：两次独立的随机选股中，平均有10%重合。）

当组合规模N为100时，两种深度学习选股策略平均每期选到的股票有53.3%重合。（参考：两次独立的随机选股中，平均有20%重合。）





04

| 总结 |



总结

- 将风险因子中性化处理后，可以构建新的深度学习选股模型，该模型受风险因子的影响较小
- 2011年以来，中证500内选股对冲策略年化收益率21.95%，最大回撤 - 5.03%，胜率为 74.6%，信息比2.92
- 即使是采用同样的特征和模型结构，风险因子中性化之后，训练的策略也会有较大的差别，可以通过此方法丰富深度学习选股策略的多样性

本文旨在对所研究问题的主要关注点进行分析，因此对市场及相关交易做了一些合理假设，但这样会导致建立的模型以及基于模型所得出的结论并不能完全准确地刻画现实环境。而且由于分析时采用的相关数据都是过去的时间序列，因此可能会与未来真实的情况出现偏差。本文内容并不是适合所有的投资者，客户在制定投资策略时，必须结合自身的环境和投资理念。

广发证券股份有限公司（以下简称“广发证券”）具备证券投资咨询业务资格。本报告只发送给广发证券重点客户，不对外公开发布，只有接收客户才可以使用，且对于接收客户而言具有相关保密义务。广发证券并不因相关人员通过其他途径收到或阅读本报告而视其为广发证券的客户。本报告的内容、观点或建议并未考虑个别客户的特定状况，不应被视为对特定客户关于特定证券或金融工具的投资建议。本报告发送给某客户是基于该客户被认为有能力独立评估投资风险、独立行使投资决策并独立承担相应风险。


本报告所载资料的来源及观点的出处皆被广发证券股份有限公司认为可靠，但广发证券不对其准确性或完整性做出任何保证。报告内容仅供参考，报告中的信息或所表达观点不构成所涉证券买卖的出价或询价。广发证券不对因使用本报告的内容而引致的损失承担任何责任，除非法律法规有明确规定。客户不应以本报告取代其独立判断或仅根据本报告做出决策。

广发证券可发出其它与本报告所载信息不一致及有不同结论的报告。本报告反映研究人员的不同观点、见解及分析方法，并不代表广发证券或其附属机构的立场。报告所载资料、意见及推测仅反映研究人员于发出本报告当日的判断，可随时更改且不予通告。

本报告旨在发送给广发证券的特定客户及其它专业人士。未经广发证券事先书面许可，任何机构或个人不得以任何形式翻版、复制、刊登、转载和引用，否则由此造成的一切不良后果及法律责任由私自翻版、复制、刊登、转载和引用者承担。

THANKS - 谢谢

“慧博资讯”是中国领先的投资研究大数据分享平台，更多分享请访问“慧博资讯”

点击进入  <http://www.hibor.com.cn>



35

Thanks !
谢谢

地址: 广州市天河北路183号大都会广场 P.C.510075 电话: 020-87555888 传真: 020-87553600 WWW.GF.COM.CN