

基于机器学习的牛股精选

-----国泰君安金融工程投资策略

姓名：刘富兵
邮箱：liufubing008481@gtjas.com
电话：021-38676673
证书编号：S0880511010017

姓名：陈奥林
邮箱：chenaolin@gtjas.com
电话：021-38674835
证书编号：S0880114110077

1

本文介绍了基于机器学习的决策树算法框架

2

相比神经网络、支持向量机等方法，其优点在于它是易于理解的“白箱”模型，可理解性更高

3

决策树算法在处理非线性解释变量时，其表现要优于线性模型

4

组合在各年份的收益率及信息比都比较稳定，年化收益21%

策略逻辑

CART决策树算法简介

基于机器学习的策略构建

决策树模型选股表现

总结与后续研究展望

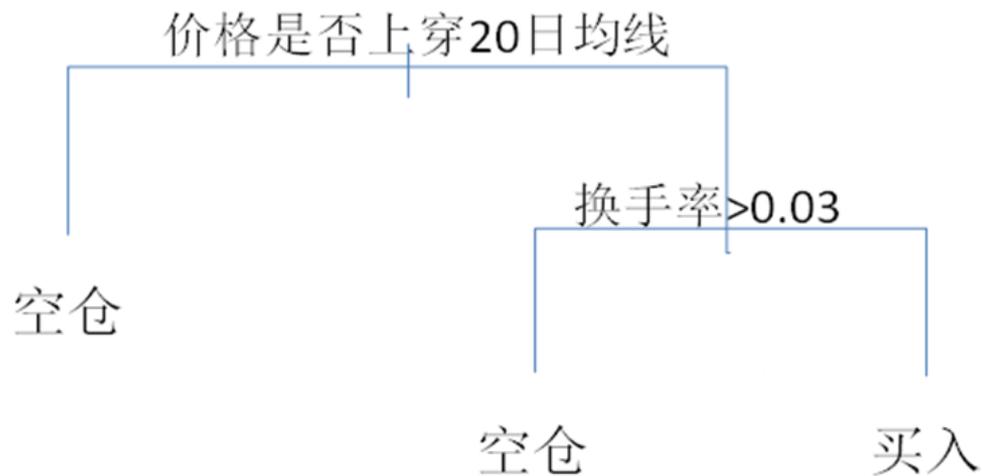


核心优势：

- ◆ 模型的可视化+可理解性
- ◆ 放松了模型变量线性的假定

- (1) 投资者决策过程是串联而不是并联的
- (2) 投资中对每个指标分配的权重是不一样的
- (3) 投资者决策过程并不满足严格的线性特征
- (4) 决策树对投资者投资决策的过程进行了系统化的展示

图1 投资决策过程展示。



资料来源：国泰君安证券研究。

策略逻辑

CART决策树算法简介

基于机器学习的策略构建

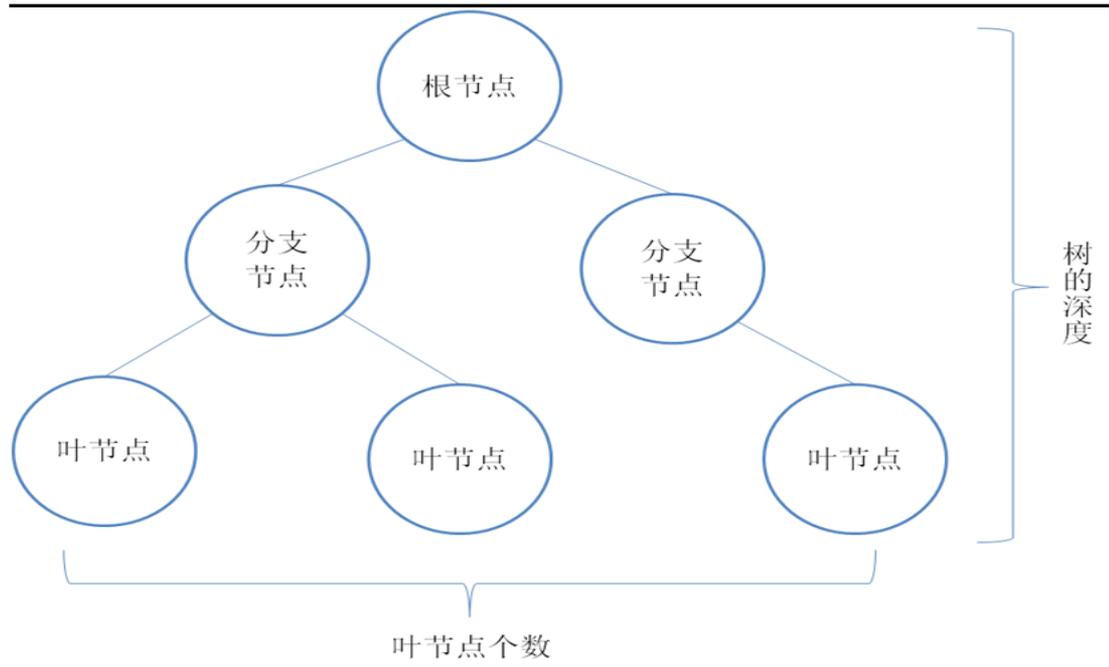
决策树模型选股表现

总结与后续研究展望

决策树算法综述

- (1) 一颗标准的决策树由叶节点、分支节点和根节点构成
- (2) 决策树算法设置区域划分的形状为矩形
- (3) 决策树算法采用了贪婪的递归二元划分法

图 1 决策树示意图。



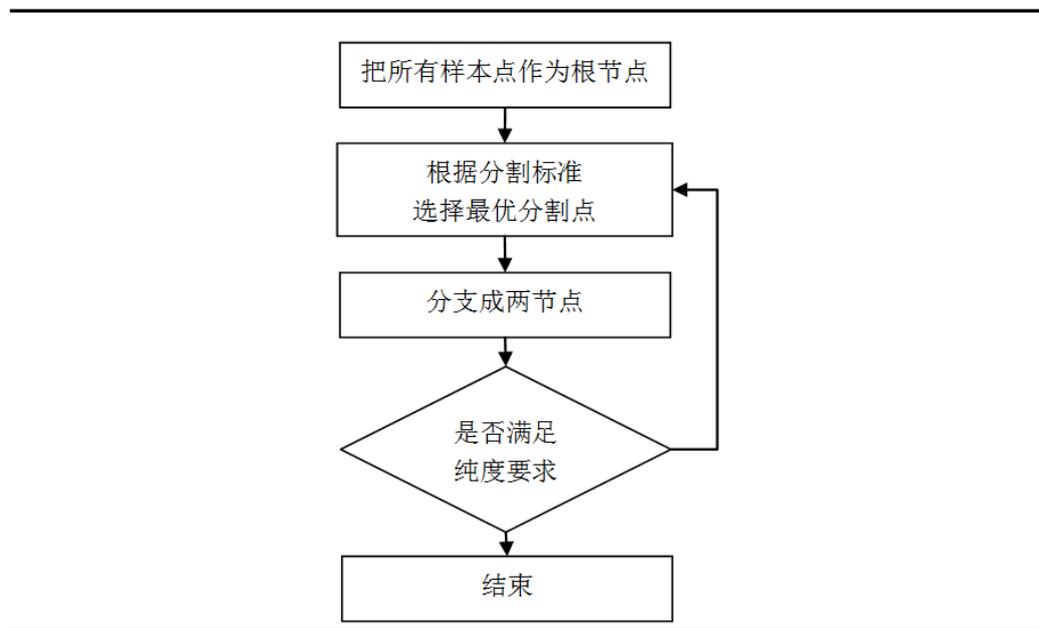
资料来源:国泰君安证券研究。

(1) 选择最优指标，在最优点进行分割

(2) 如何评估切割纯度--- $Gini = \sum_{k=1}^K P(m,k)(1 - P(m,k))$

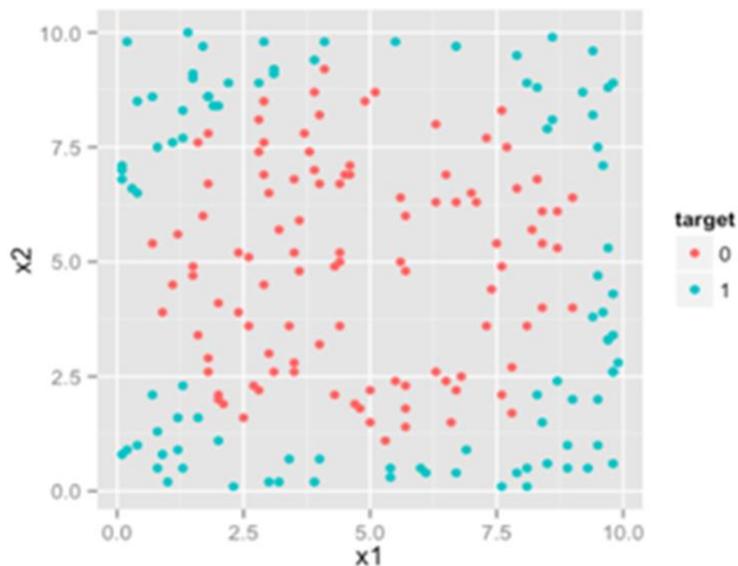
(3) 设置人工结束条件-剪枝

图2 决策树学习过程

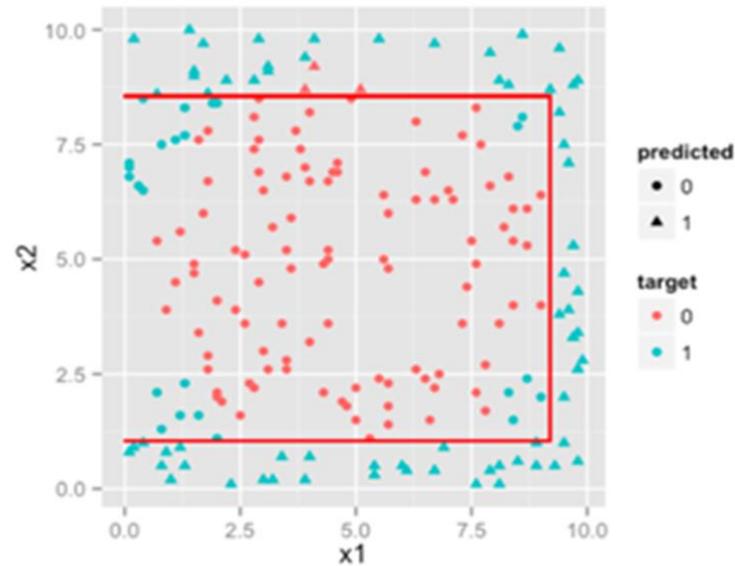


资料来源:国泰君安证券研究

算法实例



资料来源:国泰君安证券研究



资料来源:国泰君安证券研究

- (1) 算法基于 x_1 , x_2 和Y取值的关系对原样本进行不断的切分
- (2) 核心目的在于找出Y变量取值为1的区域
- (3) 模型可以容纳变量的非线性特征

例如：当 x_2 在处于1.05和8.55的范围内的時候，Y变量的取值由 x_1 决定

策略逻辑

CART决策树算法简介

基于机器学习的策略构建

决策树模型选股表现

总结与后续研究展望

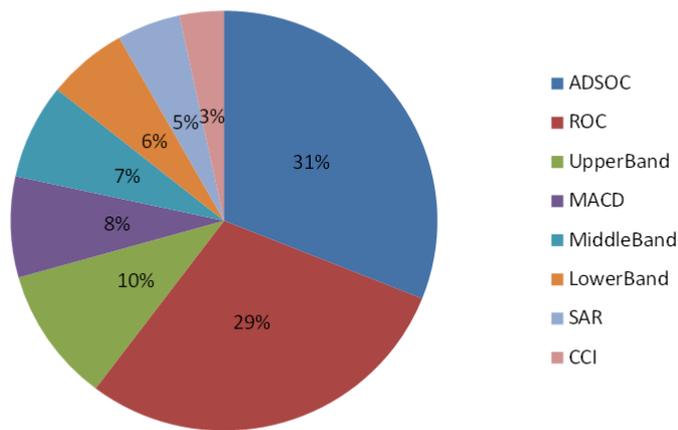
备选指标的评估与选择

- 变动速率ROC
- 累积派发指标AD
- 布林线上轨
- 平滑异同移动平均线MACD
- 顺势指标CCI
- 抛物线指标SAR



- ◆ 训练期：2006年12月至2010年12月
- ◆ 指标贡献度评价：根据解释变量x分割的节点带来的不纯度下降的累加值（标准化）

指标贡献评估。



资料来源:国泰君安证券研究。

决策树模型构建

- ◆ 样本空间：中证**500**成分股，剔除停牌股票、剔除**ST**股票
- ◆ 训练样本：**T-100**周到**T-4**周
- ◆ 换仓频率：月度换仓，年换手率**9**倍左右
- ◆ 分支节点纯度定义：指标为**1**的个数占叶节点样本总数的比例
- ◆ 剪枝参数设定：叶节点最大个数为**6**个
- ◆ 手续费：单边千**2**，均价买入
- ◆ 特殊情况：当日成交量约等于**0**，则已有持仓或待购股票默认无法交易

策略逻辑

CART决策树算法简介

基于机器学习的策略构建

决策树模型选股表现

总结与后续研究展望

决策树模型选股表现

- ▶组合累计收益率为165%
- ▶年化超额收益21.55%，信息比率2.11

图 7 对冲中证 500 指数的累计超额收益。

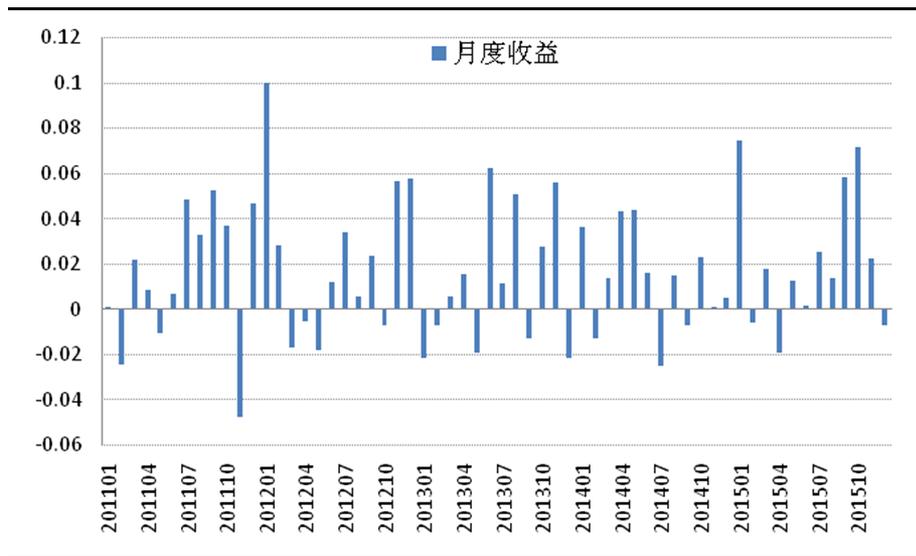


资料来源:国泰君安证券研究。

决策树模型选股表现

- ▶组合月度胜率可达70%
- ▶月度最大负收益仅4.3%

图 8 月度超额收益



资料来源:国泰君安证券研究。

决策树模型选股表现

- ▶组合在各年度超额收益显著且稳定
- ▶决策树模型与传统的多因子模型有一定互补性

表 1 组合年度超额收益

年份	2011	2012	2013	2014	2015
收益率	17.81%	30.24%	14.45%	14.88%	28.77%
信息比率	2.04	3.13	1.52	1.98	2.08
最大回撤	5.58%	5.22%	4.60%	5.32%	9.33%

资料来源:国泰君安证券研究

策略逻辑

CART决策树算法简介

基于机器学习的策略构建

决策树模型选股表现

总结与后续研究展望

核心贡献：

- 介绍了基于机器学习的决策树算法框架
- 机器学习算法使得多个指标的综合运用效率有所提升
- 模型能够容纳变量非线性的特征

未来研究计划：

- ✓ 从投资逻辑的角度出发，深度挖掘新的备选指标
- ✓ 从算法精度的角度出发，对决策树算法进行改进

本公司具有中国证监会核准的证券投资咨询业务资格

分析师声明

作者具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，保证报告所采用的数据均来自合规渠道，分析逻辑基于作者的职业理解，本报告清晰准确地反映了作者的研究观点，力求独立、客观和公正，结论不受任何第三方的授意或影响，特此声明。

免责声明

本报告仅供国泰君安证券股份有限公司（以下简称“本公司”）的客户使用。本公司不会因接收人收到本报告而视其为本公司的当然客户。本报告仅在相关法律许可的情况下发放，并仅为提供信息而发放，概不构成任何广告。

本报告的信息来源于已公开的资料，本公司对该等信息的准确性、完整性或可靠性不作任何保证。本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值及投资收入可升可跌。过往表现不应作为日后的表现依据。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。本公司不保证本报告所含信息保持在最新状态。同时，本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

本报告中所指的投资及服务可能不适合个别客户，不构成客户私人咨询建议。在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的投资建议。在任何情况下，本公司、本公司员工或者关联机构不承诺投资者一定获利，不与投资者分享投资收益，也不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。投资者务必注意，其据此做出的任何投资决策与本公司、本公司员工或者关联机构无关。

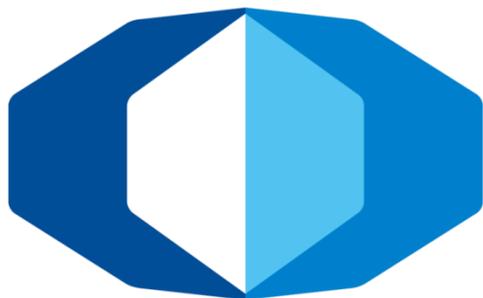
本公司利用信息隔离墙控制内部一个或多个领域、部门或关联机构之间的信息流动。因此，投资者应注意，在法律许可的情况下，本公司及其所属关联机构可能会持有报告中提到的公司所发行的证券或期权并进行证券或期权交易，也可能为这些公司提供或者争取提供投资银行、财务顾问或者金融产品等相关服务。在法律许可的情况下，本公司的员工可能担任本报告所提到的公司的董事。

市场有风险，投资需谨慎。投资者不应将本报告作为作出投资决策的唯一参考因素，亦不应认为本报告可以取代自己的判断。在决定投资前，如有需要，投资者务必向专业人士咨询并谨慎决策。

本报告版权仅为本公司所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制、发表或引用。如征得本公司同意进行引用、刊发的，需在允许的范围内使用，并注明出处为“国泰君安证券研究”，且不得对本报告进行任何有悖原意的引用、删节和修改。

若本公司以外的其他机构（以下简称“该机构”）发送本报告，则由该机构独自为此发送行为负责。通过此途径获得本报告的投资者应自行联系该机构以要求获悉更详细信息或进而交易本报告中提及的证券。本报告不构成本公司向该机构之客户提供的投资建议，本公司、本公司员工或者关联机构亦不为该机构之客户因使用本报告或报告所载内容引起的任何损失承担任何责任。

诚信 亲和 专业 创新



国泰君安证券

GUOTAI JUNAN SECURITIES