

2018年11月30日

回测曲线



研发部

分析师：程毅敏
SAC 执业证书编号：S1340511010001
联系电话：010-67017788
Email: chengyimin@cnpsec.com

基于机器学习的量化投资策略

——系列四：随机森林

- 作为经典的机器学习算法之一,随机森林在很多领域有着广泛的应用,其更加侧重于分类场景的使用。
- 就目前的策略而言,其在回撤控制方面还需要进一步改善,还需要叠加因子挖掘等研究。

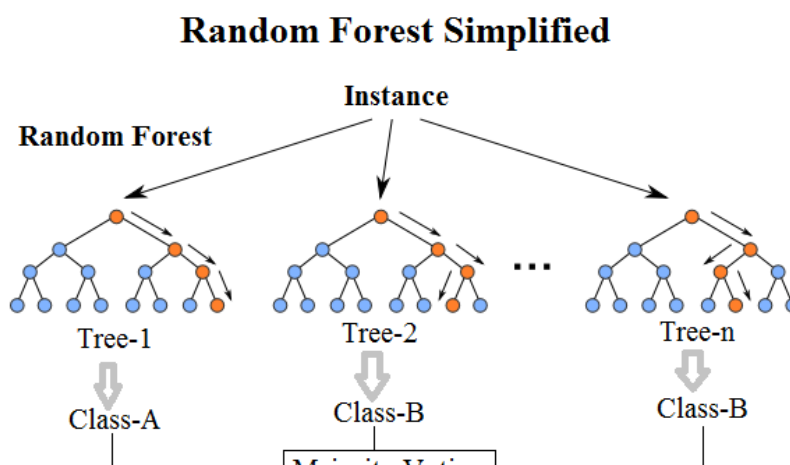
一、什么是随机森林

（一）随机森林简单介绍

作为经典的机器学习算法之一，随机森林在很多领域有着广泛的应用。随机森林是指利用多棵决策树对样本进行训练并预测的一种算法。也就是说随机森林算法是一个包含多个决策树的算法，其输出的类别是由个别决策树输出的类别的众数来决定的。

在 Sklearn 模块库中，与随机森林算法相关的函数都位于集成算法模块 ensemble 中，相关的算法函数包括：随机森林算法（RandomForestClassifier）、袋装算法（BaggingClassifier）、完全随机树算法（ExtraTreesClassifier）、迭代算法（Adaboost）、GBT 梯度 Boosting 树算法（GradientBoostingClassifier）、梯度回归算法（GradientBoostingRegressor）、投票算法（VotingClassifier）。

图表 1：随机森林示意图



资料来源：WIND 资讯，中邮证券研发部

分类、聚类和回归是机器学习的最基本主题。而随机森林主要是应用于回归和分类这两种场景，又侧重于分类。研究表明，组合分类器比单一分类器的分类效果好，在上述中我们知道，随机森林是指利用多棵决策树对样本数据进行训练、分类并预测的一种方法，它在对数据进行分类的同时，还可以给出各个变量（基因）的重要性评分，评估各个变量在分类中所起的作用。随机森林的构建大致如下：首先利用 bootstrap 方法又放回的从原始训练集中随机抽取 n 个样本，并构建 n 个决策树；然后假设在训练样本数据中有 m 个特征，那么每次分裂时选择最好的特征进行分裂 每棵树都一直这样分裂下去，直到该节点的所有训练样例都属于同一类；接着让每颗决策树在不做任何修剪的前提下最大限度的生长；最后将生成的多棵分类树组成随机森林，用随机森林分类器对新的数据进行分类与回归。对于分类问题，按多棵树分类器投票决定最终分类结果；而对于回归问题，则由多棵树预测值的均值决定最终预测结果。

（二）几个重要的参数

1、`n_estimators`: 它表示建立的树的数量。一般来说，树的数量越多，性能越好，预测也越稳定，但这也会减慢计算速度。一般来说在实践中选择数百棵树是比较好的选择，因此，一般默认是 100。2、`n_jobs`: 超参数表示引擎允许使用处理器的数量。若值为 1，则只能使用一个处理器。值为 -1 则表示没有限制。设置 `n_jobs` 可以加快模型计算速度。3、`oob_score`: 它是一种随机森林交叉验证方法，即是否采用袋外样本来评估模型的好坏。默认是 False。推荐设置为 True，因为袋外分数反应了一个模型拟合后的泛化能力。

（三）小结

随机森林的优缺点在于：1、优点：对于大部分的数据，它的分类效果比较好；能处理高维特征，不容易产生过拟合，模型训练速度比较快，特别是对于大数据而言；在决定类别时，它可以评估变数的重要性；对数据集的适应能力强：既能处理离散型数据，也能处理连续型数据，数据集无需规范化。2、缺点：对少量数据集和低维数据集的分类不一定可以得到很好的效果；计算速度比单个的决策树慢；当我们需要推断超出范围的独立变量或非独立变量，随机森林做得并不好。

二、随机森林在股票市场的简单应用

我们基于随机森林做一个简单的策略，回测标的为沪深 300 成分股，区间为：2017-01-01 至 2018-11-01，回测指标如下：

回测曲线如下：

图表 4：回测曲线



资料来源：WIND 资讯，中邮证券研发部

三、总结

随机森林是机器学习的常用算法之一，基于随机森林算法，我们研究了其在 A 股市场的应用，可以看出其在回撤控制上还有很大的研究潜力；另外可以和多因子相结合，充分挖掘因子，剔除无关因子进行预测。

中邮证券投资评级标准

股票投资评级标准：

- 推荐： 预计未来 6 个月内，股票涨幅高于沪深 300 指数 20%以上；
谨慎推荐： 预计未来 6 个月内，股票涨幅高于沪深 300 指数 10%—20%；
中性： 预计未来 6 个月内，股票涨幅介于沪深 300 指数-10%—10%之间；
回避： 预计未来 6 个月内，股票涨幅低于沪深 300 指数 10%以上；

行业投资评级标准：

- 强于大市： 预计未来 6 个月内，行业指数涨幅高于沪深 300 指数 5%以上；
中性： 预计未来 6 个月内，行业指数涨幅介于沪深 300 指数-5%—5%之间；
弱于大市： 预计未来 6 个月内，行业指数涨幅低于沪深 300 指数 5%以上；

可转债投资评级标准：

- 推荐： 预计未来 6 个月内，可转债涨幅高于中信标普可转债指数 10%以上；
谨慎推荐： 预计未来 6 个月内，可转债涨幅高于中信标普可转债指数 5%—10%；
中性： 预计未来 6 个月内，可转债涨幅介于中信标普可转债指数-5%—5%之间；
回避： 预计未来 6 个月内，可转债涨幅低于中信标普可转债指数 5%以上；

分析师声明

撰写此报告的分析师（一人或多人）承诺本机构、本人以及财产利害关系人与所评价或推荐的证券无利害关系。

本报告所采用的数据均来自我们认为可靠的目前已公开的信息，并通过独立判断并得出结论，力求独立、客观、公平，报告结论不受本公司其他部门和人员以及证券发行人、上市公司、基金公司、证券资产管理公司、特定客户等利益相关方的干涉和影响，特此声明。

免责声明

中邮证券有限责任公司具备经中国证监会批准的开展证券投资咨询业务的业务资格。

本报告信息均来源于公开资料或者我们认为可靠的资料，我们力求但不保证这些信息的准确性和完整性。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，中邮证券有限责任公司不对因使用本报告的内容而引致的损失承担任何责任。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。

中邮证券有限责任公司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。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。

中邮证券有限责任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头寸并进行交易，也可能为这些公司提供或者争取提供投资银行、财务顾问或者其他金融产品等相关服务。

《证券期货投资者适当性管理办法》于 2017 年 7 月 1 日起正式实施，本报告仅供本公司客户中的专业投资者使用，若您非本公司客户中的专业投资者，为控制投资风险，请取消接收、订阅或使用本报告中的任何信息。本公司不会因接收人收到、阅读或关注本报告中的内容而视其为客户。

本报告版权仅为中邮证券有限责任公司所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制和发布。如引用发布，需注明出处为中邮证券有限责任公司研发部，且不得对本报告进行有悖原意的引用、删节和修改。

中邮证券有限责任公司对于本免责声明条款具有修改权和最终解释权。

公司简介

中邮证券有限责任公司，2002年9月经中国证券监督管理委员会批准设立，注册资本40.6亿元人民币。中邮证券是中国邮政集团公司绝对控股的证券类金融子公司。

中邮证券的经营经营范围包括证券经纪、证券投资咨询、证券投资基金销售、融资融券、代销金融产品、证券资产管理、证券承销与保荐、证券自营和与证券交易、证券投资活动有关的财务顾问等。中邮证券目前已经在北京、陕西、深圳、山东、江苏、四川、江西、湖北、湖南、福建、辽宁、吉林、黑龙江、广东、浙江、贵州、新疆、河南、山西等地设有分支机构。

中邮证券紧紧依托中国邮政集团公司雄厚的实力，坚持诚信经营，践行普惠服务，为社会大众提供全方位专业化的证券投、融资服务，帮助客户实现价值增长。中邮证券努力成为客户认同、社会尊重，股东满意，员工自豪的优秀企业。

业务简介

■ 证券经纪业务

公司经中国证监会批准，开展证券经纪业务。业务内容包括：证券的代理买卖；代理证券的还本付息、分红派息；证券代保管、鉴证；代理登记开户；

公司为投资者提供现场、自助终端、电话、互联网、手机等多种委托通道。公司开展网上交易业务已经中国证监会核准。

公司全面实行客户交易资金第三方存管。目前存管银行有：中国邮政储蓄银行、中国工商银行、中国农业银行、中国银行、中国建设银行、交通银行、民生银行、兴业银行、招商银行、北京银行、华夏银行。

■ 证券自营业务

公司经中国证监会批准，开展证券自营业务。使用自有资金和依法筹集的资金，以公司的名义开设证券账户买卖依法公开发行或中国证监会认可的其他有价证券的自营业务。自营业务内容包括权益类投资和固定收益类投资。

■ 证券投资咨询业务

公司经中国证监会批准开展证券投资咨询业务。为证券投资者或客户提供证券投资的相关信息、分析、预测或建议。

■ 证券投资基金销售业务：公司经中国证监会批准开展证券投资基金销售业务。代理发售各类基金产品，办理基金份额申购、赎回等业务。

■ 证券资产管理业务：公司经中国证监会批准开展证券资产管理业务。

■ 证券承销与保荐业务：公司经中国证监会批准开展证券承销与保荐业务。

■ 财务顾问业务：公司经中国证监会批准开展财务顾问业务。