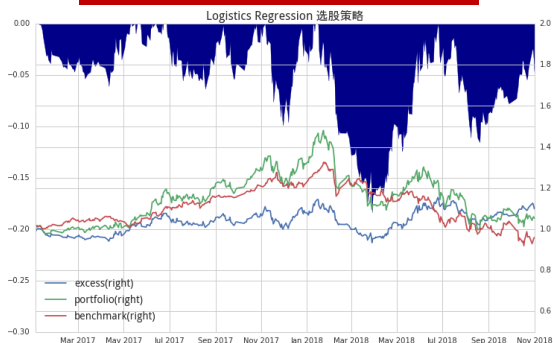


2018年11月16日

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

——系列三：Logistics regression 的探索

- 同线性回归类似，逻辑回归也是监督学习的一种，主要用于对样本进行分类。
- 基于 Logistics Regression，我们选取诸如 PB、PE、RSI 等因子作为训练数据，训练周期为 T-1 至 T-60，用 T 日的数据预测下一个换仓周期的涨跌。

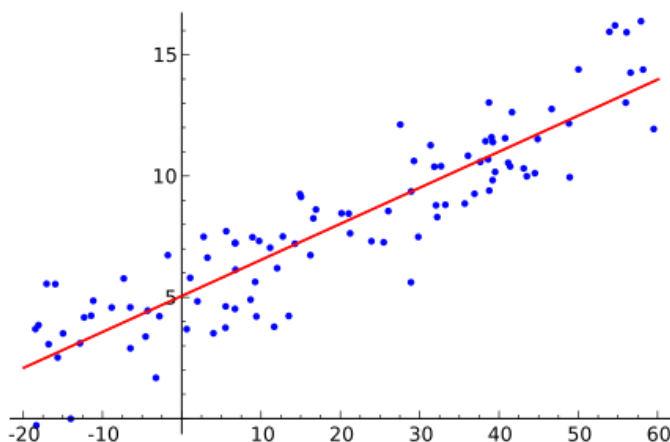
一、什么是 Logistics Regression?

(一) Logistics Regression

可能大家对线性回归比较熟悉，同线性回归类似，逻辑回归也是监督学习的一种，主要用于对样本进行分类。通常，Logistic Regression 用于二分类问题，当然它也可以用于多分类问题。

假设，我们有一些数据点，我们可以利用一条直线对这些点进行拟合(该线称为最佳拟合直线)，此过程就称为回归，如图表 1 所示。

图表 1：线性回归示意图



资料来源：WIND 资讯，中邮证券研发部

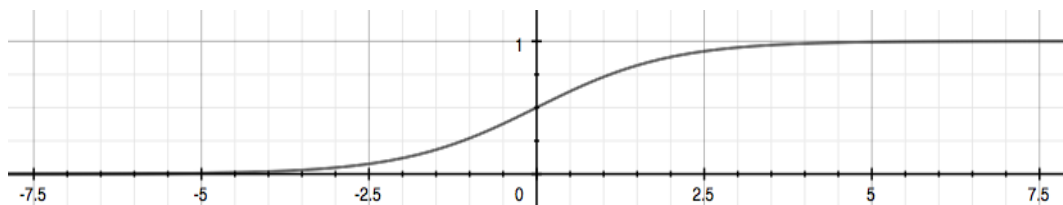
与一般的线性回归不同，Logistic 回归是分类方法，它利用的是 Sigmoid 函数阈值在[0,1]这个特性。其主要思想是：根据现有数据对分类边界线建立回归公式，以此进行分类。Logistic 本质上是一个基于条件概率的判别模型(Discriminative Model)。

从 Sigmoid 函数入手，来看 Logistic 回归：

$$g(z) = \frac{1}{1 + e^{-z}}$$

其图像如下：

图表 2：Sigmoid 函数示意图



资料来源：WIND 资讯，中邮证券研发部

对于一般的线性回归，我们有：

$$f(x) = \omega_0 x_0 + \omega_1 x_1 + \dots + \omega_n x_n + b$$

进一步，可以写为：

$$f(x) = \omega^T \mathbf{x} + b$$

从而，可以得到“广义线性回归”的一般模型：

$$y = g^{-1}(\omega^T \mathbf{x} + b)$$

其中 g^{-1} 为单调可微函数。

我们知道线性回归模型只能够进行回归学习，但是若要是做分类任务该如何做？答案便是在“广义线性回归”模型中：需找一个单调可微函数将分类任务的真实标记 y 与线性回归模型的预测值联系起来。

logistic 回归是处理二分类问题的，输出的标记 $y = \{0, 1\}$ ，并且线性回归模型产生的预测值 $z = \omega^T \mathbf{X} + b$ 是一个实值，所以我们将实值 z 转化成 0/1 值便可，这样有一个可选函数便是“单位阶跃函数”：

$$y = \begin{cases} 1, z > 0 \\ 0.5, z = 0 \\ 0, z < 0 \end{cases}$$

回想前面的 Sigmoid 函数，其在一定程度上近似单位阶跃函数，同时单调可微。这样，我们可以在原来的线性回归模型上套用 sigmoid 函数，便形成了 logistic 回归模型的预测函数，可以用于二分类问题：

$$y = \frac{1}{1 + e^{-(\omega^T \mathbf{X} + b)}}$$

进一步，可以得到：

$$\ln \frac{y}{1-y} = \omega^T \mathbf{X} + b$$

换一种思路：若将 y 视为样本 \mathbf{X} 作为正例的可能性，则 $1-y$ 便是其反例的可能性。二者的比值便被称为“几率”，反映了 \mathbf{X} 作为正例的相对可能性，这也是 logistic 回归又被称为对数几率回归的原因。

总结一下线性回归模型和 logistic 回归的关系：

logistic 回归分类模型的预测函数是在用线性回归模型的预测值的结果去逼近真实标记的对数几率，这样也将线性回归的预测值和分类任务的真实标记联系在一起。

（二）似然函数

对于式子：

$$\ln \frac{y}{1-y} = \omega^T \mathbf{X} + b$$

我们可以将其中的 y 视为后验概率 $P(y=1|\mathbf{X})$ ，因有：

$$\ln \frac{P(y=1|x)}{1-P(y=1|x)} = \omega^T \mathbf{X} + b = \ln \frac{P(y=1|x)}{P(y=0|x)}$$

求解可得：

$$P(y=1|x) = \frac{e^{\omega^T x + b}}{1 + e^{\omega^T x + b}} = h_{\omega}(x)$$

$$P(y=0|x) = \frac{1}{1 + e^{\omega^T x + b}} = 1 - h_{\omega}(x)$$

$$h_{\omega}(x) = g(\omega^T x) = \frac{1}{1 + e^{-(\omega^T x + b)}}$$

将上面两个式子合并，可以得到：

$$P(y|\omega, x) = (h_{\omega}(x))^y (1 - h_{\omega}(x))^{1-y}$$

因问题本身是一个二分类问题，因而式中 y 的取值为 0 或者 1。

利用最大似然估计，可以得到：

$$L(\omega) = \prod_{i=1}^m p(y^i|x^i\omega) = \prod_{i=1}^m (h_{\omega}(x^i))^{y^i} (1 - h_{\omega}(x^i))^{1-y^i}$$

其中 m 为样本总数。

取对数，可以有：

$$J(\omega) = \frac{1}{m} \ln L(\omega) = \frac{1}{m} \sum_{i=1}^m (y^i \ln h_{\omega}(x^i) + (1 - y^i) \ln (1 - h_{\omega}(x^i)))$$

我们的目标便是最大化似然函数，也就是最大化准则函数：

$$\max_{\omega} J(\omega)$$

(三) 梯度上升法求解参数 ω

我们利用梯度上升法来求解参数 ω ，参数 ω 的迭代形式如下：

$$\omega_{j+1} = \omega_j + \alpha \nabla J(\omega_j)$$

其中： α 为“学习率”

对准则函数 $J(\omega)$ 进行微分，可以得到下列表达式：

$$\frac{\partial}{\partial \omega_j} J(\omega) = \frac{\partial J(\omega)}{\partial g(\omega^T x)} * \frac{\partial g(\omega^T x)}{\partial \omega^T x} * \frac{\partial \omega^T x}{\partial \omega_j}$$

对于右侧各部分，我们有：

$$\frac{\partial J(\omega)}{\partial g(\omega^T x)} = y * \frac{1}{g(\omega^T x)} + (y - 1) * \frac{1}{1 - g(\omega^T x)}$$

$$\frac{\partial g(\omega^T x)}{\partial \omega^T x} = g(\omega^T x) * (1 - g(\omega^T x))$$

$$\frac{\partial \omega^T x}{\partial \omega_j} = x^j$$

因而，可以有：

$$\frac{\partial}{\partial \omega_j} J(\omega) = (y - h_\omega(x)) x^j$$

于是（此处去掉 $1/m$ 并不影响最终结果）：

$$\omega_{j+1} = \omega_j + \alpha \sum_{i=1}^m (y^i - h_\omega(x^i)) x_j^i$$

（四）小结

Logistic 回归的一般过程：

1、收集数据：采用任意方法收集数据；2、准备数据：由于需要进行距离计算，因此要求数据类型为数值型。另外，结构化数据格式则最佳；3、分析数据：采用任意方法对数据进行分析；4、训练算法：大部分时间将用于训练，训练的目的是为了找到最佳的分类回归系数；5、测试算法：一旦训练步骤完成，分类将会很快；6、使用算法：首先，我们需要输入一些数据，并将其转换成对应的结构化数值；接着，基于训练好的回归系数，就可以对这些数值进行简单的回归计算，判定它们属于哪个类别；在这之后，我们就可以在输出的类别上做一些其他分析工作。

二、探索 Logistics Regression 在股票市场的应用

我们基于 logistics 回归做一个简单的策略，用 T-1 天股票的因子值（比如 PB、PE、RSI 等），预测股票的涨跌（未来 10 个交易日，调仓周期为 10 个交易日）；训练数据为 T-1 到 T-60 之间的滚动时间窗口。

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：

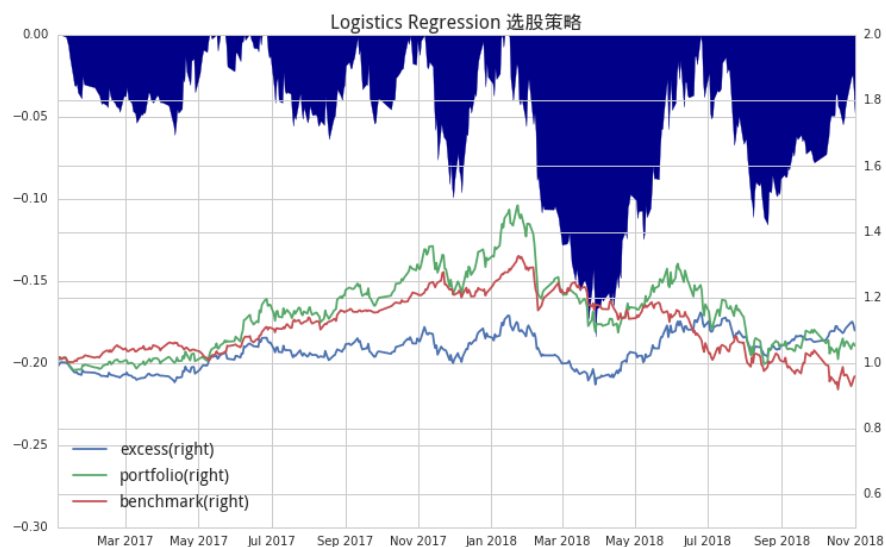
图表 3：回测指标

策略年化收益	2.9%
基准年化收益	-2.3%
α	5.5%
Sharp Ratio	-0.02
最大回撤	32.7%

资料来源：WIND 资讯，中邮证券研发部

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

Logistics regression 是简单的机器学习算法之一，其背后的数学含义较为简单；基于此，我们研究了逻辑回归在 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。