

# 股票数据的分析与预测

组员： 姚明宇  
于畅泳  
于文楠

2016-6-29

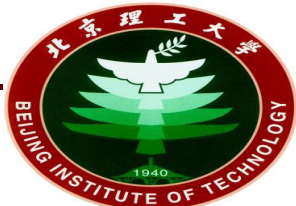


# 数据获取

从财经网站上下载得到下列所示的股票指标：上海证券交易所的招商银行股票最新的交易数据。

日期	股票代码	名称	收盘价	最高价	最低价	开盘价	前收盘	涨跌额	涨跌幅	换手率	成交量	成交金额	总市值	流通市值
2016/6/14	600036	招商银行	17.38	17.5	17.29	17.33	17.32	0.06	0.3464	0.0531	10953458	190753279	4.38E+11	3.59E+11
2016/6/13	600036	招商银行	17.32	17.79	17.29	17.58	17.77	-0.45	-2.5324	0.0786	16220713	284764728	4.37E+11	3.57E+11
2016/6/8	600036	招商银行	17.77	17.84	17.65	17.75	17.8	-0.03	-0.1685	0.0461	9514926	168924714	4.48E+11	3.67E+11
2016/6/7	600036	招商银行	17.8	17.87	17.59	17.76	17.64	0.16	0.907	0.0545	11250355	199544447	4.49E+11	3.67E+11
2016/6/6	600036	招商银行	17.64	17.83	17.56	17.73	17.78	-0.14	-0.7874	0.0619	12770352	225235543	4.45E+11	3.64E+11

将获取的原始数据转换成时间序列数据 (xts)，方便之后的计算和建模



# 模型建立

定义指标变量T，假设在未来k天内我们的目标是获得b%的利润，一天的平均价格可以用当天的收盘价、最高价和最低价的均值来表示：

$$\bar{P}_i = \frac{C_i + H_i + L_i}{3}$$

$V_i$ ：表示未来k天的平均价格相对今天收盘价的百分比变化：

$$V_i = \frac{\bar{P}_{i+j} - C_i}{C_i} \quad j = 1, 2, \dots, k$$

把动态变化绝对值超过目标收益p%的变化进行累加，作为我们的指标变量T

$$T_i = \sum_v \{ v \in V_i : v > b\% \vee v < -b\% \}$$



---

指标T用来找出在k天内，日平均价格明显高于目标变化的那些日期的变化之和。

若T值为正，并且值较大则表明有几天的日平均报价高于今天收盘价的b%，这种情况说明对于未持该股票的股民可以进行买入；

相反，若T值为负，表明价格可能下降，这种情况对应着持有该股票的股民可以进行卖出；

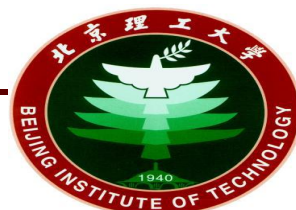
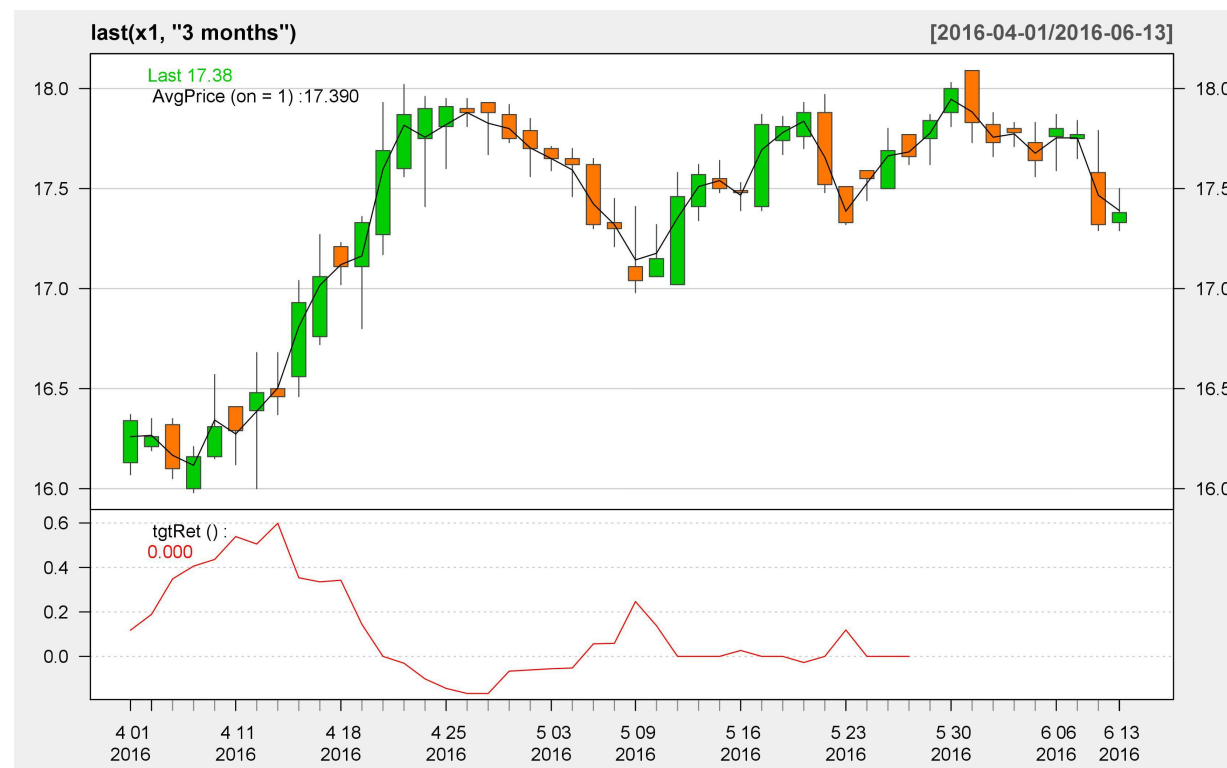
T值趋于0则表示价格相对平稳，可以进行买入或卖出。在实验中，我们设计的b值为2.0。



导入股票数据之后，我们首先计算数据的T指标，并画出K线图和指标线图

K线图中柱条：表示当天的最高、最低价格；  
框：表示开盘价和收盘价；  
橙色：一天中呈下降趋势；  
绿色：一天中呈上升趋势；

本实验中计算的是未来 10 天内的T值，如 Figure1中下方的图线所示 根据未来 10 天内股价的变化来得到T值，在股价大跌时取得较大的负值，大涨时取得较大的正值。



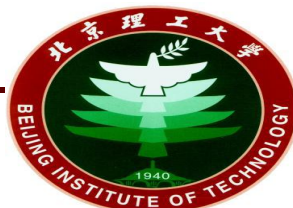
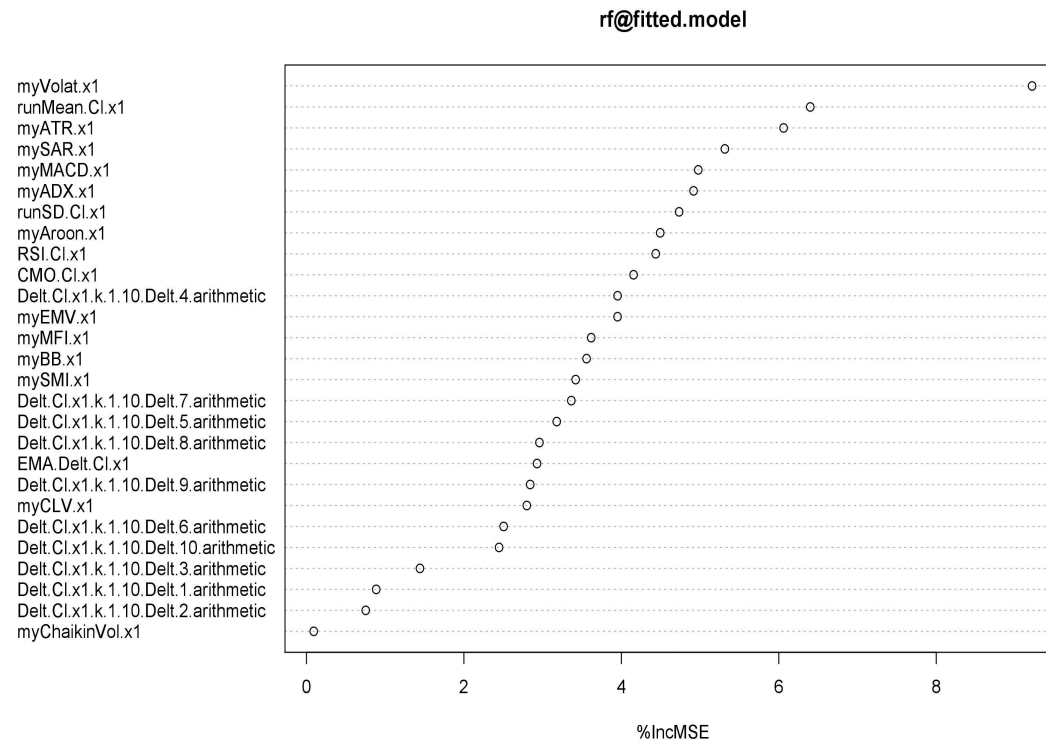
# 股票评价指标

TTR程序包中给出了许多种评价股票的指标：

ATR(Average True Range,平均真实波幅)  
SMI(Stochastic Momentum Index,随机动量指数)  
ADX(Average Directional Index,平均趋向指标)  
MACD(Moving Average Convergence/Divergence, 指数平滑异同移动平均线)等。

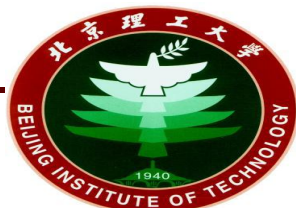
面对诸多指标，我们引入随机森林算法，用它来检测指标的重要程度，选择几个重要的指标，计算结果如下图：

我们筛选出重要性大于5%的指标，用这些大于5%的指标的线性组合来表示股票最终的评价。



通过实验，我们选出了6种评价股票的指标：

- SAR: Stop and Reverse, 又叫抛物线指标或停损转向操作点指标，由美国技术分析大师 Wells Wilder所创造。
- ADX: Average Directional Index, 平均趋向指标，ADX无法告诉我们趋势的发展方向。可是如果趋势存在, ADX 可以衡量趋势的强度。
- runMean: 滑动平均值，由一个长度为10的滑动窗口来计算平均值。
- ATR: Average True Range, 平均真实波幅，取一定时间周期内的股价波动幅度的移动平均值，主要用于判断买卖时机，是显示市场变化率的反趋向指标, 由J. Welles Wilder在1978年提出。
- Aroon: 1995年由Tushar Chande发明，它通过计算自价格达到近期最高值和最低值以来所经过的期间数，帮助投资者预测证券价格趋势、强弱以及趋势的反转等。
- runSD: 滑动标准差，由一个长度为10的滑动窗口来计算标准差。



---

在能够表示股票的指标之后，我们便可以使用相关算法进行回归、分类。

实验中，我们对比了两个比较常用的机器学习算法：SVM和人工神经网络。

通过将数据集平分成两部分，前半部分作为训练集、后半部分作为测试集，得到一组预测精确度和回溯精确度的值

我们可以发现，人工神经网络的准确率远高于SVM，因此我们最终选择使用人工神经网络进行建模。





---

# Thank you !

