

投资决策流程

所有过程均可复现

1: 安装工具包

A: ClementineV12

B: pycharm

C: 区分度算法工具

2: 流程

甄选行业-》挑选股票-》特征工程-》精选股票

3: 模型参数评估指标（见流程里面）

4: 相关文件

5: GB 的算法原理及算法设计

6: 关于模型优化

7: 风险预警及换股策略

以上流程全部实现后会形成一个模型部署文件，以后只要调用这个部署文件就可以，目前这个文件是：**function.py**

甄选行业

1: 区分度算法

2: 神经网络算法

3: 线性回归算法

1: 区分度算法模型评价指标

区分度模型之所以能用于选股是建立在道氏理论的第一假设：股价走势反应一切市场信息。我们通过收集尽可能多的股票信息来挑选股票。

需要用知测软件来做模型（已经给吴老师安装好了，也教会她如何使用了。）

首先使用大智慧把所有行业及行业相关的数据全部导出到 excel，并做好预处理。见文件

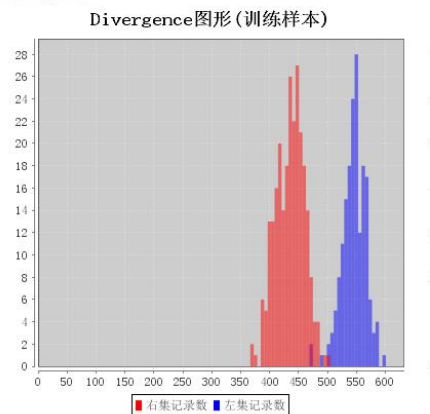
（行业 20191124 - 副本.txt 和行业 20191124 - 副本.xls，其中 txt 用于知测软件使用，xls 用于 clementine 使用），然后用知测做模型，模型评估参数如下：

评分模型及主要统计量

| 样本类型 | 区分度(Divergence) | ROC面积 | Gini系数 | KS值 | KS 评分 / KS 尺度化评分 | 通过率(KS Pass Rate) | 判定系数(R-squared) |
|----------------|-----------------|-------|--------|--------|------------------|-------------------|-----------------|
| 训练样本(Training) | 24.693 | 0.999 | 0.999 | 98.315 | 2.044 / 499.000 | 40.509 | 0.727 |

区分度见下图

Divergence图



然后导出文件就能得到每个行业的打分情况。见文件 **export-区分度.txt**
其中前 50 个行业及得分如下

| SERIES | 代码 | 名称 | UNSCALED_SCORE | SCALINGSCORE |
|--------|---------|-------|----------------|--------------|
| 1 | 994453' | 谷歌概念 | 30.67367322 | 601 |
| 2 | 994164' | 网红经济 | 26.77335233 | 587 |
| 3 | 994238' | 智慧停车 | 25.72912207 | 583 |
| 4 | 994440' | 京东概念 | 25.65738297 | 583 |
| 5 | 994126' | 汽车电子 | 25.64096905 | 583 |
| 8 | 994177' | 血液制品 | 24.82939153 | 580 |
| 10 | 994062' | 国产软件 | 24.74506669 | 579 |
| 11 | 994421' | 独角兽概念 | 24.18004624 | 577 |
| 12 | 994431' | 自主可控 | 23.81574445 | 576 |
| 13 | 994353' | 无人零售 | 23.56582127 | 575 |
| 14 | 992044' | 贵州 | 23.55295858 | 575 |
| 16 | 994283' | 特种玻璃 | 23.39981903 | 575 |
| 17 | 994086' | 健康中国 | 23.28783265 | 574 |
| 20 | 994489' | 超高清 | 23.26752355 | 574 |
| 21 | 994419' | 富士康概念 | 21.91140811 | 569 |
| 23 | 993939' | 白酒 | 21.87810597 | 569 |
| 24 | 994176' | 信息安全 | 21.85397821 | 569 |
| 25 | 994420' | 工业互联网 | 21.82241778 | 569 |
| 26 | 994483' | 新基建 | 21.81207014 | 569 |
| 27 | 993063' | 抗癌药物 | 21.81134575 | 569 |
| 28 | 991006' | 电子设备 | 21.78574932 | 569 |
| 29 | 994228' | SAAS | 21.56043153 | 568 |
| 30 | 994097' | 壳资源 | 21.5145308 | 568 |
| 31 | 994082' | 集成电路 | 21.42258499 | 568 |
| 32 | 994142' | 生物识别 | 21.26997208 | 567 |
| 33 | 994318' | 新版医保 | 21.23589771 | 567 |
| 34 | 994428' | 仿制药 | 21.14758941 | 567 |

| | | | | |
|----|---------|------|-------------|-----|
| 35 | 994537' | 医用耗材 | 20.92803295 | 566 |
| 36 | 994512' | 休闲食品 | 20.82957404 | 566 |
| 37 | 993075' | 次新股 | 20.64711601 | 565 |
| 38 | 993756' | LED | 20.61567111 | 565 |
| 39 | 994387' | 定制家居 | 20.44390275 | 564 |
| 40 | 992058' | 山东 | 20.28275175 | 564 |
| 42 | 993979' | 医疗美容 | 20.20898779 | 563 |
| 43 | 991043' | 互联网 | 20.13341929 | 563 |
| 44 | 994487' | 纳米银线 | 19.96393034 | 562 |
| 45 | 994198' | 在线旅游 | 19.96048072 | 562 |
| 46 | 994286' | 体外诊断 | 19.95956819 | 562 |
| 48 | 994384' | 贵州国资 | 19.74345079 | 562 |
| 49 | 994203' | 智能交通 | 19.74046583 | 562 |
| 50 | 994479' | 光刻胶 | 19.66566018 | 561 |
| 51 | 993980' | 人工智能 | 19.40280424 | 560 |

2: 神经网络算法

使用文件 **20191124 - 副本.xls** 作为导入 **clementine** 的数据，对数据做神经网络算法（已经给吴老师安装好 **clementine** 并教会了她如何使用），模型输出的参数如下：

| | |
|----------------|--------|
| 输出字段 GB 的结果 | |
| 比较 \$N-GB 与 GB | |
| 最小误差 | -0.193 |
| 最大误差 | 0.334 |
| 平均误差 | 0.007 |
| 绝对平均误差 | 0.041 |
| 标准差 | 0.056 |
| 线性相关 | 0.774 |
| 发生率 | 540 |

然后导出文件就能得到每个行业的打分情况。见文件 **export-neuro.xls**

其中前 50 个行业及得分如下：

| no | 代码 | 名称 | \$N-GB |
|-----|---------|-------|------------|
| 1 | 994247' | 昨日涨停 | 0.43730929 |
| 2 | 994453' | 谷歌概念 | 0.34486257 |
| 6 | 994450' | 脸书概念 | 0.27738447 |
| 7 | 994472' | 抖音概念 | 0.26847779 |
| 11 | 994384' | 贵州国资 | 0.24796489 |
| 17 | 994286' | 体外诊断 | 0.22672166 |
| 5 | 994454' | 眼科概念 | 0.22297132 |
| 39 | 993386' | 预盈预增 | 0.2008617 |
| 3 | 994123' | 苹果产业链 | 0.18959628 |
| 104 | 994427' | 无感支付 | 0.18463997 |
| 54 | 991020' | 酒及饮料 | 0.1805187 |

| | | | |
|-----|---------|--------|------------|
| 53 | 994537' | 医用耗材 | 0.17857906 |
| 108 | 994079' | 机器视觉 | 0.17841791 |
| 48 | 993939' | 白酒 | 0.17350777 |
| 13 | 994241' | 降解材料 | 0.17343566 |
| 63 | 994475' | 芬太尼概念 | 0.16545841 |
| 72 | 994239' | 盖板玻璃 | 0.15915769 |
| 107 | 993064' | 医疗器械 | 0.15800813 |
| 65 | 991028' | 医疗卫生 | 0.15717255 |
| 31 | 993975' | 增强现实 | 0.15169452 |
| 92 | 994188' | 乙肝疫苗 | 0.15033966 |
| 66 | 991006' | 电子设备 | 0.15002677 |
| 47 | 994447' | PCB 概念 | 0.14963588 |
| 79 | 994415' | 抗流感 | 0.14833708 |
| 10 | 993971' | 细胞治疗 | 0.14714522 |
| 295 | 994442' | 网易概念 | 0.14659615 |
| 9 | 994240' | GPU | 0.1453312 |
| 87 | 994177' | 血液制品 | 0.14439125 |
| 141 | 994403' | 覆铜板 | 0.14428412 |
| 12 | 994474' | 麻醉概念 | 0.14293299 |
| 14 | 992044' | 贵州 | 0.14099113 |
| 42 | 994062' | 国产软件 | 0.13638896 |
| 8 | 994202' | 智能穿戴 | 0.13573997 |
| 38 | 994428' | 仿制药 | 0.13421157 |
| 43 | 994397' | 芯片概念 | 0.12848366 |
| 34 | 994095' | 抗癌治癌 | 0.12809189 |
| 37 | 993063' | 抗癌药物 | 0.1276954 |
| 19 | 993053' | 苹果三星 | 0.12560147 |
| 55 | 994479' | 光刻胶 | 0.12508324 |
| 71 | 994081' | 激光概念 | 0.12417827 |
| 91 | 994356' | 华为概念 | 0.12381396 |
| 240 | 993933' | 疫苗 | 0.12351831 |
| 36 | 994318' | 新版医保 | 0.1229208 |
| 504 | 994233' | 养鸡 | 0.12178947 |
| 88 | 994144' | 生物医药 | 0.12008959 |
| 124 | 994518' | 操作系统 | 0.11929451 |
| 114 | 993047' | 安防 | 0.11877935 |
| 30 | 994424' | 6G 概念 | 0.11867425 |
| 23 | 994168' | 无线充电 | 0.11853995 |
| 61 | 994416' | 小米概念 | 0.11828613 |

3: 线性回归算法

使用文件 **20191124 - 副本.xls** 作为导入 **clementine** 的数据，对数据做线性回归算法，模型输出的参数如下：

| | |
|----------------|-------|
| 输出字段 GB 的结果 | |
| 比较 \$E-GB 与 GB | |
| 最小误差 | -0.18 |
| 最大误差 | 0.241 |
| 平均误差 | 0.0 |
| 绝对平均误差 | 0.037 |
| 标准差 | 0.052 |
| 线性相关 | 0.816 |
| 发生率 | 540 |

然后导出文件就能得到每个行业的打分情况。见文件 **export_regress2.xls**

其中前 50 个行业及得分如下：

| no | 代码 | 名称 | \$N-GB |
|-----|---------|--------|-------------|
| 5 | 994454' | 眼科概念 | 0.258944173 |
| 48 | 993939' | 白酒 | 0.197181366 |
| 53 | 994537' | 医用耗材 | 0.195625735 |
| 108 | 994079' | 机器视觉 | 0.17125673 |
| 55 | 994479' | 光刻胶 | 0.168930624 |
| 15 | 993951' | 智能手机 | 0.160259207 |
| 95 | 994002' | 区块链 | 0.15519826 |
| 47 | 994447' | PCB 概念 | 0.150622375 |
| 4 | 994351' | 智能音箱 | 0.149226831 |
| 69 | 994512' | 休闲食品 | 0.145988728 |
| 66 | 991006' | 电子设备 | 0.139098536 |
| 28 | 991043' | 互联网 | 0.134363334 |
| 71 | 994081' | 激光概念 | 0.133985218 |
| 43 | 994397' | 芯片概念 | 0.131091208 |
| 8 | 994202' | 智能穿戴 | 0.129358245 |
| 91 | 994356' | 华为概念 | 0.128777768 |
| 121 | 994431' | 自主可控 | 0.127922101 |
| 474 | 994317' | 次新开板 | 0.125885265 |
| 132 | 993057' | 网络安全 | 0.121948513 |
| 127 | 993978' | 人脸识别 | 0.121685228 |
| 33 | 994347' | 腾讯云 | 0.121047456 |
| 77 | 993980' | 人工智能 | 0.118922832 |
| 67 | 994050' | 高送转 | 0.118793996 |
| 61 | 994416' | 小米概念 | 0.117387858 |
| 41 | 994292' | 金融机具 | 0.11733688 |
| 49 | 993970' | 基因测序 | 0.117042093 |
| 87 | 994177' | 血液制品 | 0.11568743 |

| | | | |
|-----|---------|--------|-------------|
| 29 | 994086' | 健康中国 | 0.11542714 |
| 107 | 993064' | 医疗器械 | 0.114824337 |
| 50 | 993974' | 虚拟现实 | 0.11466687 |
| 74 | 991004' | 软件 | 0.114555851 |
| 277 | 994231' | 猪肉概念 | 0.114113477 |
| 114 | 993047' | 安防 | 0.112722037 |
| 60 | 994500' | 流媒体 | 0.11132089 |
| 124 | 994518' | 操作系统 | 0.11107297 |
| 247 | 993752' | 融资融券 | 0.110832175 |
| 23 | 994168' | 无线充电 | 0.108464529 |
| 240 | 993933' | 疫苗 | 0.107826357 |
| 37 | 993063' | 抗癌药物 | 0.106826573 |
| 65 | 991028' | 医疗卫生 | 0.106635277 |
| 111 | 993979' | 医疗美容 | 0.105749691 |
| 94 | 994082' | 集成电路 | 0.105550603 |
| 20 | 994525' | ETC 概念 | 0.105264397 |
| 233 | 991013' | 农副食品 | 0.101557032 |
| 76 | 993059' | 手游 | 0.100203617 |
| 31 | 993975' | 增强现实 | 0.099254123 |
| 39 | 993386' | 预盈预增 | 0.098780607 |
| 51 | 991045' | 技术服务 | 0.098597733 |
| 19 | 993053' | 苹果三星 | 0.096911492 |
| 85 | 994242' | 半导体 | 0.0963662 |

然后我们合并三种算法共同的行业，并按排名高低甄选了 13 个行业

| 序号 | ★ | 代码 | 名称 |
|----|----|----------|-------|
| 田 | 1 | ★ 994453 | 谷歌概念 |
| 田 | 2 | ★ 994454 | 眼科概念 |
| 田 | 3 | ★ 994537 | 医用耗材 |
| 田 | 4 | ★ 994079 | 机器视觉 |
| 田 | 5 | ★ 993939 | 白酒 |
| 田 | 6 | ★ 994177 | 血液制品 |
| 田 | 7 | ★ 994447 | PCB概念 |
| 田 | 8 | ★ 991006 | 电子设备 |
| 田 | 9 | ★ 994431 | 自主可控 |
| 田 | 10 | ★ 994512 | 休闲食品 |
| 田 | 11 | ★ 994479 | 光刻胶 |
| 田 | 12 | ★ 994384 | 贵州国资 |
| 田 | 13 | ★ 994286 | 体外诊断 |

然后分别把每个行业的所有股票放到一个股票池，总共 583 只股票



| 行情 | 序号 | 代码 | 名称 | 最新 | 涨幅 | 总手 | 73组涨幅 | 每股收益 | 市净率 | 净利润 |
|-----|----|--------|------|------|-------|-------|--------|--------|-------|-------|
| 574 | ★ | 300685 | 艾德生物 | 7158 | 0.11 | 19357 | 54.50% | 0.748 | 12.22 | 1.09亿 |
| 575 | ★ | 600645 | 中源协和 | 1619 | -1.34 | 16751 | -8.48% | 0.200 | 2.26 | 9350万 |
| 576 | ★ | 601010 | 文峰股份 | 297 | 0.68 | 17785 | -8.90% | 0.132 | 1.17 | 2.44亿 |
| 577 | ★ | 603108 | 润达医疗 | 1005 | -0.79 | 19072 | -3.74% | 0.500 | 2.12 | 2.92亿 |
| 578 | ★ | 603387 | 基蛋生物 | 2266 | -1.56 | 20064 | -2.50% | 0.770 | 3.98 | 1.99亿 |
| 579 | ★ | 603658 | 安图生物 | 9958 | 0.35 | 24737 | 52.99% | 1.280 | 19.63 | 5.38亿 |
| 580 | ★ | 603716 | 塞力斯 | 1621 | -1.28 | 9428 | -0.18% | 0.450 | 2.10 | 8882万 |
| 581 | ★ | 603718 | 海利生物 | 1184 | -2.07 | 15018 | -7.06% | 0.0364 | 7.11 | 2341万 |
| 582 | ★ | 688068 | 热景生物 | 4198 | -1.39 | 5166 | | 0.280 | 4.18 | 1323万 |
| 583 | ★ | 688389 | 普门科技 | 1596 | -2.80 | 28980 | | 0.157 | 6.64 | 6610万 |

可生成一个板块数据:自选股 4.BLK。该文件可直接拷贝到大智慧后就能看到所选的所有股票。

第二步：特征工程

1: 特征选取

首先从大智慧对这个板块的所有股票抽取所需数据，然后保持到 excel 中，见文件：feature.xls

2: 特征融合算法

可以使用多种算法（DecisionTree，GradientBoosting，ExtraTree，AdaBoost，randomforest，Lightgbm,XGB,DeepRandomforest）随意组合然后并再融合对 feature 中的 63 个特征做特征抽取，算法代码见文件 mul_algorithm.py.

目前特征融合算法改为并行计算，能很快出结果。之前因为采用 pipeline 方式是串行化，所以计算比较慢。这次改进后极大的提升了效率。

3: 入选股票和特征

第三步：精选股票

1: 选股算法

神经网络算法

首先从大智慧对这个板块的所有股票抽取所需数据，然后保持到 excel 中，见文件：stock.xls

然后剔除相关性较强的几个变量，见下表：

| 字段 | 类型 | 值 | 缺失 | 检查 | 方向 |
|--------|-----|----------------|----|----|----|
| no | 范围 | [1.0,567.0] | 无 | 无 | 无 |
| 代码 | 无类型 | | 无 | 无 | 无 |
| 名称 | 无类型 | | 无 | 无 | 无 |
| GB | 范围 | [-0.175,0.... | 无 | 无 | 输出 |
| gb2 | 范围 | [0.0,1.0] | 无 | 无 | 无 |
| 每股收益 | 范围 | [-0.44,13.... | 无 | 无 | 输入 |
| 每股经营现金 | 范围 | [-5.04,12.... | 无 | 无 | 输入 |
| 每股未分配 | 范围 | [-6.194,83.... | 无 | 无 | 输入 |
| 股东权益比 | 范围 | [-9.936,75.... | 无 | 无 | 输入 |
| 净利润同比 | 范围 | [-1768.78.... | 无 | 无 | 输入 |
| 销售毛利率 | 范围 | [-31.676,6.... | 无 | 无 | 输入 |
| 股东权益 | 范围 | [271043.0.... | 无 | 无 | 输入 |
| 经营现金流量 | 范围 | [-5.55562.... | 无 | 无 | 输入 |
| 筹资现金流量 | 范围 | [-1928815.... | 无 | 无 | 输入 |
| 现金增加额 | 范围 | [-866856.... | 无 | 无 | 输入 |
| BIAS | 范围 | [-25.073,2.... | 无 | 无 | 输入 |
| CCI | 范围 | [-393.241.... | 无 | 无 | 输入 |
| RSI | 范围 | [15.078,7.... | 无 | 无 | 输入 |
| CR | 范围 | [29.495,2.... | 无 | 无 | 输入 |
| PSY | 范围 | [8.333,83.... | 无 | 无 | 输入 |
| MACD | 范围 | [-5.782,2.... | 无 | 无 | 输入 |
| KD | 范围 | [4.913,84.... | 无 | 无 | 输入 |
| W&R | 范围 | [0.0,100.0] | 无 | 无 | 输入 |
| 30组换手 | 范围 | [0.0483,5.... | 无 | 无 | 输入 |
| 30组涨幅 | 范围 | [-0.7548,0.... | 无 | 无 | 无 |
| 30组振幅 | 范围 | [0.062699.... | 无 | 无 | 无 |
| 30组总量 | 范围 | [129507.0.... | 无 | 无 | 输入 |
| 30组总额 | 范围 | [12034.35.... | 无 | 无 | 输入 |
| 45组换手 | 范围 | [0.0864,7.... | 无 | 无 | 输入 |
| 45组涨幅 | 范围 | [-0.80650.... | 无 | 无 | 无 |
| 45组振幅 | 范围 | [0.0726,1.... | 无 | 无 | 无 |
| 45组总量 | 范围 | [174964.0.... | 无 | 无 | 输入 |
| 45组总额 | 范围 | [15062.53.... | 无 | 无 | 输入 |
| 60组换手 | 范围 | [0.122400.... | 无 | 无 | 输入 |
| 60组涨幅 | 范围 | [-0.82129.... | 无 | 无 | 无 |
| 60组振幅 | 范围 | [0.0868,1.... | 无 | 无 | 无 |
| 60组总量 | 范围 | [247736.0.... | 无 | 无 | 输入 |
| 60组总额 | 范围 | [21356.66.... | 无 | 无 | 输入 |

然后带入模型计算，得到统计参数如下：

| [GB] 的分析 #51 | |
|----------------|--------|
| 文件 编辑 | |
| 折叠全部 展开全部 | |
| 输出字段 GB 的结果 | |
| 比较 \$N-GB 与 GB | |
| 最小误差 | -0.617 |
| 最大误差 | 1.058 |
| 平均误差 | -0.008 |
| 绝对平均误差 | 0.082 |
| 标准差 | 0.136 |
| 线性相关 | 0.853 |
| 发生率 | 567 |
| 分析 注解 | |
| 确定 | |

区分度算法：

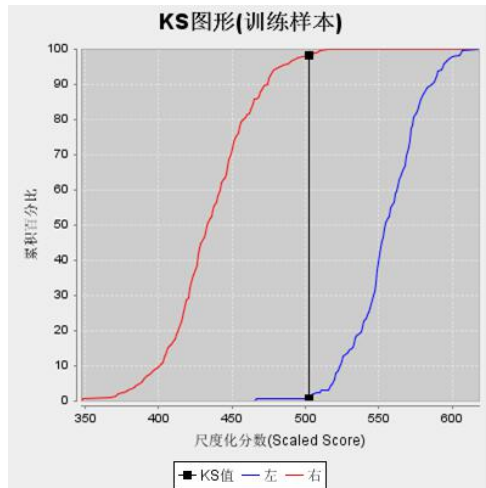
首先从大智慧对这个板块的所有股票抽取所需数据，然后保持到 excel 中，见文件：
stock.xls

然后剔除相关性较强的几个变量，见下表：
 然后带入模型计算，得到统计参数如下：

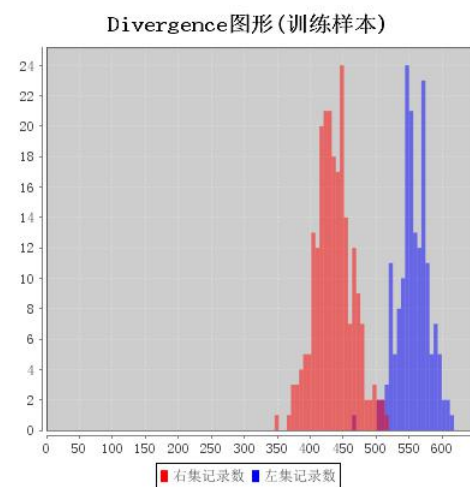
评分模型及主要统计量

| 样本类型 | 区分度(Divergence) | ROC面积 | Gini系数 | KS值 | KS 评分 / KS 尺度化评分 | 通过率(KS Pass Rate) | 判定系数(R-squared) |
|----------------|-----------------|-------|--------|--------|------------------|-------------------|-----------------|
| 训练样本(Training) | 21.112 | 0.999 | 0.998 | 97.658 | 1.128 / 502.000 | 43.073 | 0.722 |

K-s 值

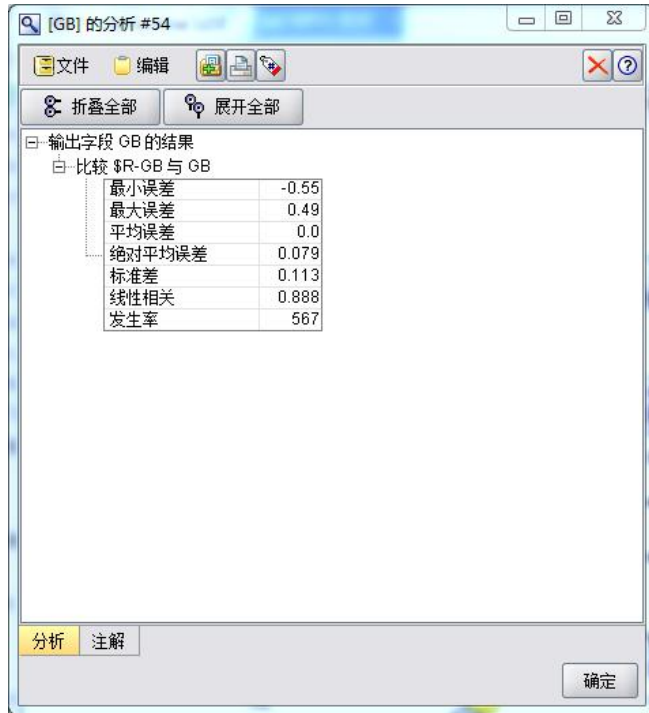


Divergence 图



决策树算法

数据处理方式同神经网络算法一致，直接看输出参数如下：



4: 相关文件

以上所有数据都放在 70 服务器的 zxh/stock/data 中

5: GB 的算法原理及算法设计

1: 算法原理

1.1: 首先把股价涨跌的原因分为外部因素和内部因素。

外部因素：外部因素有很多，但总体上还是反应在资金供求关系上，当流动性过剩，股市上扬，当流动性紧缩，股市下跌。详细论证可参看高盛高华的《股市与资金供求》报告。内部因素：内部因素也有很多，但总体上还是反应在财务数据上。当整体财务数据良好时，股价普涨，当整体财务数据不好时，股价普跌。

1.2: 在预测股价是否会上涨时首先剔除外部因素。

股市涨跌受到政治，经济，行业等诸多方面影响，由于外部因素有太多不可测的情况，所以在我们预测除外部因素影响外的股价是否会涨跌时，我们需要把这些不可测的情况剔除掉，只看公司本身是否创造了价值或者创造了多少价值来判断股价是否上涨，这就相对简单，并且也容易把握。

1.3: 如何剔除外部因素对股价的影响？

反映整体股市涨跌的上证指数或者深成指的涨跌代表了所有外部因素和内部因素导致股价涨跌的成因。所以在预测个股涨跌时我们剔除上证指数或深成指某段时间的涨跌幅时，其实就是近似的剔除了外部因素对个股的影响（大家下来自己想想看是不是这样的）。

1.4: 股票是否被低估或高估（市场平均水平）由他的内在价值或者说是内部驱动因素决定。

如何评价股票是否被低估或高估？

在剔除外部因素对股价的影响后，剩下的股价的涨跌就是由股票内在价值或者说是内部驱动因素造成的，当然，它也是市场行为决定的，但是，市场的行为未必是最优选择，他往往包含了非理性交易行为，这当中的隐含风险大部分人是不清楚的。而且即便全体股民的行为都是理性的交易行为，当这当中的潜在风险大部分人也是不清楚的。假设某股票内在价值高于平均市场行为涨跌幅（此时是指已经剔除外部因素对股价的影响），那么该股票就是被低估，如果我们总是能买入被市场行为远远低估的股票，那么无论市场处于理性或非理性行为，我们都能盈利。那要如何来评价每个股票的市场行为是否都合理的反应其内在价值呢？或者说如何量化每只股票的内在价值与其市场行为的涨跌（剔除外部因素对股价的影响）是否对应呢？

GB 的算法设计

2.1、计算剔除大盘的后每个基金在某个时间段的涨跌幅

当大盘为跌的时候：

```
aa:=count(1191105<date<1191229,0);
```

```
ll:=(c-ref(c,aa))/ref(c,aa);
```

```
zf:=ll+0.0348;
```

```
zf;
```

当大盘为涨的时候：

```
aa:=count(date>1190806,0);
```

```
ll:=(c-ref(c,aa))/ref(c,aa);
```

```
if ll>0 and ll<0.0226 then zf:=0.0226-ll;
```

```
else zf:=ll-0.0226;
```

```
zf;
```

本次 GB 的值-0.045 是统计从 20190806 统计到 20191124 上证指数的涨跌幅计算的。

2.2、计算超过或者低于平价涨跌幅的 GB（把 GB 变为逻辑变量后也可以做神经网络或者其他分类算法，比如逻辑回归，fish 判别，遗传算法，决策树，森林区分度等等）

```
aa:=count(date>1190806,0);
```

```
ll:=(c-ref(c,aa))/ref(c,aa);
```

```
if ll>-0.011 then gb:=1;
```

```
else gb:=0;
```

```
gb;
```

RSI 计算公式

```
REFLINE: 0, 20, 50, 80, 100;
```

```
aa:=count(date>1191203,0);
```

```
bb:=count(date>1191126,0);
```

```
LC := REF(CLOSE, bb);
```

```
LCA:=ref(close, aa);
```

```
RSI1:SMA(MAX(LCA-LC,0),N1,1)/SMA(ABS(LCA-LC),N1,1)*100;
```

```
RSI2:SMA(MAX(LCA-LC,0),N2,1)/SMA(ABS(LCA-LC),N2,1)*100;
```

RSI3:SMA(MAX(LCA-LC,0),N3,1)/SMA(ABS(LCA-LC),N3,1)*100;

kdj 计算公式

REFLINE: 0, 20, 50, 80, 100;

aa:=count(date>1191203,0);

RSV:=(ref(c,aa)-LLV(ref(c,aa),N))/(HHV(ref(c,aa),N)-LLV(ref(c,aa),N))*100;

K:SMA(RSV,M1,1);

D:SMA(K,M2,1);

J:3*K-2*D;

BIAS 计算公式

aa:=count(date>1191203,0);

c1:=ref(close,aa);

BIAS1 : (c1-MA(c1,L1))/MA(c1,L1)*100;

BIAS2 : (c1-MA(c1,L2))/MA(c1,L2)*100;

BIAS3 : (c1-MA(c1,L3))/MA(c1,L3)*100;

MACD 计算公式

aa:=count(date>1191203,0);

c1:=ref(close,aa);

DIFF : EMA(c1,SHORT) - EMA(c1,LONG);

DEA : EMA(DIFF,M);

MACD : 2*(DIFF-DEA), COLORSTICK;

7：风险预警及换股策略

8：剔除多重共线性后的模型情况

剔除多重共线性之前的数据情况如下

系数^a

| 模型 | | 共线性统计量 | |
|----|--------|--------|---------|
| | | 容差 | VIF |
| 1 | 最新 | .328 | 3.046 |
| | 总手 | .232 | 4.310 |
| | 每股收益 | .155 | 6.454 |
| | 市净率 | .807 | 1.239 |
| | 净利润 | .025 | 40.003 |
| | 每股净资产 | .158 | 6.329 |
| | 主营收入 | .156 | 6.422 |
| | 营业利润 | .015 | 68.927 |
| | 市盈率 | .984 | 1.016 |
| | 净资产收益率 | .450 | 2.223 |
| | 每股经营现金 | .641 | 1.559 |
| | 每股公积金 | .256 | 3.911 |
| | 流动负债 | .183 | 5.460 |
| | 长期负债 | .233 | 4.291 |
| | 股东权益 | .039 | 25.607 |
| | 资本公积金 | .183 | 5.470 |
| | 经营现金流量 | .224 | 4.457 |
| | 投资收益 | .668 | 1.497 |
| | 主营收入同比 | .923 | 1.084 |
| | 净利润同比 | .986 | 1.015 |
| | 股东权益比 | .733 | 1.364 |
| | 销售毛利率 | .810 | 1.235 |
| | 阶段换手 | .031 | 31.858 |
| | 阶段涨幅 | .260 | 3.852 |
| | 阶段振幅 | .178 | 5.614 |
| | 阶段成交量 | .009 | 109.272 |
| | 阶段成交额 | .014 | 70.875 |
| | 阶段换手 | .012 | 81.582 |
| | 阶段涨幅 | .183 | 5.469 |
| | 阶段振幅 | .130 | 7.680 |
| | 阶段成交量 | .008 | 120.793 |
| | 阶段成交额 | .005 | 198.914 |
| | 阶段换手 | .005 | 193.979 |
| | 阶段涨幅 | .099 | 10.144 |
| | 阶段振幅 | .070 | 14.287 |
| | 阶段成交量 | .002 | 418.371 |
| | 阶段成交额 | .002 | 503.745 |
| | 阶段换手 | .002 | 406.991 |
| | 阶段涨幅 | .072 | 13.898 |

剔除多重共线之后的数据情况如下

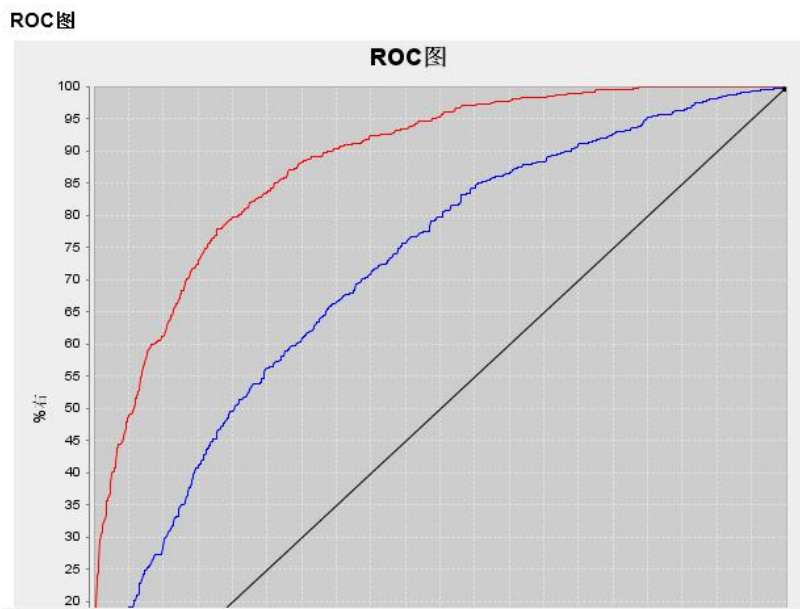
系数^a

| 模型 | | 共线性统计量 | |
|----|--------|--------|--------|
| | | 容差 | VIF |
| 1 | 最新 | .320 | 3.126 |
| | 总手 | .286 | 3.497 |
| | 每股收益 | .144 | 6.932 |
| | 市净率 | .784 | 1.276 |
| | 净利润 | .024 | 40.832 |
| | 每股净资产 | .164 | 6.104 |
| | 主营收入 | .155 | 6.469 |
| | 营业利润 | .014 | 69.542 |
| | 市盈率 | .982 | 1.018 |
| | 净资产收益率 | .410 | 2.437 |
| | 每股经营现金 | .634 | 1.577 |
| | 每股公积金 | .275 | 3.639 |
| | 流动负债 | .178 | 5.622 |
| | 长期负债 | .232 | 4.312 |
| | 股东权益 | .039 | 25.949 |
| | 资本公积金 | .178 | 5.624 |
| | 经营现金流里 | .220 | 4.555 |
| | 投资收益 | .666 | 1.503 |
| | 主营收入同比 | .923 | 1.083 |
| | 净利润同比 | .983 | 1.018 |
| | 股东权益比 | .727 | 1.376 |
| | 销售毛利率 | .813 | 1.231 |
| | 阶段换手 | .116 | 8.635 |
| | 阶段涨幅 | .400 | 2.497 |
| | 阶段振幅 | .394 | 2.535 |
| | 阶段成交量 | .065 | 15.287 |
| | 阶段成交额 | .046 | 21.828 |
| | 阶段换手 | .071 | 14.004 |
| | 阶段涨幅 | .664 | 1.506 |
| | 阶段振幅 | .409 | 2.442 |
| | 阶段成交量 | .052 | 19.371 |
| | 阶段成交额 | .029 | 34.708 |
| | 阶段换手 | .082 | 12.190 |
| | 阶段涨幅 | .628 | 1.592 |
| | 阶段振幅 | .373 | 2.679 |
| | 阶段成交量 | .056 | 17.942 |
| | 阶段成交额 | .046 | 21.636 |
| | 阶段换手 | .082 | 12.205 |
| | 阶段涨幅 | .522 | 1.917 |
| | 阶段振幅 | .282 | 3.543 |

剔除多重共线之前的模型统计量情况如下

评分模型及主要统计量

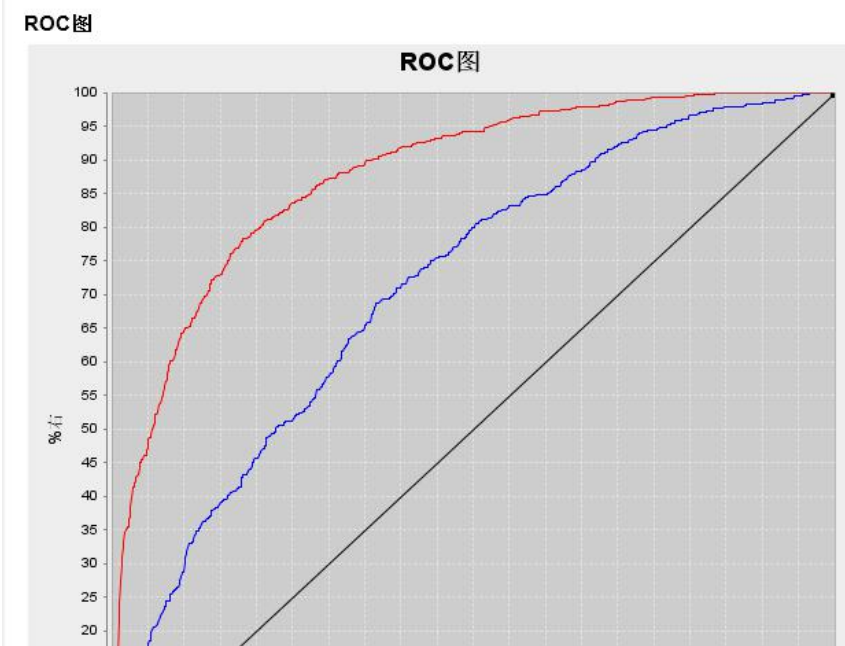
| 样本类型 | 区分度(Divergence) | ROC面积 | Gini系数 | KS值 | KS 评分 / KS |
|------------------|-----------------|-------|--------|--------|------------|
| 训练样本(Training) | 2.832 | 0.881 | 0.762 | 60.178 | -0.242 / 4 |
| 验证样本(Validation) | 0.620 | 0.715 | 0.430 | 31.892 | 0.079 / 4 |



剔除多重共线之后的模型统计量情况如下

评分模型及主要统计量

| 样本类型 | 区分度(Divergence) | ROC面积 | Gini系数 | KS值 | KS |
|------------------|-----------------|-------|--------|--------|----|
| 训练样本(Training) | 2.731 | 0.877 | 0.754 | 60.255 | |
| 验证样本(Validation) | 0.630 | 0.711 | 0.423 | 32.072 | |



这两个文件是最新一期最好模型的模型部署文件的 **java** 版本和 **python** 版

div-score 这个文件是针对 **python** 版的模型部署对每个股票进行区分度打分的程序

现在有一点小 **bug**，不过已经可以针对所有财务数据和技术指标打分并统计结果了，江恩周期的打分因为在做模型的时候软件会自动把变量名称改了，需要逐一比对，还在调试，不过明天程序可以全部出来。周日我会针对今天收盘的数据做一次计算并发到群里，我们可以在排名靠前的股票中挑选。

需要说明一点的是，之前做 **alpha** 组合的时候是挑选前 **20-50** 个股票一起买入，因为我们的模型并没有做到最优化，只能靠多选来抵消模型的不精确，如果只挑选 **4-10** 只股票就要好好研究个股的基本面。

买卖股票的具体原则：

1：买跌不买涨，卖涨不卖跌