量化选股研究与策略

X,S., C,X. and Y,Y

2017 年 2 月 28 日

目录

1 量化选股基础

1.1 因子的预处理

从各个数据源得到的因子数据称为原始因子。在构建Alpha策略前，要对截面原始因子进行一定的处理。

假设有k个行业（一般使用申万一级行业），每个行业均有m个股票(仅为计算方便如此假设)。用$f_{i,t}^j$表示t时刻第i个行业的j 支股票的因子。因为是针对截面数据做处理(t作为一个固定值)，下面的表示将隐去t角标，因子就以f_i^j表示。

1.1.1 中性化

一般来说，因子需要对行业进行中性化处理，以去除行业对因子的影响。方法是将原始因子对行业虚拟变量进行回归，取得回归的残差作为因子值。

根据假设，将这些股票的因子合并成一个$(km)\times 1$的向量, $f=(f_1^1,...f_i^j...f_k^m)^T, i=1,2...k, j=1,2,..m$, 对应的$(k\times 1)$ 回归系数向量$\beta=(\beta_1,...\beta_j...\beta_k)^T$。

回归变量可以表示为$(km)\times k$的矩阵$X=\left(u_1^1,...u_i^j...u_k^m\right)^T$, 其中$u_i^j=(0,..,1,0,...0)$ 为第i个元素为1总长度为k的单位行向量。

$$\begin{pmatrix} f_1^1 \\ f_1^2 \\ ... \\ f_i^j \\ ... \\ f_k^m \end{pmatrix}_{(km)\times 1} = \begin{pmatrix} 1 & 0 & ... & 0 \\ 1 & 0 & ... & 0 \\ ... & & & \\ 0 & 0 & 1_{(i-1)m+j,i} & 0 \\ ... & & & \\ 0 & ... & 0 & 1 \end{pmatrix}_{(km)\times k} \begin{pmatrix} \beta_1 \\ \beta_2 \\ ... \\ \beta_i \\ ... \\ \beta_k \end{pmatrix}_{k\times 1} + \begin{pmatrix} \varepsilon_1^1 \\ \varepsilon_1^2 \\ ... \\ \varepsilon_i^j \\ ... \\ \varepsilon_k^m \end{pmatrix}_{(km)\times 1} \tag{1}$$

公式(1)中的残差向量$(\varepsilon_1^1,...\varepsilon_k^m)$就是行业中性化后的因子

[

公式(1)的含义是，行业中性化处理等价于每个股票的因子减去行业内股票该类因子的平均值。因为对指示函数$\mathbf{1}_{industry}$回归，那么回归系数β等于所有因变量的均值。

]

还有一种中性化的方法是同时对行业以及风格（如市值对数）进行中性化，用$(km)\times 1$ 的向量$\left(\log M_1^1,\log M_1^2...\log M_k^m\right)^T$表示所有股票在该截面时刻的市值对数，原始因子对行业虚拟矩阵和市值对数向量同时做回归可得

$$\begin{pmatrix} f_1^1 \\ ... \\ f_i^j \\ ... \\ f_k^m \end{pmatrix}_{(km)\times 1} = \begin{pmatrix} 1 & ... & 0 \\ ... & & \\ 0 & 1_{(i-1)m+j,i} & 0 \\ ... & & \\ 0 & ... & 1 \end{pmatrix}_{(km)\times k} \begin{pmatrix} \beta_1 \\ ... \\ \beta_i \\ ... \\ \beta_k \end{pmatrix}_{k\times 1} + \gamma \begin{pmatrix} \log M_1^1 \\ ... \\ \log M_i^j \\ ... \\ \log M_k^m \end{pmatrix}_{(km)\times 1} + \begin{pmatrix} \varepsilon_1^1 \\ ... \\ \varepsilon_i^j \\ ... \\ \varepsilon_k^m \end{pmatrix}_{(km)\times 1} \tag{2}$$

公式(2)中的残差向量$(\varepsilon_1^1,...\varepsilon_k^m)$就是行业中性化以及市值中性化后的因子。在实际处理中通常要把公式(2)转换成公式(3)表示的形式：

$$\begin{pmatrix} f_1^1 \\ ... \\ f_i^j \\ ... \\ f_k^m \end{pmatrix}_{(km)\times 1} = \begin{pmatrix} 1 & ... & 0 & \log M_1^1 \\ ... & ... & ... & ... \\ 0 & 1_{(i-1)m+j,i} & 0 & \log M_i^j \\ ... & ... & ... & ... \\ 0 & ... & 1 & \log M_k^m \end{pmatrix}_{(km)\times(k+1)} \begin{pmatrix} \beta_1 \\ ... \\ \beta_k \\ \gamma \end{pmatrix}_{(k+1)\times 1} + \begin{pmatrix} \varepsilon_1^1 \\ ... \\ \varepsilon_i^j \\ ... \\ \varepsilon_k^m \end{pmatrix}_{(km)\times 1} \tag{3}$$

1.1.2 去极值化

因子序列中要去除某些极值（异常值）的影响，一般做法是将原始因子值调整到3倍绝对偏离中位数（或者是均值）的范围内。

1.1.3 标准化

对于原始因子序列$\{f_i, i = 1, ...N\}$, 用μ, σ表示这一序列的均值和标准差, 则标准化后的因子序列为

$$\{f_i^s = \frac{f_i - \mu}{\sigma}, i = 1, ...N\}$$

1.2 主成分分析(PCA)

[主成分分析原理详解，http://blog.csdn.net/yansmile1/article/details/46954089]

[主成分分析简介及其python实现，http://blog.csdn.net/yansmile1/article/details/46954089]

[Peter Harrington，《机器学习实战》，人民邮电出版社，2013]

主成分分析（Principal Component Analysis, PCA）是一种常见的数据降维方法，其目的是在信息量损失较小的前提下，将高维的数据转换到低维，从而减小计算量。实质就是找到一些投影方向，使得数据在这些投影方向上包含的信息量最大，而且这些投影方向是相互正交的。选择其中一部分包含最多信息量的投影方向作为新的数据空间，同时忽略包含较小信息量的投影方向，从而达到降维的目的。

样本的“信息量”可以理解为是样本在特征方向上投影的方差。方差越大，则样本在该特征上的差异就越大，因此该特征就越重要。参见《机器学习实战》上的图，在分类问题里，样本的方差越大，越容易将不同类别的样本区分开。

PCA的数学原理，就是对原始的空间中顺序地找一组相互正交的坐标轴，第一个轴是使得方差最大的，第二个轴是在与第一个轴正交的平面中使得方差最大的，第三个轴是在与第1、2个轴正交的平面中方差最大的，这样假设在N维空间中，可以找到N个这样的坐标轴，取前r 个去近似这个空间，这样就从一个N维的空间压缩到r维的空间了，但是最终选择的r个坐标轴能够使得数据的损失最小。

假设存在一组原始数据$\{z_1, z_2, ..., z_n\}$, 其中z_i为$p \times 1$维向量，即每个数据有p个特征。

(i) 去除平均值，即中心化，得到新的数据$\{x_1, x_2, ..., x_n\}$, $x_i = z_i - \mu$, $\mu = \frac{1}{n}\sum_{i=1}^{n} z_i$

(ii) 计算$\{x_i\}$协方差矩阵$\Sigma_{p\times p}$

(iii) 计算协方差矩阵的特征向量$\{\xi_j\}$和特征值$\{\lambda_j\}$, $j = 1..p$

(iv) 将特征值排序

(v) 保留前q个特征向量, 假设为$\{\lambda_j^*\}$, $j = 1..q$, 对应的特征向量构成的矩阵为$\Lambda = [\xi_1^* .. \xi_q^*]_{p\times q}$

(vi) 将数据转换到上述N个特征向量构建的新的空间中, 得到新的数据集$X^* = X\Lambda = \begin{pmatrix} x_1 \\ x_2 \\ ... \\ x_i \\ ... \\ x_n \end{pmatrix}_{n\times p} [\xi_1^*..\xi_q^*]_{p\times q}$

数学原理请见附录(3.1)。

1.3 奇异值分解(SVD)

[从PCA和SVD的关系拾遗，http://blog.csdn.net/dark_scope/article/details/53150883]

[Jonathon Shlens，A Tutorial on Principal Component Analysis, https://arxiv.org/pdf/1404.1100.pdf, 2014]

1.3.1 概述

奇异值分解是一个能适用于任意矩阵(如非方阵)的一种分解的方法。对于$m \times n$的矩阵X:

$$X = U\Sigma V^T$$

其中U为$m \times m$正交阵，Σ为$m \times n$非负实对数对角矩阵，V为$n \times n$正交阵。

[构造过程可以描述如下：

X为$m \times n$矩阵，假设X^TX为秩为r的$n \times n$实对称阵。

设$\{v_1, \ldots, v_r\}$为X^TX的标准正交特征向量，v_i为$(n \times 1)$的列向量，对应特征值为$\{\lambda_1, \ldots, \lambda_r\}$，根据特征值定义有

$$(X^TX)v_i = \lambda_i v_i$$

由于X^TX为半正定矩阵，$\lambda_i \geq 0$，$i = 1, ...r$。令$\sigma_i = \sqrt{\lambda_i}$，可知$\sigma_i$为非负实数，称为奇异值。

用$\{u_1, \ldots, u_r\}$表示一组列向量，其中u_i为$(m \times 1)$的列向量，且

$$u_i = \frac{1}{\sigma_i} X v_i \tag{4}$$

由定理(8)可知

$$< u_i, u_j >= \begin{cases} 1 & \text{i=j} \\ 0 & \text{i≠j} \end{cases}$$

$$||Xv_i|| = \sigma_i$$

由公式(4)可知

$$Xv_i = \sigma_i u_i \tag{5}$$

令$\Sigma = \begin{pmatrix} \sigma_1 & 0 & \ldots & 0 \\ \vdots & \ddots & \ldots & \vdots \\ \vdots & 0 & \sigma_r & \vdots \\ \vdots & \ldots & \ldots & \vdots \\ 0 & \ldots & \ldots & 0 \end{pmatrix}$，$\sigma_1 \geq \sigma_2 \ldots \geq \sigma_r$

$$V = (v_1, v_2, \ldots, v_n)$$

$$U = (u_1, u_2, \ldots, u_n)$$

$v_{r+1}, \ldots, v_n$为X^TX的特征值为0对应的特征向量。$u_{r+1}, \ldots, u_n$通过4计算得到的向量。
故用矩阵表示公式(5)为$XV = U\Sigma$
因为V为正交阵, 所以$V^{-1} = V^T$
$\Rightarrow X = U\Sigma V^T$
]

1.3.2 性质与用途

在X的奇异值分解中

(i) V的列向量组成一套对X的正交“输入”或“分析”基向量，这些向量是X^TX的特征向量

(ii) U的列向量组成一套对X的正交“输入”基向量，这些向量是XX^T的特征向量

(iii) Σ对角线上的元素是奇异值，可视为输入与输出间进行的标量的“膨胀控制”。这些是XX^T及X^TX的特征值的非零平方根，并与U和V的行向量对应。

基于这些性质，可以得到定理(1)，即X^TX的奇异值分解的结果和特征值分解一致。证明见附录(3.2)。由此可以得到以下结论

(i) 根据定理(1)，可以使用$X^TX = U\Sigma V^T$奇异值分解得到的Σ和U作为X^TX的特征值和特征向量矩阵

(ii) 根据性质1，也可以使用$X = U\Sigma V^T$奇异值分解得到的Σ^2和V作为X^TX的特征值和特征向量矩阵

从章节(1.2)可知，PCA的计算中需要对X^TX(等价于协方差矩阵, 为表述方便，此处X^T就是(1.2)中的X，每一行是一个样本)进行特征值分解，通过奇异值分解可以得到同样的结果，并且从数值计算的角度考虑，奇异值分解的效果更好：没有计算X^TX这一步，而矩阵中一些非常小的数容易在平方中丢失，另外奇异值分解的速度更快。

1.4 策略表现的指标计算

1.4.1 对冲组合的净值计算方法

一个对冲组合的净值计算主要涉及到调仓的设定。此处调仓是指使得组合的股票多头和对应期货（以指数代替）的空头市值相同，即完全对冲。在下一个调仓日之前，组合的股票多头和期货空头数量不再调整，随着市场的波动，组合会出现多空头寸的暴露。

假设有$t = 0, 1, 2, ...N$共$(N+1)$个交易日，调仓日为$\{t_i^*, i = 1, 2...k\}$共k个。调仓时间为调仓日的收盘前。

用$\{r_t^s\}$和$\{r_t^f\}$表示策略和对冲标的的日简单收益序列。用V_t, V_t^s, V_t^f表示t时刻的组合价值，组合中股票账户市值，以及期货账户市值。

假设组合的收益为累计复利计算方式，另外假设组合中期货账户占用资金比例为p（用于保证金账户）,那么组合中$1-p$的部分可投资于股票账户。对冲组合的净值NPV_t变化可以演示为

– 在$t = 0$时刻建立组合，该时刻也是第一个调仓日。此时$V_0^s = V_0^f$, $NPV_0 = 1.0$

- 在$t=1$时刻，组合的价值变动为$V_0^s(1+r_1^s)-V_0^f(1+r_1^f)$，此时

$$\mathrm{NPV}_1 = NPV_0\left(1+(1-p)[(1+r_1^s)-(1+r_1^f)]\right)$$

[

Proof. 根据股票账户以及期货账户的资金比例假设以及$V_0^s=V_0^f$可知

$$\begin{aligned}\mathrm{NPV}_1 &= \mathrm{NPV}_0\left(1+\frac{V_0^s(1+r_1^s)-V_0^f(1+r_1^f)}{V_0}\right)\\ &= \mathrm{NPV}_0\left(1+\frac{V_0(1-p)(1+r_1^s)-V_0(1-p)(1+r_1^f)}{V_0}\right)\\ &= \mathrm{NPV}_0\left(1+(1-p)[(1+r_1^s)-(1+r_1^f)]\right)\end{aligned}$$

□

]

- 在下一个调仓日之前(包括调仓日)的任意$t=n$时刻，净值为

$$\mathrm{NPV}_n = \mathrm{NPV}_0\left(1+(1-p)[\prod_{i=1}^{n}(1+r_i^s)-\prod_{i=1}^{n}(1+r_i^f)]\right) \tag{6}$$

- 在第一个调仓日t_1^*之后，$t=t_1^*+1$ 时刻，净值为

$$\mathrm{NPV}_{t_1^*+1} = NPV_{t_1^*}\left(1+(1-p)[(1+r_{t_1^*+1}^s)-(1+r_{t_1^*+1}^f)]\right) \tag{7}$$

从公式(6)和公式(7)可以得出以下结论

(i) 任意两个调仓日t_i^*, t_{i+1}^*之间，对冲组合的净值为

$$\mathrm{NPV}_{t_1^*+n} = \mathrm{NPV}_{t_1^*}\left(1+(1-p)[\prod_{i=1}^{n}(1+r_{t_1^*+i}^s)-\prod_{i=1}^{n}(1+r_{t_1^*+i}^f)]\right) \tag{8}$$

(ii) 期货保证金账户的资金占用降低了组合净值的变化。假设期货保证金占用为零，此时对冲组合的净值为

$$\mathrm{NPV}_{t_1^*+n} = \mathrm{NPV}_{t_1^*}\left(1+[\prod_{i=1}^{n}(1+r_{t_1^*+i}^s)-\prod_{i=1}^{n}(1+r_{t_1^*+i}^f)]\right) \tag{9}$$

(iii) 有一种特殊的情况为每日调仓，保证每日结算时多空敞口为零，此种情况下

$$\mathrm{NPV}_{n+1} = \mathrm{NPV}_n\left(1+[(1+r_{n+1}^s)-(1+r_{n+1}^f)]\right) \tag{10}$$

2 量化选股策略

2.1 动态情景多因子策略

[朱建涛，东方证券，动态情景多因子Alpha模型，2016]

2.1.1 概述

传统多因子模型大多是全市场范围对所有股票一视同仁打分评价，而忽视了个股的基本面情况差异以及选股因子对于不同风格股票组合的适用性。原文的动态情景多因子模型(Dynamic Contextual Alpha Model)将全市场的股票按照规模、估值、成长等因子进行了划分，在此基础上对不同分档的股票组合中对个股再赋予一定权重；通过这一方法大幅提升模型对于市场风格切换的适应能力。

2.1.2 算法

假设调仓频率为月度，调仓时点为每个月末。假设有N个分层因子和M个α因子。

(i) 在调仓日根据风格因子(可称为分层因子)将股票池分组：方法为因子值从高到低排列，对半分组。每个分层因子对应一个股票池的分组

风格维度	因子名称和含义
规模	总市值
价值	最近财报的净资产/总市值
成长	净资产同比增长率
盈利	净资产收益率
流动性水平	季度日均换手率

表 1: 分层因子表

(ii) 在不分层的情况下，在调仓日对每个α因子进行去极值化和标准化处理, 参见(1.1.2)和(1.1.3)

(iii) 在不分层的情况下，在调仓日对每个α因子和标的阶段收益率R进行行业和市值中性化处理（原文中称为风险调整），参见(1.1.1)。

$f_{i,t}$表示调仓日t时刻的原始因子i向量，R_t表示调仓日$t-1$到t之间的股票收益向量，M_t为调仓日t时刻的市值向量，X 为行业虚拟矩阵，参见公式(2)。调整后的因子和收益向量为

$$f_t^a = f_t^i - X\beta_1 - \beta_2 * \ln M_t$$
$$R_t^a = R_t - X\beta_1 - \beta_2 * \ln M_t$$

(iv) 在每个分层空间内，对每个α 因子计算风险调整IC, 用$\mathrm{IC}_{i,t}^a$表示调仓日t时刻的基于因子i的风险调整因子

$$\mathrm{IC}_{i,t}^a = \mathrm{Corr}\left(f_{i,t-1}^a, R_t^a\right)$$

(v) 用过去十二个月（不包括调仓的月末，因为需要预测的是下个月的收益率）的各个α因子在分层空间内的风险调整IC，计算因子加权矩阵。用$\omega_{i,t}$ 表示在某分层中，因子i在t时刻的权重

$$\omega_{i,t} = \frac{\mu(\mathrm{IC}^a_{i,t-12}, \mathrm{IC}^a_{i,t-11} ... \mathrm{IC}^a_{i,t-1})}{\sigma(\mathrm{IC}^a_{i,t-12}, \mathrm{IC}^a_{i,t-11} ... \mathrm{IC}^a_{i,t-1})} \tag{11}$$

其中，分子为因子的均值，分母为因子的标准差。

(vi) 使用月末的风险调整后的因子值，在分层空间内对股票进行排序打分。

当因子加权矩阵值为正时，按照从小到大排序，表示因子值越大越好，因子值越大得分越高；当因子加权矩阵值为负时，按照从大到小排序，表示因子值越小越好，因子值越小得分越高。

(vii) 在每个分层内，计算每个股票的加权得分。用$\omega_{i,c,t}$表示在t时刻，根据公式(11)计算的在分层因子c空间内的因子i的权重；用$s^j_{i,c,t}$ 表示在t 时刻, 股票j 在c 分层因子空间内的α 因子i 的排序分数；那么该时刻在该分层因子下股票j的α因子加权得分可表示为

$$s^j_{\alpha,c,t} = \sum_{i=1}^{M} s^j_{i,c,t} w_{i,c,t}$$

(viii) 另外还需要对股票在不同情景分层上的属性进行定量刻画（比如全市场估值最高的10%和45%的股票可能存在一些区别）。原文中根据在分层中的分位数定义了一个距离函数来描述股票的情景相似度，此处本文定义该函数为一拉伸过的sigmoid函数，使得函数曲线更平滑。

用$d^j_{c,t}$表示在t时刻股票j的分层因子c的情景得分，p为该股票在分层因子c划分的分层空间内排序的百分比

$$d^j_{c,t} = 18\left(\frac{1}{1+e^{-10(p-0.5)}} - 0.5\right)$$

(ix) 最后计算t时刻股票j的总得分为

$$\gamma^j_t = \sum_{c=1}^{N} |d^j_{c,t}| s^j_{\alpha,c,t} \tag{12}$$

根据公式(12)计算的个股总得分进行由高到低的排序和筛选。此处有两点需要注意

(a) 公式(12)中的距离函数取绝对值，含义是策略风格中性: 如分层因子中特别小的和特别大的距离函数值应该相同

(b) 如果是行业中性策略则还需要根据标的指数的行业比例进行选股

3 附录

3.1 主成分分析的相关推导

假设存在一组原始数据$\{z_1, z_2, ..., z_n\}$, 其中z_i为$p \times 1$向量，即每个数据有p个特征。
首先，将数据中心化$\{x_1, x_2, ..., x_n\}$, $x_i = Z_i - \mu$, $\mu = \frac{1}{n}\sum_{i=1}^n z_i$
用$u_{p\times 1}$表示某投影方向上的单位向量，那么x_i在u上的投影可以表示为

$$< x_i, u >= (x_i^T u)$$

令$X = (x_1, x_2, ...x_n)_{p\times n}$，那么所有数据在该方向上的投影的方差为$\frac{1}{n}u^T XX^T u$。
[

Proof. 根据方差的定义，可得投影的方差

$$\frac{1}{n}\sum_{i=1}^n (x_i^T u)^2 - [\frac{1}{n}\sum_{i=1}^n (x_i^T u)]^2 \tag{13}$$

公式(13)中的第二项为零，因为

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n x_i^T u = \frac{1}{n}\sum_{i=1}^n x_i^T u &= \frac{1}{n}\sum_{i=1}^n (z_i^T - \mu^T)u \\ &= \frac{1}{n}(\sum_{i=1}^n z_i^T - n\mu^T)u \\ &= \frac{1}{n}(\sum_{i=1}^n z_i^T - \sum_{i=1}^n z_i^T)u \\ &= 0 \end{aligned} \tag{14}$$

第一项可以简化为

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n (x_i^T u)^2 &= \frac{1}{n}\sum_{i=1}^n (x_i^T u)^T (x_i^T u) \qquad \left(x_i^T u\text{是一个数}\right) \\ &= \frac{1}{n}\sum_{i=1}^n u^T x_i x_i^T u \qquad (u\text{与i无关}) \\ &= \frac{1}{n}u^T \left(\sum_{i=1}^n x_i x_i^T\right) u \end{aligned} \tag{15}$$

因为$X = (x_1, x_2, ...x_n)_{p\times n}$，则$X^T = \begin{pmatrix} x_i^T \\ \vdots \\ x_n^T \end{pmatrix}_{n\times p}$

那么

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n (x_i^T u)^2 - [\frac{1}{n}\sum_{i=1}^n (x_i^T u)]^2 &= \frac{1}{n}u^T \left(\sum_{i=1}^n x_i x_i^T\right) u \qquad (\text{根据公式 }(14)\text{和公式}(15)) \\ &= \frac{1}{n}u^T XX^T u \qquad (\text{根据X的定义})\end{aligned}$$

实际上方差矩阵省略$\frac{1}{n}$对特征向量没有影响，对特征值大小关系也没有影响。

□

]

根据定理(9), XX^T是半正定对称矩阵, 所有特征值均大于等于零。
作为对称矩阵，根据定理(8), XX^T的特征向量相互正交，构成了空间的正交基。
假设对称矩阵$XX^T \in C^{p\times p}$的p 个特征值为

$$\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p \geq 0$$

对应的单位特征向量为$\xi_1, \xi_2, ..., \xi_p$。这些特征向量表示投影方向的单位向量u, 假设$u = \sum_{i=1}^{p} \alpha_i \xi_i$

$$\begin{aligned}
u^T XX^T u =< u, XX^T u > &= < \sum_{i=1}^{p} \alpha_i \xi_i, \sum_{i=1}^{p} \alpha_i XX^T \xi_i > \\
&= < \sum_{i=1}^{p} \alpha_i \xi_i, \sum_{i=1}^{p} \alpha_i \lambda_i \xi_i > \\
&= \lambda_1 \alpha_1^2 + ... + \lambda_n \alpha_n^2 \qquad \big(\text{单位特征向量}\xi_i, \xi_j\text{正交}, i \neq j\big) \\
&\leq \lambda_1(\alpha_1^2 + ... + \alpha_n^2) \\
&= \lambda_1 \|u\| \\
&= \lambda_1 \qquad \big(u\text{为单位向量}, \|u\| = 1\big)
\end{aligned}$$

故$u^T XX^T u \leq \lambda_1$
综上所述，数据集合$X = (x_1, x_2, ...x_n)_{p\times n}$的投影方差最大为$\frac{\lambda_1}{n}$, 该投影方向为$\lambda_1$对应的特征向量的方向。

3.2 奇异值分解的相关推导

性质1和3: V的列向量是$X^T X$的特征向量，Σ对角线上的元素是$X^T X$的特征值的非零平方根，并与V的行向量对应。

Proof. 由奇异值分解定义

$$X = U\Sigma V^T$$

则

$$\begin{aligned}
X^T X &= V\Sigma U^T U\Sigma V^T \\
&= V\Sigma^2 V^T \\
&= V\Sigma^2 V^{-1} \qquad (16)
\end{aligned}$$

又因为$X^T X$为实对称矩阵，可得

$$X^T X = Q\Lambda Q^{-1} \qquad (17)$$

公式(16)和(16)形式一致，当限定特征值顺序后，这样的组合是唯一的。所以V是$X^T X$的特征向量且奇异值和特征值是平方关系

$$V = Q$$

$$\Lambda = \Sigma^2$$

□

性质2的证明同理可得。

Theorem 1. X^TX的奇异值分解结果和特征值分解结果一致，即奇异值分解的矩阵U等于特征向量矩阵，Σ等于特征值对角矩阵。

Proof. 对X^TX进行SVD分解(为加以区分，下标为2）

$$X^TX = U_2\Sigma_2V_2^T$$

由性质2，U_2为X^TXX^TX的特征向量

$$\begin{aligned} X^TXXX &= U_2\Sigma_2V_2^T(U_2\Sigma_2V_2^T)^T \\ &= U_2\Sigma^2U_2^T \\ &= U_2\Sigma^2U_2^{-1} \end{aligned} \tag{18}$$

根据特征值分解, 用Q_2表示特征向量矩阵，Σ_2表示特征值对角矩阵，可得

$$X^TX = Q_2\Lambda_2Q_2^{-1}$$

进一步可知

$$X^TXX^TX = Q_2\Lambda_2^2Q_2^{-1} \tag{19}$$

由(18)(19)可知$U_2 = Q_2$，$\Sigma^2 = \Lambda^2$。

有这样的结果是因为X^TX为半正定矩阵。 □

3.3 线性代数的重要定理

3.3.1 矩阵的特征值与特征向量

Theorem 2. n阶矩阵A可对角化的充要条件是A有n个线性无关的特征向量。

Proof. 必要性：

若A可对角化，则∃ 可逆矩阵P，使得

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \tag{20}$$

将矩阵P用列向量表示$P = (x_1, \ldots, x_n)$，x_i为列向量, 根据公式(20)

$$\begin{aligned} AP &= A\,(x_1, \ldots, x_n) \\ &= (x_1, \ldots, x_n)\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \\ \Rightarrow Ax_i &= \lambda_i x_i \quad \text{i=1,...,n} \end{aligned}$$

所以x_i是A属于特征值λ_i的特征向量。又因为P可逆，故$(x_1, \ldots, x_n)$线性无关。故A有n个线性无关的特征向量。

充分性：

设$(x_1, \ldots, x_n)$是A的n个线性无关的特征向量，对应特征值为$(\lambda_1, \ldots, \lambda_n)$，则有$Ax_i = \lambda x_i$, i=1,...,n

令$P=(x_1,\dots,x_n)$, 则P是一个可逆矩阵。

$$\begin{aligned} AP &= A(x_1,\dots,x_n) \\ &= (\lambda_1 x_1,\dots,\lambda_n x_n) \\ &= (x_1,\dots,x_n)\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \\ \Rightarrow P^{-1}AP &= \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \qquad (\text{因为}P=(x_1,\dots,x_n)\text{可逆}) \end{aligned}$$

□

Theorem 3. 一个矩阵的属于不同特征值的特征向量线性无关

3.3.2 实对称矩阵

Definition 1. 若A为实对称矩阵，则A满足$\overline{A}=A$且$(\overline{A})^T=A^T=A$

Theorem 4. 实对称矩阵的特征值都是实数

Theorem 5. 实对称矩阵的属于不同特征值的特征向量是正交的

Proof. 设λ_1和λ_2是实对称矩阵A的两个不同特征值，ξ_1,ξ_2分别是属于λ_1,λ_2的特征向量，则有

$$\begin{aligned} \lambda_1<\xi_1,\xi_2> &= <\lambda_1\xi_1,\xi_2> \\ &= <A\xi_1,\xi_2> \quad (\text{根据定义}A\xi_1=\lambda_1\xi_1) \\ &= (A\xi_1)^T\xi_2 \\ &= \xi_1^T A^T \xi_2 \\ &= \xi_1^T A \xi_2 \qquad \left(A\text{为实对称矩阵，}A^T=A\right) \\ &= \xi_1^T \lambda_2 \xi_2 \qquad (\text{根据定义}A\xi_2=\lambda_2\xi_2) \\ &= \lambda 2<\xi_1,\xi_2> \\ \Rightarrow (\lambda_1-\lambda_2) &< \xi_1,\xi_2>=0 \end{aligned}$$

因为$\lambda_1\neq\lambda_2$, 所以$<\xi_1,\xi_2>=0$, ξ_1和ξ_2正交。 □

Theorem 6. 设A为n阶实对称矩阵，λ是A的特征方程的r重根，则$(A-\lambda I)$的秩为n-r，从而对应特征值λ恰有r个线性无关的特征向量

Theorem 7. 设A为n阶实对称矩阵，则必存在正交阵P，使得

$$P^{-1}AP=\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

，其中$\lambda_1,...,\lambda_n$为A的n个特征值。

Proof. 设A互不相同的特征值为$\lambda_1,...,\lambda_s$，它们的重数依次为$r_1,...,r_s$，则有$r_1+...+r_s=n$

由定理(4)和(6)可知：λ_i恰有r_i个线性无关的实特征向量，将其标准正交化就可以得到r_i个单位正交的特征向量组，由$r_1+...+r_s=n$可知，这样的特征向量共有n个。

又由定理(5)可知，A的属于不同特征值的特征向量是正交的，故这n个单位特征向量两两正交，以它们为列向量构成正交矩阵P并有

$$P^{-1}AP=\Lambda=\begin{pmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{pmatrix}$$

□

Theorem 8. 对于任意$m\times n$实矩阵X, 实对称阵X^TX有一组标准正交特征向量$\xi_1,...,\xi_n$和对应的特征值$\lambda_1,...,\lambda_n$, $\{X\xi_1,X\xi_2,...,X\xi_n\}$构成了一组标准正交基，且$||X\xi_i||=\sqrt{\lambda_i}$，*i=1,2,...,n*。

Proof. X^TX为实对称矩阵，根据定理(7), 其必存在一组标准正交特征向量$\xi_1,...,\xi_n$和对应的特征值$\lambda_1,...,\lambda_n$

$$\begin{aligned}<X\xi_i,X\xi_j> &= (X\xi_i)^TX\xi_j\\ &= \xi_i^TX^TX\xi_j\\ &= \xi_i^T\lambda_j\xi_j \qquad \Big(\text{根据特征值定义，}X^TX\xi_j=\lambda_j\xi_j\Big)\\ &= \lambda_j<\xi_i,\xi_j>\\ &= \begin{cases}\lambda_j & \text{i=j}\\ 0 & \text{i}\neq\text{j}\end{cases} \qquad \Big(\text{根据定理(5)，不同特征值的特征向量正交}\Big)\end{aligned}$$

所以$X\xi_1,...,X\xi_2$ 构成一组标准正交基，且$||X\xi_i||=\sqrt{\lambda_i}$

□

Theorem 9. 对于任意$m\times n$实矩阵X, 实对称阵X^TX为半正定矩阵

Proof. 假设X^TX某特征值为λ, 对应特征向量为ξ

由特征向量的定义可知$(X^TX)\xi=\lambda\xi$

两边转置后同乘以ξ有

$$(X^TX\xi)^T\xi=(\lambda\xi)^T\xi=\lambda\xi^T\xi \tag{21}$$

又因为

$$(X^TX\xi)^T\xi=\xi^TX^TX\xi=(X\xi)^T(X\xi)=\|X\xi\|^2 \tag{22}$$

由公式(21)和公式(22)可知

$$\begin{aligned}\|X\xi\|^2 &= \lambda\xi^T\xi\\ &= \lambda\|\xi\|^2\geq 0\\ &\Longrightarrow \lambda\geq 0\\ &\Longrightarrow X^TX\text{半正定}\end{aligned}$$

□