协方差矩阵特性

rinterested

2020/12

1 Gramian 矩阵

矩阵 $A^\top A$(Gramian 矩阵) 具有以下性质：

- $A^\top A$ 是一个关键的矩阵结构，因为它在正交投影中起着重要的作用。协方差矩阵只是特例。
- $A^\top A$ 是协方差矩阵—你可以定义多元正态分布，其中 $A^\top A$ 是协方差矩阵，参见这里。
- 这相当于讨论对称半正定矩阵 (symmetric positive semidefinite matrices, s.p.s.d.)—对于某些矩阵 A，每个对称半正定矩阵都可以写成 $A^\top A$。

特性列表：

1. 对称性
2. 半正定性 (可为零)
3. 实特征值和正特征值
4. 矩阵迹 (trace) 为正 (矩阵迹为特征值之和)
5. 行列式是正的 (行列式是特征值的乘积)
6. 对角线条目都是正数
7. 正交特征向量
8. 可对角化为 $Q\Lambda Q^T$
9. 可以得到 Cholesky 分解。
10. $A^\top A$ 的秩与 A 的秩相同。
11. $\ker(A^\top A) = \ker(A)$

2 协方差矩阵

如果列向量的条目：

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$$

是具有有限方差的随机变量，则协方差矩阵 Σ 是其 (i,j) 项为协方差的矩阵

$$\Sigma_{ij} = cov(X_i, X_j) = E\left[(X_i - \mu_i)(X_j - \mu_j)\right] = E[X_i, X_j] - E[X]E[Y]$$

其中 $\mu_i = E(X_i)$ 是向量 X 中第 i 项的期望值。换句话说，

$$\Sigma = \begin{bmatrix} E\left[(X_1-\mu_1)(X_1-\mu_1)\right] & E\left[(X_1-\mu_1)(X_2-\mu_2)\right] & \cdots & E\left[(X_1-\mu_1)(X_n-\mu_n)\right] \\ E\left[(X_2-\mu_2)(X_1-\mu_1)\right] & E\left[(X_2-\mu_2)(X_2-\mu_2)\right] & \cdots & E\left[(X_2-\mu_2)(X_n-\mu_n)\right] \\ \vdots & \vdots & \ddots & \vdots \\ E\left[(X_n-\mu_n)(X_1-\mu_1)\right] & E\left[(X_n-\mu_n)(X_2-\mu_2)\right] & \cdots & E\left[(X_n-\mu_n)(X_n-\mu_n)\right] \end{bmatrix}$$

对于具有均值向量 μ 的随机向量 $\mathbf{X} \in \mathbb{R}^n$，更简洁的定义是 $\mathbb{E}\big((\mathbf{X}-\mu)(\mathbf{X}-\mu)^\top\big)$。

这与维基百科的另一个定义是一致的：

$$\Sigma = E\left[\left(X - E[X]\right)\left(X - E[X]\right)^\top\right]$$

从这篇文章和另一篇文章可知：当数据居中 (零均值) 时，协方差矩阵为 $\frac{1}{n-1}\mathbf{X}\mathbf{X}^\top$。

因为协方差矩阵是对称的，所以矩阵是可对角化的，并且特征向量可以归一化，使得它们是正交的：

$$\mathbf{X}\mathbf{X}^\top = \mathbf{W}\mathbf{D}\mathbf{W}^\top$$

另一方面，对数据矩阵 $\mathbf{X}$ 应用 SVD 如下：

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$$

同时尝试从这个分解构造协方差矩阵得到

$$\mathbf{X}\mathbf{X}^\top = (\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top)(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top)^\top$$
$$\mathbf{X}\mathbf{X}^\top = (\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top)(\mathbf{V}\mathbf{\Sigma}\mathbf{U}^\top)$$

并且因为 $\mathbf{V}$ 是一个正交矩阵 $(\mathbf{V}^\top\mathbf{V} = \mathbf{I})$，

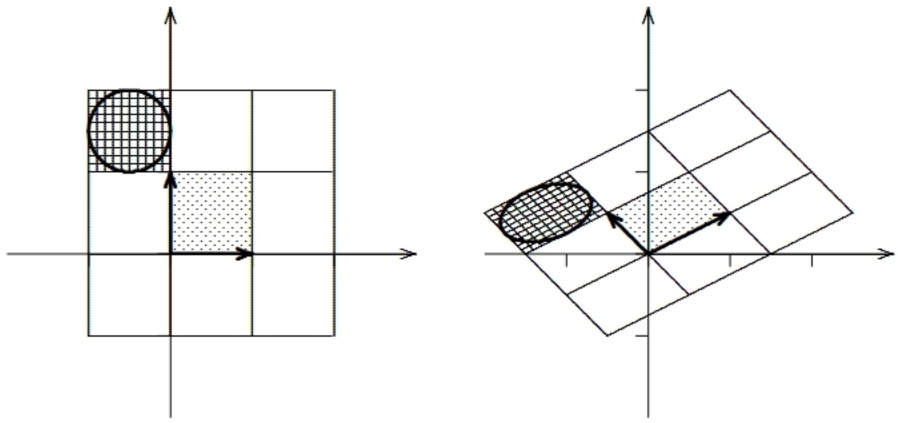
$$\mathbf{X}\mathbf{X}^\top = \mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^\top$$

并且相关对应很容易看出 ($\mathbf{X}\mathbf{X}^\top$ 的特征值的平方根是 $\mathbf{X}$ 的奇异值，等等)。

3 几何解释

正如单变量方差是平均值的平均平方距离一样，$\mathrm{trace}(\hat{\mathbf{\Sigma}})$ 是到质心的平均平方距离：以 $\dot{\mathbf{X}}$ 做为中心变量的矩阵，则 $\hat{\mathbf{\Sigma}} = \frac{1}{n}\dot{\mathbf{X}}'\dot{\mathbf{X}}$，其中 $\dot{\mathbf{X}}'\dot{\mathbf{X}}$ 是 $\dot{\mathbf{X}}$ 列的点积矩阵。其对角线元素为 $\dot{\mathbf{X}}'_{\cdot i}\dot{\mathbf{X}}_{\cdot i} = (\mathbf{X}_{\cdot \mathbf{i}} - \overline{\mathbf{X}}_{\cdot \mathbf{i}})'(\mathbf{X}_{\cdot \mathbf{i}} - \overline{\mathbf{X}}_{\cdot \mathbf{i}})$，即变量 i 与其平均值的平方距离。因此，$\mathrm{trace}(\hat{\mathbf{\Sigma}})$ 是单变量方差的自然推广。

第二个推广是 $\det(\hat{\boldsymbol{\Sigma}})$：这是描述分布的椭球体体积的度量。更准确地说，$|\det(\hat{\boldsymbol{\Sigma}})|$ 是应用线性变换 $\hat{\boldsymbol{\Sigma}}$ 后单位立方体体积变化的因子，(见此解释)。以下是行列式为 0.75 的矩阵 $\left(\begin{smallmatrix}1 & -.5\\ .5 & .5\end{smallmatrix}\right)$ 的图示 (左：变换前，右：变换后)：



你可以试试矩阵的 Frobenius 范数 (Frobenius norm)，它基本上是 $\mathrm{trace}(\Sigma^{\mathrm{T}}\Sigma)$。要选择阈值，可以在信息增益表达式中添加正则化 (Regularization) 项。

4 样本协方差与相关性

样本协方差矩阵为：

$$\mathbf{S} = \frac{1}{n-1}\sum_{j=1}^{n}(\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})^T$$

对于 $i = 1, \ldots, n$，标准化随机变量的协方差矩阵为 $X_i/\sigma(X_i)$。

在 $G^T G$ 运算中，G^T 是一个 6×8 矩阵，G 是一个 8×6 矩阵。因此，矩阵乘法将产生一个 6×6 矩阵。

针对评论和根本问题，让我们假设我们有一个对应于不同股票 (数据列中) 与连续 5 年 (数据行中) 回报率的矩阵—完全虚构的股票和年份。我们把矩阵称为 A：

$$A = \begin{array}{cc} & \begin{array}{ccc} \text{yah(y)} & \text{goog(g)} & \text{ms(m)} \end{array} \\ \begin{array}{c} \text{Yr.1} \\ \text{Yr.2} \\ \text{Yr.3} \\ \text{Yr.4} \\ \text{Yr.5} \end{array} & \begin{bmatrix} 1 & 8 & 1 \\ -4 & 9 & 3 \\ 5 & 10 & 4 \\ 7 & 3 & 5 \\ 8 & 7 & 6 \end{bmatrix} \end{array}$$

我们想计算不同回报向量之间的相关性，每个公司一个，在矩阵 A 中“打包”。

假设持有量相等，投资组合的方差协方差矩阵为：

$$\sigma(A) = \frac{G^T G}{n-1}$$

其中 G 是以均值为中心的观测值，$n-1$ 对应于观测次数减 1。

均值中心 (或去均值) 矩阵 G 为：

$$
\begin{array}{c}
\begin{array}{cccc} & \color{red}{\text{y}} & \color{blue}{\text{g}} & \color{green}{\text{m}} \end{array} \\
\begin{array}{c} \text{Yr.1} \\ \text{Yr.2} \\ \text{Yr.3} \\ \text{Yr.4} \\ \text{Yr.5} \end{array}
\begin{bmatrix}
-2.4 & 0.6 & -2.8 \\
-7.4 & 1.6 & -0.8 \\
1.6 & 2.6 & 0.2 \\
3.6 & -4.4 & 1.2 \\
4.6 & -0.4 & 2.2
\end{bmatrix}
\end{array}
$$

并且方差-协方差矩阵为：

$$
\begin{array}{c}
\begin{array}{cccc} & \color{red}{y} & \color{blue}{\text{g}} & \color{green}{\text{m}} \end{array} \\
\begin{array}{c} \color{red}{y} \\ \color{blue}{\text{g}} \\ \color{green}{\text{m}} \end{array}
\begin{bmatrix}
24.30 & -6.70 & 6.85 \\
-6.70 & 7.30 & -2.15 \\
6.85 & -2.15 & 3.70
\end{bmatrix}
\end{array}
$$

所以它把 5×3 的 A 矩阵变成了 3×3 矩阵。

计算相关矩阵所涉及的操作类似，但在数据点居中后，通过减去协方差矩阵中的列平均值，将每个数据点除以每个公司 (列向量) 收益的标准差 (standard deviation)，从而对数据点进行标准化：

$$
cor(A) = \frac{1}{n-1}
\underset{3 \times 5\,\text{matrix}}{\begin{bmatrix}
\frac{y_1-\bar{y}}{sd(y)} & \frac{y_2-\bar{y}}{sd(y)} & \frac{y_3-\bar{y}}{sd(y)} & \frac{y_4-\bar{y}}{sd(y)} & \frac{y_5-\bar{y}}{sd(y)} \\
\frac{g_1-\bar{g}}{sd(g)} & \frac{g_2-\bar{g}}{sd(g)} & \frac{g_3-\bar{g}}{sd(g)} & \frac{g_4-\bar{g}}{sd(g)} & \frac{g_5-\bar{g}}{sd(g)} \\
\frac{m_1-\bar{m}}{sd(m)} & \frac{m_2-\bar{m}}{sd(m)} & \frac{m_3-\bar{m}}{sd(m)} & \frac{m_4-\bar{m}}{sd(m)} & \frac{m_5-\bar{m}}{sd(m)}
\end{bmatrix}}
\underset{5 \times 3\,\text{matrix}}{\begin{bmatrix}
\frac{y_1-\bar{y}}{sd(y)} & \frac{g_1-\bar{g}}{sd(g)} & \frac{m_1-\bar{m}}{sd(m)} \\
\frac{y_2-\bar{y}}{sd(y)} & \frac{g_2-\bar{g}}{sd(g)} & \frac{m_2-\bar{m}}{sd(m)} \\
\frac{y_3-\bar{y}}{sd(y)} & \frac{g_3-\bar{g}}{sd(g)} & \frac{m_3-\bar{m}}{sd(m)} \\
\frac{y_4-\bar{y}}{sd(y)} & \frac{g_4-\bar{g}o}{sd(g)} & \frac{m_4-\bar{m}}{sd(m)} \\
\frac{y_5-\bar{y}}{sd(y)} & \frac{g_5-\bar{g}}{sd(g)} & \frac{m_5-\bar{m}}{sd(m)}
\end{bmatrix}}
$$

为了完整性起见，还有一件事要做：到目前为止，我们有一个笨拙的矩阵作为结果，但通常我们要估计投资组合方差：1 个投资组合；1 个方差。为此，我们将 A 的方差-协方差矩阵左乘和右乘每个股票中包含比例或权重的向量-W。因为我们希望得到一个标量单个数字，所以代数将是：$W^T \sigma(A) W$，权重向量 (分数) 也就不足为奇了，在这种情况下，为完全匹配 3×1 向量，左侧为 W^T，右侧为 W。