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 $\mathbb{E}\big((\mathbf{X} - \mu)(\mathbf{X} - \mu)^T\big)$。

这与维基百科的另一个定义是一致的：

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

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 $\frac{1}{n-1}\mathbf{X}^\top\mathbf{X}$。

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：

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

同时尝试从这个分解构造协方差矩阵得到

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

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

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

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，trace($\hat{\mathbf{\Sigma}}$) 是到质心的平均平方距离：以 $\dot{\mathbf{X}}$ 为中心变量的矩阵，$\hat{\mathbf{\Sigma}} = \frac{1}{n}\dot{\mathbf{X}}'\dot{\mathbf{X}}$，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，trace($\hat{\mathbf{\Sigma}}$) 是单变量方差的自然推广。

第二个推广是 det($\hat{\mathbf{\Sigma}}$)：这是描述分布的椭球体体积的度量。更准确地说，$|\det(\hat{\mathbf{\Sigma}})|$ 是应用线性变换 $\hat{\mathbf{\Sigma}}$ 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)：