STAT243 Lecture 10.2 Statistical Interpretations of Matrix Properties

1 Linear Independence, Rank, and Basis

1.1 线性无关与秩

一组向量 $v_1, \dots, v_n$ 是**线性无关**的，当：

$$\sum_i c_i v_i = 0 \Rightarrow c_i = 0 \ \forall i$$

矩阵的 **秩 (rank)** = 线性无关的列（或行）的数量。
最大值为 $\min(\text{行数}, \text{列数})$。

1.2 基向量与张成空间

任意一组线性无关向量 v_i 构成的线性组合：

$$y = \sum_i c_i v_i$$

定义了该组向量张成的空间 (span)。
v_i 即为该空间的**基 (basis)**。

若 v_i 是**正交归一化的**，则：

$$c_i = \langle y, v_i \rangle = v_i^\top y$$

1.3 回归模型中的秩与空间

设计矩阵 X 的列为协变量向量：

- 若 $q = \text{rank}(X) < p$
  则有 $p - q$ 列是其他列的线性组合
- 空间维数为 q，对应 q 个基向量
- 在回归中：
  - 若 $n = p = q \rightarrow$ 唯一且精确解（无残差）
  - 若 $n < p \rightarrow$ 欠定系统（多解）
  - 若 $n > p \rightarrow$ 过定系统（最小二乘求近似解）

回归的几何解释：
Y 被投影到由 X 的列张成的空间上。

1.4 Invertibility, Singularity, and Positive Definiteness

1.4.1 特征分解 (Eigendecomposition)

对于可对角化矩阵：

$$A = \Gamma\Lambda\Gamma^{-1}$$

若 A 对称：

$$A = \Gamma\Lambda\Gamma^\top$$

其中：

- Γ：特征向量矩阵
- Λ：对角特征值矩阵
- 非零特征值数量 = 矩阵秩

1.4.2 矩阵分类与性质

性质	定义	推论
非奇异 (nonsingular)	A^{-1} 存在	秩满 (full rank)
正定 (positive definite)	$x^\top Ax > 0$	所有 $\lambda_i > 0$
半正定 (positive semi-definite)	$x^\top Ax \ge 0$	一部分 $\lambda_i = 0$
奇异 (singular)	不可逆	存在 $\lambda_i = 0$

若 A 为协方差矩阵 $\text{Cov}(y)$：

$$x^\top Ax = \text{Var}(x^\top y) \ge 0$$

因此协方差矩阵必为半正定矩阵。

1.4.3 结论总结

代数性质	统计解释
A p.d. ⇔ $x^\top Ax > 0$ ⇔ $\lambda_i > 0$	协方差矩阵，方差均为正
A p.s.d. ⇔ $x^\top Ax \ge 0$ 且部分 $\lambda_i = 0$	受约束协方差矩阵，部分方向方差为 0
$\lambda_i = 0$	某线性组合无方差信息（不识别）

---

2 Eigen-decomposition Interpretation

生成随机向量 y 的思路：

$$y = \Gamma\Lambda^{1/2}z, \quad \text{where } \text{Cov}(z) = I$$

解释：

- $\Gamma_{\cdot i}$ 为方差主方向（特征向量）
- $\sqrt{\lambda_i}$ 为对应标准差
- z_i 为该方向上的标准化系数

2.1.1 问题与解释

1. **若 $\lambda_i = 0$：**
   该方向上方差为零，y 在该方向无变化。
2. **若将极小 λ_i 设为 0：**
   去除对应的低方差分量，相当于“平滑”或“降维”。
3. **若 $(\Lambda^{-1})_{ii}$ 很大：**
   对应方向方差很小（高精度）。
   若很小 → 对应方向方差极大（低精度）。

---

3 Generalized (Pseudo) Inverses (optional)

若 A 不可逆，仍可定义广义逆 A^- 使：

$$AA^-A = A$$

Moore–Penrose 伪逆 (pseudo-inverse)：唯一满足附加约束的广义逆，记为 A^+。
其最小化解为：

$$x = A^+b$$

即满足 $Ax = b$ 且 $|x|$ 最小。

3.1 由特征分解求伪逆

$$A^+ = \Gamma\Lambda^+\Gamma^\top, \quad \Lambda^+_{ii} = \begin{cases} 1/\lambda_i, & \lambda_i > 0 \\ 0, & \lambda_i = 0 \end{cases}$$

统计解释：
若 $\lambda_i = 0$，对应方向无信息（无限方差），伪逆将该方向方差设为 0。

3.2 示例：AR(1) 与 AR(2) 精度矩阵

- 一阶自回归 (AR(1)) 精度矩阵：

  $$(1 \quad -1 \quad 0 \quad 0 \quad 0 - 1 \quad 2 \quad -1 \quad 0 \quad 0\,0 \quad -1 \quad 2 \quad -1 \quad 0\,0 \quad 0 \quad -1 \quad 2 \quad -1\,0 \quad 0 \quad 0 \quad -1 \quad 1)$$

  零特征值对应常数向量（均值不可识别）。
- 二阶自回归 (AR(2)) 矩阵中有两个零特征值，对应：
  - 总和不确定性 (constant component)
  - 线性趋势不确定性 (linear component)

伪逆操作即为：
将 $\lambda_i = 0$ 的分量方差设为零，相当于约束 $\sum y_i = 0$ 或去除线性趋势。

---

4 Regression 中的 Matrices

在线性回归中：

$$\hat{\beta} = (X^\top X)^{-1}X^\top Y$$

- $X^\top X$：
  - 对称
  - 半正定 (n.n.d.)
  - 若列向量线性无关 → 正定 (p.d.)

4.1.1 Hat Matrix (投影矩阵)

$$H = X(X^\top X)^{-1}X^\top$$

性质：

1. 投影：$\hat{Y} = HY$
2. 幂等性：$HH = H$
3. 奇异性：H 不可逆（因为其秩 < n）
4. 唯一情况 H 可逆：$H = I$（仅当 X 为满秩方阵且 $n = p$）

几何解释：
H 将 Y 投影到由 X 列张成的空间；
投影后再投影不改变结果，因此 $HH = H$。