

# 第十一讲

王伟文 暨南大学

## 1 线性代数回顾

### 定理 1.1: 对称实矩阵谱分解

任意  $n$  阶实矩阵  $\mathbf{A}$  可以表示为

$$\mathbf{A} = \sum_{i=1}^n \lambda_i \mathbf{u}_i \mathbf{u}_i^T,$$

其中  $(\lambda_i, \mathbf{u}_i)$  为矩阵  $\mathbf{A}$  的特征值及相应的正交化单位特征向量,  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ .

令  $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r] \in \mathbb{R}^{n \times r}$ ,  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_r)$ , 则

$$\mathbf{A} = \mathbf{U} \Lambda \mathbf{U}^T.$$

从优化的视角看特征值

$$\lambda_1 = \langle \mathbf{A} \mathbf{u}_1, \mathbf{u}_1 \rangle = \max_{\|\mathbf{x}\|=1} \mathbf{x}^T \mathbf{A} \mathbf{x}, \quad \mathbf{u}_1 = \arg \max_{\|\mathbf{x}\|=1} \mathbf{x}^T \mathbf{A} \mathbf{x}$$

$$\lambda_2 = \langle \mathbf{A} \mathbf{u}_2, \mathbf{u}_2 \rangle = \max_{\substack{\mathbf{x} \perp \mathbf{u}_1 \\ \|\mathbf{x}\|=1}} \mathbf{x}^T \mathbf{A} \mathbf{x}, \quad \mathbf{u}_2 = \arg \max_{\substack{\mathbf{x} \perp \mathbf{u}_1 \\ \|\mathbf{x}\|=1}} \mathbf{x}^T \mathbf{A} \mathbf{x}$$

$$\lambda_3 = \langle \mathbf{A} \mathbf{u}_3, \mathbf{u}_3 \rangle = \max_{\substack{\mathbf{x} \perp \{\mathbf{u}_1, \mathbf{u}_2\} \\ \|\mathbf{x}\|=1}} \mathbf{x}^T \mathbf{A} \mathbf{x}, \quad \mathbf{u}_3 = \arg \max_{\substack{\mathbf{x} \perp \{\mathbf{u}_1, \mathbf{u}_2\} \\ \|\mathbf{x}\|=1}} \mathbf{x}^T \mathbf{A} \mathbf{x}$$

...

$$\lambda_n = \langle \mathbf{A} \mathbf{u}_n, \mathbf{u}_n \rangle = \min_{\|\mathbf{x}\|=1} \mathbf{x}^T \mathbf{A} \mathbf{x}, \quad \mathbf{u}_n = \arg \min_{\|\mathbf{x}\|=1} \mathbf{x}^T \mathbf{A} \mathbf{x}$$

### 命题 1.1

对任意  $n$ -阶对称矩阵  $\mathbf{A}$ , 有

$$\lambda_1 = \langle \mathbf{A} \mathbf{u}_1, \mathbf{u}_1 \rangle = \max_{\|\mathbf{x}\|=1} \mathbf{x}^T \mathbf{A} \mathbf{x}, \quad \mathbf{u}_1 = \arg \max_{\|\mathbf{x}\|=1} \mathbf{x}^T \mathbf{A} \mathbf{x}$$

$$\lambda_n = \langle \mathbf{A} \mathbf{u}_n, \mathbf{u}_n \rangle = \min_{\|\mathbf{x}\|=1} \mathbf{x}^T \mathbf{A} \mathbf{x}, \quad \mathbf{u}_n = \arg \min_{\|\mathbf{x}\|=1} \mathbf{x}^T \mathbf{A} \mathbf{x}$$

证明. 由定理 1.1 知, 矩阵  $\mathbf{A}$  可表示谱分解的形式  $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$ , 故

$$\begin{aligned}\max_{\|\mathbf{x}\|=1} \mathbf{x}^T \mathbf{A} \mathbf{x} &= \max_{\|\mathbf{x}\|=1} \mathbf{x}^T \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T \mathbf{x} \\ &= \max_{\|\mathbf{y}\|^2=1} \mathbf{y}^T \mathbf{\Lambda} \mathbf{y} = \max_{\|\mathbf{y}\|^2=1} \sum_{i=1}^n y_i^2 \lambda_i \quad (\mathbf{y} = \mathbf{U}^T \mathbf{x}) \\ &= \max_{\xi \in \Delta_n} \sum_{i=1}^n \xi_i \lambda_i = \lambda_1 \quad (y_i^2 = \xi_i)\end{aligned}$$

余下结论显然成立. □

### 定理 1.2: 奇异值分解 (Singular Value Decomposition)

$\forall \mathbf{A} \in \mathbb{R}^{m \times n}$ , 可以表示为

$$\mathbf{A} = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$$

其中  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$  为奇异值,  $\mathbf{u}_i$  为左奇异向量,  $\mathbf{v}_i$  为右奇异向量, 满足

$$\sigma_i = \sqrt{\lambda_i(\mathbf{A}\mathbf{A}^T)} = \sqrt{\lambda_i(\mathbf{A}^T\mathbf{A})};$$

$\{\mathbf{u}_i\}$  为矩阵  $\mathbf{A}\mathbf{A}^T$  的正交化特征向量;

$\{\mathbf{v}_i\}$  为矩阵  $\mathbf{A}^T\mathbf{A}$  的正交化特征向量.

且  $r = \text{rank}(\mathbf{A})$ .

奇异值分解的矩阵形式

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T,$$

其中  $\mathbf{U} \in \mathbb{R}^{m \times n}$ ,  $\mathbf{V} \in \mathbb{R}^{n \times n}$  为正交矩阵, 分别为矩阵  $\mathbf{A}\mathbf{A}^T$  和  $\mathbf{A}^T\mathbf{A}$  的正交化特征向量构成的列矩阵,  $\mathbf{\Sigma} \in \mathbb{R}^{m \times n}$  的对角元为奇异值  $(\sigma_1, \sigma_2, \dots, \sigma_r)$ , 其他位置为 0.

紧凑形式

$$\mathbf{A} = \mathbf{U}_r \mathbf{\Sigma}_r \mathbf{V}_r^T,$$

其中  $\mathbf{U}_r \in \mathbb{R}^{m \times r}$  和  $\mathbf{V}_r \in \mathbb{R}^{n \times r}$  分别为  $\mathbf{U}$  和  $\mathbf{V}$  的前  $r$  列, 满足  $\mathbf{U}_r^T \mathbf{U}_r = \mathbf{V}_r^T \mathbf{V}_r = \mathbf{I}_r$ ,  $\mathbf{\Sigma}_r = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$ .

从优化的视角看奇异值

$$\begin{aligned}\sigma_1 &= \max_{\|x\|=1} \|Ax\|_2, & v_1 &= \arg \max_{\|x\|=1} \|Ax\|_2 \\ \sigma_2 &= \max_{\substack{x \perp v_1 \\ \|x\|=1}} \|Ax\|_2, & v_2 &= \arg \max_{\substack{x \perp v_1 \\ \|x\|=1}} \|Ax\|_2 \\ \sigma_3 &= \max_{\substack{x \perp \{v_1, v_2\} \\ \|x\|=1}} \|Ax\|_2, & v_3 &= \arg \max_{\substack{x \perp \{v_1, v_2\} \\ \|x\|=1}} \|Ax\|_2 \\ &\dots \\ \sigma_n &= \min_{\|x\|=1} \|Ax\|_2, & v_n &= \arg \min_{\|x\|=1} \|Ax\|_2\end{aligned}$$

### 命题 1.2

设  $A$  为  $n$ -阶对称实矩阵, 则有

$$\sigma_1(A) = \max_{\|x\|=1} |x^T Ax| = \max_{i \in [n]} |\lambda_i(A)|.$$

## 2 矩阵范数

考虑矩阵  $A \in \mathbb{R}^{m \times n}$ , Frobenius 范数

$$\|A\|_F = \sqrt{\sum_{i,j} A_{i,j}^2}$$

矩阵内积

$$\langle A, B \rangle = \text{trace}(B^T A)$$

$$\|A\|_F^2 = \langle A, A \rangle$$

### 命题 2.1: $\|\cdot\|_F$ 的正交不变性

对任意正交矩阵  $U \in \mathbb{R}^{m \times m}$  或  $V \in \mathbb{R}^{n \times n}$ , 有

$$\|UA\|_F = \|AV\|_F = \|A\|_F$$

证明.  $\|UA\|_F = \sqrt{\langle UA, UA \rangle} = \sqrt{\text{trace}(A^T U^T U A)} = \sqrt{\text{trace}(A^T A)} = \|A\|_F$ .  $\square$

### 命题 2.2

$$\|A\|_F^2 = \sum \sigma_i^2(A).$$

证明. 将矩阵  $\mathbf{A}$  表示为 SVD 分解的形式,  $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ .

$$\|\mathbf{A}\|_F^2 = \|\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\| \stackrel{(i)}{=} \|\mathbf{\Sigma}\|_F^2 = \text{trace}(\mathbf{\Sigma}^T\mathbf{\Sigma}) = \sum \sigma_i^2(\mathbf{A}),$$

其中 (i) 由 Frobenius 范数的正交不变性得到. □

### 定义 2.1: 矩阵谱范数

$$\|\mathbf{A}\| = \max_{\|\mathbf{x}\|_2=1} \|\mathbf{A}\mathbf{x}\|_2 = \max_{\|\mathbf{x}\|_2 \leq 1} \|\mathbf{A}\mathbf{x}\|_2 = \max_{\mathbf{x}} \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2}$$

关于矩阵谱范数有

- (a) 对任意正交矩阵  $\mathbf{U} \in \mathbb{R}^{m \times m}$  或  $\mathbf{V} \in \mathbb{R}^{n \times n}$ ,  $\|\mathbf{U}\mathbf{A}\| = \|\mathbf{A}\mathbf{V}\| = \|\mathbf{A}\|$ ;
- (b) 若矩阵  $\mathbf{A}$  可逆,  $\|\mathbf{A}^{-1}\| = \frac{1}{\sigma_n(\mathbf{A})}$
- (c)  $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$
- (d)  $\|\mathbf{A}\mathbf{B}\| \leq \|\mathbf{A}\| \|\mathbf{B}\|$  ( $\|\mathbf{A}\mathbf{B}\mathbf{x}\|_2 \leq \|\mathbf{A}\| \|\mathbf{B}\mathbf{x}\|_2 \leq \|\mathbf{A}\| \|\mathbf{B}\|$ )
- (e)  $\|\mathbf{A}\| \leq \|\mathbf{A}\|_F \leq \sqrt{r} \|\mathbf{A}\|$  ( $\|\mathbf{A}\| = \sigma_1(\mathbf{A}) \leq \sqrt{\sum_{i=1}^r \sigma_i^2(\mathbf{A})} = \|\mathbf{A}\|_F \leq \sqrt{r \sigma_1^2(\mathbf{A})} = \sqrt{r} \|\mathbf{A}\|$ )  
这里  $r = \text{rank}(\mathbf{A})$ .
- (f)  $\|\mathbf{A}\| = \max_{\|\mathbf{x}\|=1 \text{ 且 } \|\mathbf{y}\|=1} \mathbf{y}^T \mathbf{A} \mathbf{x}$   
 $(\|\mathbf{z}\|_2 = \max_{\|\mathbf{y}\|_2=1} \mathbf{y}^T \mathbf{z}, \max_{\|\mathbf{x}\|=1 \text{ 且 } \|\mathbf{y}\|=1} \mathbf{y}^T \mathbf{A} \mathbf{x} = \max_{\|\mathbf{x}\|=1} \max_{\|\mathbf{y}\|_2=1} \mathbf{y}^T \mathbf{A} \mathbf{x} = \max_{\|\mathbf{x}\|_2=1} \|\mathbf{A}\mathbf{x}\|_2 = \sigma_1(\mathbf{A}) = \|\mathbf{A}\|)$

## 3 随机向量

设  $X$  为一个  $d$ -维随机向量, 其均值为  $\mathbb{E}X = \boldsymbol{\mu}$ .

协方差矩阵为

$$\text{cov}(X) = \mathbb{E}(X - \boldsymbol{\mu})(X - \boldsymbol{\mu})^T = \mathbb{E}[XX^T] - \boldsymbol{\mu}\boldsymbol{\mu}^T$$

- $(\text{cov}(X))_{i,j} = \mathbb{E}(X_i - \mu_i)(X_j - \mu_j) = \text{cov}(X_i, X_j)$
- $(\text{cov}(X))_{i,i} = \text{var}(X_i)$

### 命题 3.1

$\text{cov}(X)$  是一个正半定矩阵.

## 4 主成分分析 (Principal Component Analysis, PCA)

设  $X \in \mathbb{R}^d$  是以 0 为均值的随机向量, 协方法为  $\text{cov}(X) = \Sigma$ . 要找一个单位向量将  $X$  投影到 1 维空间中, 并使得方差最大, 即

$$\max_{\|v\|_2=1} \mathbb{E}[(\langle X, v \rangle)^2] = \max_{\|v\|_2=1} v^T \Sigma v$$

由矩阵特征值性质知, 最优的投影变换为  $\Sigma$  的最大特征对应的单位化特征向量  $v_1(\Sigma)$ , 此时

$$\lambda_1(\Sigma) = \max_{\|v\|_2=1} v^T \Sigma v.$$

接下来寻找第二个“最佳投影方向”, 满足

$$\max_{\substack{v \perp v_1 \\ \|v\|_2=1}} v^T \Sigma v.$$

得到最优的投影变换为  $\Sigma$  的最二大特征对应的单位化特征向量  $v_2(\Sigma)$ , 此时

$$\lambda_2(\Sigma) = \max_{\substack{v \perp v_1 \\ \|v\|_2=1}} v^T \Sigma v.$$

依次类推, 协方差  $\Sigma$  的前  $k$  个特征值对应的单位正交化特征向量构成主成分分析的“主成分”, 即随机变量  $X$  投影到  $k$ -为空间的线性变换.

称比率

$$\frac{\sum_{i=1}^k \lambda_i(\Sigma)}{\sum_{j=1}^d \lambda_j(\Sigma)}$$

为这  $k$  个主成分的累积解释度.

然而在实践中, 随机变量  $X \in \mathbb{R}^d$  的协方差往往是未知的. 给定  $n$  个来自与  $X$  同一总体的样本  $X_1, X_2, \dots, X_n$ , 样本协方差为

$$\Sigma_n = \frac{1}{n} \sum_{i=1}^n X_i X_i^T.$$

此时, 主成分分析建立在  $\Sigma_n$  上.

一个自然的问题是,  $PCA(\Sigma_n)$  在多大程度上近似  $PCA(\Sigma)$ , 即能否确保

$$\lambda_i(\Sigma_n) \approx \lambda_i(\Sigma) \quad \text{且} \quad v_i(\Sigma_n) \approx v_i(\Sigma),$$

相应的样本量应该是多少?

我们首先来看看能否在谱范数的度量下控制  $\Sigma$  与  $\Sigma_n$  的差别. 由谱范数性质

$$\begin{aligned}\|\Sigma_n - \Sigma\| &= \max_{\mathbf{v} \in \mathcal{S}^{d-1}} |\mathbf{v}^T (\Sigma_n - \Sigma) \mathbf{v}| \\ &= \max_{\mathbf{v} \in \mathcal{S}^{d-1}} |\mathbf{v}^T \Sigma_n \mathbf{v} - \mathbf{v}^T \Sigma \mathbf{v}| \\ &= \max_{\mathbf{v} \in \mathcal{S}^{d-1}} \left| \frac{1}{n} \sum_{i=1}^n \langle X_i, \mathbf{v} \rangle^2 - \mathbb{E}[\langle X, \mathbf{v} \rangle^2] \right|\end{aligned}$$

其中  $\mathcal{S}^{d-1} \triangleq \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_2 = 1\}$

定义随机变量

$$Z(\mathbf{v}) = \frac{1}{n} \sum_{i=1}^n \langle X_i, \mathbf{v} \rangle^2 - \mathbb{E}[\langle X, \mathbf{v} \rangle^2],$$

此时

$$\|\Sigma_n - \Sigma\| = \max_{\mathbf{v} \in \mathcal{S}^{d-1}} |Z(\mathbf{v})|$$

#### 命题 4.1: 谱范数与球面上的 $\epsilon$ -覆盖

对于矩阵  $\mathbf{A} \in \mathbb{R}^{m \times n}$ , 设  $\mathcal{N}$  为单位球面  $\mathcal{S}^{n-1}$  的一个  $\epsilon$ -覆盖, 则

$$\|\mathbf{A}\| \leq \frac{1}{1 - \epsilon} \max_{\mathbf{x} \in \mathcal{N}} \|\mathbf{A}\mathbf{x}\| \quad \forall \epsilon > 0.$$

**证明.** 由  $\epsilon$ -覆盖定义,  $\forall \mathbf{v} \in \mathcal{S}^{n-1}$ ,  $\exists \mathbf{x} \in \mathcal{N}$ , 满足

$$\|\mathbf{v} - \mathbf{x}\|_2 \leq \epsilon$$

由三角不等式及谱范数定义

$$\|\mathbf{A}\mathbf{v}\|_2 - \|\mathbf{A}\mathbf{x}\|_2 \leq \|\mathbf{A}\mathbf{v} - \mathbf{A}\mathbf{x}\|_2 \leq \|\mathbf{A}\| \|\mathbf{v} - \mathbf{x}\|_2 \leq \|\mathbf{A}\| \epsilon$$

即对任意  $\mathbf{v} \in \mathcal{S}^{d-1}$ , 存在  $\mathbf{x} \in \mathcal{N}$ , 使得

$$\|\mathbf{A}\mathbf{v}\|_2 \leq \|\mathbf{A}\| \epsilon + \|\mathbf{A}\mathbf{x}\|_2$$

因此

$$\max_{\mathbf{v} \in \mathcal{S}^{n-1}} \|\mathbf{A}\mathbf{v}\|_2 \leq \|\mathbf{A}\| \epsilon + \max_{\mathbf{x} \in \mathcal{N}} \|\mathbf{A}\mathbf{x}\|_2$$

即

$$\|\mathbf{A}\| \leq \epsilon \|\mathbf{A}\| + \max_{\mathbf{x} \in \mathcal{N}} \|\mathbf{A}\mathbf{x}\|_2$$

移项得到

$$\|\mathbf{A}\| \leq \frac{1}{1 - \epsilon} \max_{\mathbf{x} \in \mathcal{N}} \|\mathbf{A}\mathbf{x}\|_2$$

□

**命题 4.2: 对称矩阵谱范数与球面上的  $\epsilon$  覆盖**

设  $\mathbf{A}$  为  $n$ -阶对称实矩阵,  $\mathcal{N}$  为单位球面  $\mathcal{S}^{n-1}$  上的一个  $\epsilon$ -覆盖, 则

$$\|\mathbf{A}\| \leq \frac{1}{1-2\epsilon} \max_{\mathbf{x} \in \mathcal{N}} |\mathbf{x}^T \mathbf{A} \mathbf{x}| \quad \forall \epsilon \in (0, \frac{1}{2}).$$

**证明.**  $\forall \mathbf{x} \in \mathcal{S}^{n-1}, \exists \mathbf{u} \in \mathcal{N}$ , 满足  $\|\mathbf{x} - \mathbf{u}\|_2 \leq \epsilon$ .

$$\begin{aligned} |\mathbf{x}^T \mathbf{A} \mathbf{x} - \mathbf{u}^T \mathbf{A} \mathbf{u}| &= |\mathbf{x}^T \mathbf{A} \mathbf{x} - \mathbf{x}^T \mathbf{A} \mathbf{u} + \mathbf{x}^T \mathbf{A} \mathbf{u} - \mathbf{u}^T \mathbf{A} \mathbf{u}| \\ &\leq |\langle \mathbf{A} \mathbf{x}, \mathbf{x} - \mathbf{u} \rangle| + |\langle \mathbf{A} \mathbf{u}, \mathbf{u} - \mathbf{x} \rangle| \\ &\leq \|\mathbf{x} - \mathbf{u}\|_2 \|\mathbf{A} \mathbf{x}\|_2 + \|\mathbf{x} - \mathbf{u}\|_2 \|\mathbf{A} \mathbf{u}\|_2 \\ &\leq \epsilon \|\mathbf{A} \mathbf{x}\|_2 + \epsilon \|\mathbf{A} \mathbf{u}\|_2 \end{aligned}$$

故有

$$\begin{aligned} |\mathbf{x}^T \mathbf{A} \mathbf{x}| &\leq \epsilon \|\mathbf{A} \mathbf{x}\|_2 + \epsilon \|\mathbf{A} \mathbf{u}\|_2 + |\mathbf{u}^T \mathbf{A} \mathbf{u}| \\ &\leq \epsilon \|\mathbf{A} \mathbf{x}\|_2 + \epsilon \max_{\mathbf{u} \in \mathcal{N}} \|\mathbf{A} \mathbf{u}\|_2 + \max_{\mathbf{u} \in \mathcal{N}} |\mathbf{u}^T \mathbf{A} \mathbf{u}| \\ &\leq \epsilon \|\mathbf{A} \mathbf{x}\|_2 + \epsilon \max_{\mathbf{u} \in \mathcal{S}^{n-1}} \|\mathbf{A} \mathbf{u}\|_2 + \max_{\mathbf{u} \in \mathcal{N}} |\mathbf{u}^T \mathbf{A} \mathbf{u}| \end{aligned}$$

上述不等式关于  $\mathbf{x}$  最大化

$$\max_{\mathbf{x} \in \mathcal{S}^{n-1}} |\mathbf{x}^T \mathbf{A} \mathbf{x}| \leq \epsilon \max_{\mathbf{x} \in \mathcal{S}^{n-1}} \|\mathbf{A} \mathbf{x}\|_2 + \epsilon \max_{\mathbf{u} \in \mathcal{S}^{n-1}} \|\mathbf{A} \mathbf{u}\|_2 + \max_{\mathbf{u} \in \mathcal{N}} |\mathbf{u}^T \mathbf{A} \mathbf{u}|$$

即

$$\|\mathbf{A}\| \leq 2\epsilon \|\mathbf{A}\| + \max_{\mathbf{u} \in \mathcal{N}} |\mathbf{u}^T \mathbf{A} \mathbf{u}|.$$

□

**命题 4.3: 单位球面的最小覆盖数**

$\forall \epsilon > 0$ , 在  $\|\cdot\|_2$  度量下有

$$\mathcal{N}(\epsilon; \mathcal{S}^{d-1}, \|\cdot\|_2) \leq \left(\frac{2}{\epsilon} + 1\right)^d$$

#### 4.1 协方差一致收敛性分析

##### 定理 4.1: Bernstein 不等式

设  $X_1, X_2, \dots, X_N$  为相互独立、均值为  $\mathbb{E}[X_i] = \mu_i$  的 sub-exponential 随机变量, 相应的参数为  $(\nu_i, \alpha_i)$ , 则有

$$\mathbb{P} \left[ \sum_{i=1}^N (X_i - \mu_i) \geq t \right] \leq \exp \left( -\frac{1}{2} \min \left( \frac{t^2}{\tilde{\nu}}, \frac{t}{\tilde{\alpha}} \right) \right) \quad \forall t > 0,$$

其中  $\tilde{\nu}^2 = \sum_{i=1}^N \nu_i^2$ ,  $\tilde{\alpha} = \max_{i=1, \dots, N} \alpha_i$ .

特别地, 若  $X_i$  独立同分布, Bernstein 不等式可改写为

$$\mathbb{P} \left[ \frac{1}{N} \sum_{i=1}^N X_i - \mathbb{E}[X] \geq \delta \right] \leq \exp \left( -\frac{n}{2} \min \left( \frac{\delta^2}{\nu}, \frac{\delta}{\alpha} \right) \right) \quad \forall t > 0,$$

##### 定理 4.2: 协方差估计

设  $X_i \sim N(\mathbf{0}, \Sigma)$ , 若  $n \geq \frac{64d \log 3}{\delta^2}$  且  $0 < \delta \leq \frac{1}{2}$ , 则

$$\|\Sigma_n - \Sigma\| \leq \delta \|\Sigma\|$$

至少以概率  $1 - 2 \exp(-\frac{n\delta^2}{32})$  成立.

**证明.** 不失一般性, 设  $\|\Sigma\| = 1$ .

(尾界)

取任一固定  $\mathbf{v} \in \mathcal{S}^{d-1}$ , 则  $\langle X, \mathbf{v} \rangle$  是以零为均值的 Gaussian 随机变量, 其方差为

$$\sigma^2 = \mathbb{E}[\langle X, \mathbf{v} \rangle^2] = \mathbb{E}[\mathbf{v}^T X X^T \mathbf{v}] = \mathbf{v}^T \Sigma \mathbf{v} \leq \|\Sigma\| = 1$$

故  $\langle X, \mathbf{v} \rangle / \sqrt{\sigma} \sim N(0, 1)$  且  $\langle X, \mathbf{v} \rangle^2 / \sigma$  是以  $(2, 4)$  为参数的 sub-exponential 随机变量.

因为  $\sigma \leq 1$ , 故

$$\frac{1}{n} \left| \sum_{i=1}^n \frac{\langle X_i, \mathbf{v} \rangle^2}{\sigma} - \frac{\mathbb{E}[\langle X, \mathbf{v} \rangle^2]}{\sigma} \right| \geq \frac{1}{n} \left| \sum_{i=1}^n \langle X_i, \mathbf{v} \rangle^2 - \mathbb{E}[\langle X, \mathbf{v} \rangle^2] \right|$$

因此有

$$\begin{aligned} \mathbb{P} \left[ \frac{1}{n} \left| \sum_{i=1}^n \langle X_i, \mathbf{v} \rangle^2 - \mathbb{E}[\langle X, \mathbf{v} \rangle^2] \right| \geq \delta \right] &\leq \mathbb{P} \left[ \frac{1}{n} \left| \sum_{i=1}^n \frac{\langle X_i, \mathbf{v} \rangle^2}{\sigma} - \frac{\mathbb{E}[\langle X, \mathbf{v} \rangle^2]}{\sigma} \right| \geq \delta \right] \\ &\leq 2 \exp \left( -\frac{n}{2} \min \left( \frac{\delta^2}{2}, \frac{\delta}{4} \right) \right) \quad \forall \delta > 0. \end{aligned}$$

(离散化)

由命题 4.2,

$$\|\Sigma_n - \Sigma\| \geq \delta \Rightarrow \frac{1}{1 - 2\epsilon} \max_{\mathbf{v} \in \mathcal{N}} |Z(\mathbf{v})| \geq \delta$$



取  $\epsilon = \frac{1}{4}$ ,

$$\begin{aligned}\mathbb{P}[\|\Sigma_n - \Sigma\| \geq \delta] &\leq \mathbb{P}\left[2 \max_{\mathbf{v} \in \mathcal{N}} |Z(\mathbf{v})| \geq \delta\right] \\ &= \mathbb{P}\left[\max_{\mathbf{v} \in \mathcal{N}} |Z(\mathbf{v})| \geq \frac{\delta}{2}\right]\end{aligned}$$

(联合界)

$$\begin{aligned}\mathbb{P}\left[\max_{\mathbf{v} \in \mathcal{N}} |Z(\mathbf{v})| \geq \frac{\delta}{2}\right] &= \mathbb{P}\left[\bigcup_{\mathbf{v} \in \mathcal{N}} \left\{|Z(\mathbf{v})| \geq \frac{\delta}{2}\right\}\right] \\ &\leq \sum_{\mathbf{v} \in \mathcal{N}} \mathbb{P}\left[Z(\mathbf{v}) \geq \frac{\delta}{2}\right] \\ &\leq |\mathcal{N}| \mathbb{P}\left[Z(\mathbf{v}) \geq \frac{\delta}{2}\right] \\ &\leq 9^d \cdot 2 \exp\left(-\frac{n}{2} \min\left(\frac{\delta^2}{8}, \frac{\delta}{8}\right)\right) \quad \forall \delta > 0.\end{aligned}$$

设  $\delta < 1$ , 若  $n \geq \frac{64d \log 3}{\delta^2}$

$$\mathbb{P}\left[\|\Sigma_n - \Sigma\| \geq \mathbb{P}\left[\max_{\mathbf{v} \in \mathcal{N}} |Z(\mathbf{v})| \geq \frac{\delta}{2}\right] \geq \delta\right] \leq 2 \exp\left(-\frac{n\delta^2}{32}\right).$$

此时

$$\|\Sigma_n - \Sigma\| \leq \delta$$

至少以概率  $1 - 2 \exp(-\frac{n\delta^2}{32})$  成立. □

## 4.2 扰动理论

### 定理 4.3: Weyl 不等式

任意  $n$ -阶级对称矩阵  $\mathbf{S}$  和  $\mathbf{T}$ , 有

$$\max_i |\lambda_i(\mathbf{S}) - \lambda_i(\mathbf{T})| \leq \|\mathbf{S} - \mathbf{T}\|$$

### 定理 4.4: Davis-Kahan 不等式

设有任意  $n$ -阶级对称矩阵  $\mathbf{S}$  和  $\mathbf{T}$ ,  $\lambda_i(\mathbf{S})$  为  $\mathbf{S}$  的特征值从大到小排列次序为  $i$  的特征值, 记  $\lambda_i(\mathbf{S})$  与余下特征值的间隙为

$$\Delta_i \triangleq \min_{j: j \neq i} |\lambda_i(\mathbf{S}) - \lambda_j(\mathbf{S})| > 0,$$

则存在  $\theta \in \{-1, +1\}$ , 使得  $\mathbf{S}$  和  $\mathbf{T}$  中次序为  $i$  的特征值对应的单位化特征向量  $\mathbf{v}_i(\mathbf{S})$  和  $\mathbf{v}_i(\mathbf{T})$  满足

$$\|\mathbf{v}_i(\mathbf{S}) - \theta \mathbf{v}_i(\mathbf{T})\|_2 \leq \frac{2\sqrt{2}\|\mathbf{S} - \mathbf{T}\|}{\Delta_i}$$

总结: PCA 有较好的拟合效果要求

- $n = \mathcal{O}(d)$ ;
- $\Sigma_n$  有较好分离度的特征值间隙.