地址：中国 安徽 合肥市金寨路96号 邮编：230026
电话：0551-63602184 传真：0551-63631760 Http://www.ustc.edu.cn

Problem 1. $A \in R^{m\times n}$ ~~rank A = r~~. $\operatorname{rank} A = r$.

$A = U\Sigma V^T = \sum_{i=1}^{r} u_i \sigma_i v_i^T$

$\sigma_i u_i = Av_i$

$= \sum u_i \sigma_i v_i^T v_i = u_i \sigma_i$

由 $\{v_j\}$ 为正交向量集合. 且 $V^T V = I_n$.

即有 $AV = U\Sigma V^T V = U\Sigma$.

比较等式两端的第 i 列，即有 $Av_i = u_i \sigma_i = \sigma_i u_i$

Problem 2. 证明 $A_k = U_k \Sigma_k V_k^T \ (k \le r)$ $\|A - A_k\|_F \le \|A - B\|_F$.
$\operatorname{rank} B \le k$.

对于一个矩阵 $H \in R^{m\times n}$ $\|H\|_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} [H]_{ji}^2$

$Tr(HH^T) = \sum_{j=1}^{m} [H]_{m,:} [H^T]_{:,m} = \sum_{j=1}^{m} [H]_{m,:} [H]_{m,:}^T$

$= \sum_{j=1}^{m} \| [H]_{m,:} \|_2^2$

$= \sum_{j=1}^{m}\sum_{i=1}^{n} [H]_{j,i}^2 = \|H\|_F^2$.

设 H 的奇异值分解为 ~~$H = U^H \Lambda V^{HT}$~~ $H = M\Lambda N^T$

$\|H\|_F^2 = Tr\{HH^T\} = Tr\{M\Lambda N^T N \Lambda^T M^T\} = Tr\{M\Lambda\Lambda^T M^T\}$

$= Tr\{\Lambda\Lambda^T M M^T\}$

$= Tr\{\Lambda\Lambda^T\} = \sum_{i=1}^{\min(m,n)} \lambda_i$

其中 λ_i 为 H 的特征值.

~~其中 λ_i 为 H 的奇异值分解中 Λ 的对角线元素~~.

不妨设 X 为 A 的最优 k 阶近似矩阵.，作 X 的奇异值分解:

$X = Q\Omega P^T$，令 $B = Q^T A P$，则 A 可表示为 QBP^T.

由此得 $\|A - X\|_F = \|QBP^T - Q\Omega P^T\|_F = \|B - \Omega\|_F$.

由于 X 为 k 阶近似，Ω 可以分块为 $\Omega = \begin{bmatrix} \Omega_k & 0 \\ 0 & 0 \end{bmatrix}$ 其中 $\Omega = diag(\theta_1, \cdots, \theta_k)$

同样将 B 分块为 $\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$ $\|A - X\|_F^2 = \|\Omega_k - B_{11}\|_F^2 + \|B_{12}\|_F^2 + \|B_{21}\|_F^2 + \|B_{22}\|_F^2$.

下面证明： ① $\|B_{12}\|_F^2=0$，反证如下：若 $\|B_{12}\|_F^2\neq 0$.

则令 $Y=Q\begin{bmatrix}B_{11}&B_{12}\\0&0\end{bmatrix}P^T$ $\quad$ rank $Y\le k$.

此时 $\|A-Y\|_F^2=$ ~~$\|\Omega_k-B_{11}\|_F^2+\|B_{21}\|_F^2$~~ $\|B_{21}\|_F^2+\|B_{22}\|_F^2$.

$<\|\Omega_k-B_{11}\|_F^2+\|B_{12}\|_F^2+\|B_{21}\|_F^2+\|B_{22}\|_F^2$.

$=\|A-X\|_F^2$.

此时Y比最优近似X更好，矛盾.

故 $\|B_{12}\|_F^2=0$.

② $\|B_{21}\|_F^2=0$.，用与①类似的 $Z=Q\begin{bmatrix}B_{11}&0\\B_{21}&0\end{bmatrix}P^T$

不能比X更优可证得

③ $\|B_{22}\|_F^2=0$.，用与①类似的 $W=Q\begin{bmatrix}B_{11}&0\\0&0\end{bmatrix}P^T$

不能比X更优证得

~~下面证明~~ ~~所以~~ 所以 ~~$B=\begin{pmatrix}B_{11}&0\\0&0\end{pmatrix}$~~. ~~所以 $\|\Omega_k-B_{11}\|_F^2=0$~~.

所以由 $\|A-Z\|_F^2=\|B_{22}\|_F^2\le\|\Omega_k-B_{11}\|_F^2+\|B_{12}\|_F^2$

$=\|A-X\|_F^2$.

且X为最优k秩近似，可得：$\|\Omega_k-B_{11}\|_F^2=0$.

即有最优k秩近似 $X=Q\Omega P^T=Q_k\Omega_kP_k^T=A_k$.

于是证得 A_k 是F范数下的最优k秩近似

Problem 3. 要证明： $\|A-A_k\|_2\le\|A-B\|_2$

由 $\|A\|_2=\max\limits_{\|x\|=1}\|Ax\|_2$. $\quad$ $\exists\, u^*\in Null(B)\cap V_{k+1}$ 其中 u^* 是单位向量

可知 $\|A-B\|_2^2=\max\limits_{\|u\|=1}\|(A-B)u\|_2^2\ge\|(A-B)u^*\|_2^2$. $\quad$ V_{k+1} 是 $(k+1)$ 维空间

$=\|Au^*\|_2^2$ $\quad$ Null(B)是B线性无关的空间

~~$=\|\sum_{i=1}^{r}u_i\sigma_iv_i^T\sum_{j=1}^{k+1}\sigma_jv_j\|_2^2$~~

~~$=\|\sum^{k+1}u_i\sigma_i$~~

$=\|\sum_{i=1}^{r}u_i\alpha_iv_i^T\sum_{j=1}^{k+1}\alpha_jv_j\|_2^2$

$=\|\sum_{i=1}^{k+1}\alpha_i\sigma_iu_i\|_2^2$

$\ge\sigma_{k+1}^2$

$=\|A-A_k\|_2^2$.

Problem 4. 对于 l_2 球. $\forall \varepsilon \le 1$.

证明：对 $\forall \varepsilon \le 1$，存在一个子集 $N_\varepsilon \subseteq S_u$，$|N_\varepsilon| = (\frac{4}{\varepsilon})^d$，使得 $\forall v \in S_u$，

$$\min_{x \in N_\varepsilon} \|v - x\| \le \varepsilon.$$

考虑一个圆的截面.

在圆周上任取一点作为直径为 ε 的圆的圆心，要找到对 $\forall v \in S_u$，能够覆盖 $\|v-x\| \le \varepsilon$ 的所有 x.

即找到能覆盖 $r = (1+\frac{\varepsilon}{2}) < 2$ 为半径的球体体积的覆盖方式.

又 $\because$

$$C \cdot (1+\frac{\varepsilon}{2})^d < 2^d C < (\frac{4}{\varepsilon})^d \cdot C(\frac{\varepsilon}{2})^d.$$

$\therefore$ 对 $\forall \varepsilon \le 1$，存在一个子集 $N_\varepsilon \subseteq S_u$，$|N_\varepsilon| = (\frac{4}{\varepsilon})^d$，使得 $\forall v \in S_u$，$\min_{x \in N_\varepsilon} \|v - x\| \le \varepsilon$

Problem 5. 对 $\forall v \in S_u$，可以把 $\vec{v}$ 分解为 $\vec{x_0}$ 和 $\vec{y_0}$.

即 $\vec{v} = \vec{x_0} + c_1 \vec{y_0}$ 其中 $\vec{x_0} \in N_\varepsilon$，$\|c_1 \vec{y_0}\|_2 \le \varepsilon$.

(由于在 S_u 上任一点能找到所属 ε-网). $\|\vec{y_0}\|_2 = 1$.

$\vec{y_0}$ 可以作类似的分解.

分解为. $\vec{y_0} = \vec{x_1} + \frac{c_2}{c_1} \vec{y_1}$，其中 $\vec{x_1} \in N_\varepsilon$. $\|\frac{c_2}{c_1} \vec{y_1}\| < \varepsilon$.

$\|\vec{y_1}\|_2 = 1$.

依次分解下去，即可得到 $\vec{v} = \vec{x_0} + c_1 \vec{x_1} + c_2 \vec{x_2} + \cdots$

其中 $c_i < \varepsilon^i$

Problem 6. 由于 $\forall v \in S_u$，v 可以表示为 $x_0 + c_1 x_1 + c_2 x_2 + \cdots$ 其中 $x_i \in N_\varepsilon$，$c_i \le \varepsilon^i$

所以有：对 N_ε 这样的 ε-net 作 JL 变换 B 后.

$\forall x \in N_\varepsilon$，$\varepsilon \le 1$ $(1-\varepsilon)\|x\|^2 \le \|B \cdot x\|^2 \le (1+\varepsilon)\|x\|^2$.

$\forall v \in S_u$，$v = x_0 + c_1 x_1 + c_2 x_2 + \cdots$

$$Bv = Bx_0 + c_1 Bx_1 + c_2 Bx_2 + \cdots$$

$$\|Bv\|^2 \le \|Bx_0\|^2 + \|c_1 Bx_1\|^2 + \cdots$$

$$\le (1+\varepsilon) \cdot \|x_0\|^2 + (1+\varepsilon) \cdot c_1 \|x_1\|^2 + \cdots$$

$$\le (1+\varepsilon) \cdot \|v\|^2.$$

$\|BV\|^2 \geq (1-\varepsilon)\|\cancel{\quad}\| \geq (1-\varepsilon)\|V\|^2$.

综上，可知对 ε-net 作 JL变换能得到 R^d 单位球上的JL变换

优点在于

Problem 7. 算法会在有限步内结束：证明.

每执行一次算法，得到k个类中心和对应的k个类 $H_1, \cdots, H_k$.

不妨设共有n个点，当确定k个类中心后，每个点所属类别也自动确定了。而从n个点中选k个类中心的选法只有 C_n^k 种。

在确定的k个类中心为最优的k个类中心（即 $\phi_X(C)=\sum_{x\in X}\min_{c\in C}\|C-x\|_2^2$ 最小）时，算法趋于稳定，算法停止。在中间的过程中，算法的执行不可能在两种不同的划分中来回横跳（否则与 $C_j \to \mu(H_j)$，应更靠近最终的类中心冲突），也不可能多次执行后仍是同一种划分（否则算法趋于稳定应当停止）。

Problem 8. 不妨设原来的类中心分别为 $\mu_1, \cdots \mu_k$，新增的一个类的类中心为 μ_{k+1}.

$\mu_1, \cdots \mu_k \in S \quad |S|=k. \quad \mu_1 \cdots \mu_{k+1} \in S' \quad |S'|=k+1.$

原损失函数：

$$\phi_X(S)=\sum_{x\in X}\min_{\mu_i\in S}\|\mu_i-x\|_2^2 \quad i=1,\dots k.$$

增加聚类数后的损失函数：$\phi_X(S)'=\sum_{x\in X}\min_{\mu_i\in S'}\|\mu_i-x\|_2^2 \quad i=1,2,\dots,k+1$

对于增加类数前后从第k类移至第(k+1)类的点x.

损失函数从 $\|x-\mu_k\|_2^2>0$ 变为 $\|x-\mu_{k+1}\|_2^2=0$.

损失函数并没有因k的增大而增大。

Problem 9. ① $\phi_X(C)=\sum_{x\in X}\min_{c\in C}\|c-x\|_2^2$. ~~不妨设使得$\phi_X(C)$最小化的C不唯一~~.

~~如C=C1、C2均可使$\phi_X(C)$最小化~~.

则 ~~$\phi_X(C_1)=\sum$~~ 由于对每个类，对类中心的损失函数.

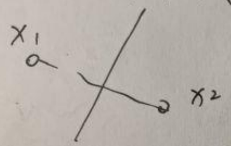
$$\phi_{X_{类1}}(c_{类1})=\sum_{i=1}^{n}\|x_i-c_{类1}\|_2^2=\sum_{i=1}^{n}\|x_i-\mu\|_2^2+n\|c_{类1}-\mu\|_2^2$$

$$\geq\sum_{i=1}^{n}\|x_i-\mu\|_2^2 \quad 其中\ \mu=\frac{\sum_{i=1}^{n}x_i}{n}.$$

所以每个类都取其重心为类中心时，单类的损失函数最小，总的损失函数也最小。使得 $\phi_X(C)$ 最小化的c唯一。

② 若取 l_1 范数，则使 $\phi_X'(C)$ 最小化的c不唯一.

如在二维平面有两点 x_1、x_2，找其聚类中心.



在 x_1、x_2 连线中垂线上各点均可使 $\phi_X'(C)$ 最小化

Problem 10. $Good_i=\{A_j \mid \phi_{A_j}(S_{i-1})\leq 10\,\phi_{A_j}(C_{opt})\}$.

$Bad_i=\{A_1,\cdots,A_k\}\setminus Good_i$.

$A=\{\phi_A(S_{i-1})\leq 20\,\phi_X(C_{opt})\}$.

$B=\{C_i\in Bad_i\}$

$$Pr[B\mid A^c]=\frac{\sum_{A_j\in Bad_i}\phi_{A_j}(S_{i-1})}{\phi_X(S_{i-1})}=1-\frac{\sum_{A_j\in Good_i}\phi_{A_j}(S_{i-1})}{\phi_X(S_{i-1})}$$

$$\geq 1-\frac{10\sum_{A_j\in Good_i}\phi_{A_j}(C_{opt})}{20\,\phi_X(C_{opt})}$$

$$\geq 1-\frac{1}{2}=\frac{1}{2}.$$

综上，$P[B\mid A^c]\geq\frac{1}{2}$.