数据科学与工程数学基础

作业提交规范及第 10 次作业

教师：黄定江
助教：陈诺、刘文辉

2022 年 5 月 8 日

作业提交规范

1. 作业提交形式：**练习本或笔记本**（建议统一使用一般的**练习本**即可，不接收以纸张的方式书写的作业）。另外，若作业包含代码部分，**请将代码文件压缩后**上传到**第 10 次作业代码传送门**。代码压缩文件命名格式："**hw10_ 代码 _ 学号 _ 姓名**"，命名示例：hw10_ 代码 _52215903014_ 刘文辉。其中，"hw10_ 代码"表示第 10 次作业代码。
2. 作业书写说明：
   (a) 可以讨论，禁止抄袭！
   (b) 练习本封面至少包含两方面信息：**姓名**和**学号**
   (c) 每一次的作业**请另起一页**，并在**第一行标明第几次作业**。例如"第 10 次作业"；
   (d) 每一题请**标注题号**，无需抄题，直接解答；
   (e) 题与题之间**请空一行**；
   (f) 不要求字好，但要求书写整体清晰易读。
3. 作业提交途径：纸质作业交给**学习委员**，由学习委员**按学号顺序**收齐后统一在截止日期前交到**助教实验室**。**单数周**布置的作业交到助教刘文辉处**数学馆西 109**；**双数周**布置的作业交到助教陈诺处**地理馆 353**。
4. 作业评分说明：正常提交作业的按照实际评分记录；逾期补交作业的根据逾期情况在实际评分基础上酌情扣分；未交作业的当次作业记为 0 分。

第 10 次作业

提交截至时间：**暂定 2022/04/01 下周五 20:00（晚上）**

理论部分

习题 1. 计算矩阵 $A = \begin{pmatrix} 5 & 2 & -4 \\ 2 & 1 & -2 \\ -4 & -2 & 5 \end{pmatrix}$ 的 *Cholesky* 分解。

解. （具体计算过程略）易求得 $L = \begin{pmatrix} \sqrt{5} & & \\ \frac{2}{\sqrt{5}} & \sqrt{\frac{1}{5}} & \\ -\frac{4}{\sqrt{5}} & -\frac{2}{\sqrt{5}} & 1 \end{pmatrix}$

则有 $A = LL^T$

习题 2. 完成以下阅读材料即可。

- **为什么需要 *LU* 分解？**

  为什么不直接求解方程组 $Ax = b$？如果了解 LU 分解所需要的计算量与高斯消去法解方程组所需的计算量，便可知计算复杂度均为 $O(n^3)$。似乎 LU 分解并没有任何优势。然而，在实际的工程问题中，我们经常会遇到的是这样的问题，需要求解这样一系列的方程组（例如时序的）：

  $$Ax = b_1,\ Ax = b_2,\ \cdots,\ Ax = b_r.$$

  其中的系数矩阵是不变的，只是右边的常数项在改变。这时，如果直接使用高斯消去法求解各个方程组的话，则对应的计算复杂度为 $O(rn^3)$。若使用 LU 分解，则只需要的运算量级为 $O(n^3 + rn^2)$，这是因为 LU 分解在解方程组时只需要 $O(n^2)$。

  或许，你会提出直接计算出 A^{-1}，然后再利用它去求解各个方程组。这时所需要的计算量的效果似乎与 LU 分解是相似的。事实上，在实际计算中往往不会直接计算矩阵的逆。这里给出其中一个原因：通常实际工程中矩阵 A 的维度很大，但有一个优点是它是稀疏的，例如一个“带状的”矩阵。使用 LU 分解，可以保持矩阵的稀疏性（因此，还可以提高运算速度）。然而，直接求解矩阵的逆，则会丢失矩阵的稀疏性。因此，在数据的存储上直接求逆存在明显的劣势。

- ***Gauss* 消去、*LU* 分解与 *Cholesky* 分解**

  还记得初中学过的求解二元一次方程组的消元法吗？这就是 *LU* 分解与 *Cholesky* 分解的全部。假设我们有如下二元一次方程组

$$\begin{cases} 2x_1 + 4x_2 = 5 \\ 4x_1 + 5x_2 = 3 \end{cases}$$

  现在我们考虑方程组的求解。在中学时，我们会通过将第一个等式乘以 -2 加到第二个等式，则得到

$$\begin{cases} 2x_1 + 4x_2 = 5 \\ 0x_1 - 3x_2 = -7 \end{cases}$$

  这个行变换对应的矩阵为 $\begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}$，它的逆矩阵就是 $\begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$。另外，变换后的系数矩阵为 $\begin{pmatrix} 2 & 4 \\ 0 & -3 \end{pmatrix}$。

  如果我们使用 *LU* 分解，则可得到

$$\begin{pmatrix} 2 & 4 \\ 4 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 4 \\ 0 & -3 \end{pmatrix}$$

  从上可以看出，高斯消去的变换矩阵的逆恰好对应 *LU* 分解的 *L* 矩阵，变换后的系数矩阵恰好对应 *U* 矩阵。

  如果我们使用 *Cholesky* 分解，则可得到

$$\begin{pmatrix} 2 & 4 \\ 4 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & -3 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$$

  从上可以看出，*Cholesky* 分解相当于将 *LU* 分解的 *U* 矩阵的对角线元素归一化。

  这个简单的例子是希望能够加深大家对 *LU* 分解和 *Cholesky* 分解。当然，上述叙述存在不严谨的地方，留给大家自己去思考。

习题 3. 求矩阵 $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}$ 的 *SVD* 分解。

解. $A^T A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$

特征值为 *1* 对应的特征向量为 $(1, -1)^T$

特征值为 *3* 对应的特征向量为 $(1, 1)^T$

所以 $\Sigma = \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$, $V^T = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$

$U = A\Sigma^{\dagger}V = \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{pmatrix}$

所以

$$A = \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

习题 4. 求矩阵 $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}$ 在 F 范数下秩为 *1* 的最优近似。(注：根据 *Eckart-Young-Mirsky* 定理，即为只保留最大的秩所对应的矩阵)

解. 最佳秩 *1* 近似：

$$\begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

习题 5. 假设 D 是一个 $n \times d$ 的矩阵，矩阵 B 是 $(n+d) \times (n+d)$ 定义为

$$B = \begin{pmatrix} 0 & D^{\mathrm{T}} \\ D & 0 \end{pmatrix}$$

显然 B 是对称矩阵。请证明矩阵 B 的对角化会产生 D 的奇异值分解所需要的所有信息。

解. D 的奇异值分解所需要的信息为 D^TD 的特征向量和 DD^T 的特征向量，以及对应的特征值。现假设对应于特征值为 $\lambda^2(\lambda > 0)$ 的 D^TD 的特征向量为 $\boldsymbol{x}_1(\|\boldsymbol{x}_1\|_2 = 1)$,$DD^T$ 的特征向量为 $\boldsymbol{x}_2(\|\boldsymbol{x}_2\|_2 = 1)$. 因此，$D^TD\boldsymbol{x}_1 = \lambda^2\boldsymbol{x}_1$ 以及 $DD^T\boldsymbol{x}_2 = \lambda^2\boldsymbol{x}_2$。故

$$(DD^T)D\boldsymbol{x}_1 = \lambda^2 D\boldsymbol{x}_1$$

所以存在 k 使得 $D\boldsymbol{x}_1 = k\boldsymbol{x}_2$。由于 $\|\boldsymbol{x}_1\|_2 = \|\boldsymbol{x}_2\|_2 = 1$，可得 $k = \lambda$，即 $D\boldsymbol{x}_1 = \lambda\boldsymbol{x}_2$。同理有 $D^T\boldsymbol{x}_2 = \lambda\boldsymbol{x}_1$。

下面证明 $\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix}$ 是矩阵 B 的特征值为 λ 的特征向量。易计算

$$B\boldsymbol{x} = \begin{pmatrix} 0 & D^{\mathrm{T}} \\ D & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix} = \begin{pmatrix} D^T\boldsymbol{x}_2 \\ D\boldsymbol{x}_1 \end{pmatrix} = \lambda \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix}$$

因此，D 的奇异值分解信息包含在 B 的对角化过程中。

实践部分

习题 6. 任选一张图片，使用 *SVD* 分解对图片进行压缩，分别展示取 *1%*、*2%*、*10%*、*50%* 奇异值的结果。提示：可在 *numpy* 包中可以使用 *'np.linalg.svd'*对一个 *'np.matrix'*对象进行 *SVD* 分解。需要上传代码和结果。