矩阵分析下篇

一：SVD理论(奇异值分解)

SVD是特征分解的广义化。特征分解本质上是对方阵而言，而SVD对任意矩阵而言。

任何$\mathbf{A} \in \mathbf{R}^{\mathbf{m \times n}}$的矩阵都可以被分解为：

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

有几点需要注意：

1. $\mathbf{U} \in \mathbf{R}^{\mathbf{m \times m}}$和$\mathbf{V} \in \mathbf{R}^{\mathbf{n \times n}}$是正交矩阵。这两个矩阵都是方阵。
2. Σ是一个$m \times n$的矩阵，和矩阵$\mathbf{A}$的样子是一样的。

定义 $p = min(m, n)$,则有

$$\Sigma(\mathrm{i}, \mathrm{j}) = \begin{cases} \sigma_{\mathrm{i}}, & \mathrm{i} = \mathrm{j} \\ 0, & \mathrm{i} \neq \mathrm{j} \end{cases}$$
$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\mathrm{p}} \geq 0$$

因此：

$$\Sigma = \begin{bmatrix} \Sigma_p & 0 \\ 0 & 0 \end{bmatrix}$$

二：SVD分解的三种形式

2.1 第一种形式：分块型

假定r表示非零奇异值的数量，即$r = rank(\mathbf{A})$，对奇异值进行排序：

$$\underbrace{\sigma_1 \geq \ldots \geq \sigma_{\mathrm{r}}}_{r} > \underbrace{\sigma_{\mathrm{r+1}} = \ldots = \sigma_{\mathrm{p}}}_{\mathrm{p-r}} = 0$$

说明：

- 矩阵本来有p个奇异值（p是m和n中的最小值），但是有r个非零奇异值。

分块式SVD可以表示为：

$$
\begin{aligned}
\mathbf{A} &= \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\top} \\
&= \underbrace{[\mathrm{U}_1 \mathrm{U}_2]}_{\mathbf{U}} \underbrace{\begin{bmatrix} \Sigma_1 & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} \mathbf{V}_1^{\top} \\ \mathbf{V}_2^{\top} \end{bmatrix}}_{\mathbf{V}^{\mathrm{T}}}
\end{aligned}
$$

其中：$\Sigma_1 = \operatorname{diag}(\sigma_1, \ldots, \sigma_\mathrm{r})$且有$\sigma_r > 0$,$\Sigma_1$是由正奇异值组成的对角矩阵。$\mathrm{U}_1 \in \mathbb{R}^{\mathrm{mxr}}$对应于r个非零奇异值的左奇异矩阵，$\mathrm{U}_2 \in \mathbb{R}^{\mathrm{m}\times(\mathrm{m-r})}$对应于0奇异值的左奇异矩阵。

2.2 第二种形式：迷你型

$$
\begin{aligned}
\mathbf{A} &= \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{T}} \\
&= \underset{r \; m-r}{[\mathbf{U}_1 \mathbf{U}_2]} \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^{\mathrm{T}} \\ \mathbf{V}_2^{\mathrm{T}} \end{bmatrix} \\
&= \mathbf{U}_1 \boldsymbol{\Sigma}_1 \mathbf{V}_1^{\mathrm{T}}
\end{aligned}
$$

2.3 第三种形式：外积型

$$
\begin{aligned}
\mathbf{A} &= \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathbf{T}} \\
&= \sum_{i=1}^{r} \sigma_i \mathbf{u}_{\mathbf{i}} \mathbf{v}_{\mathbf{i}}^{\mathbf{T}}
\end{aligned}
$$

三：SVD和特征分解的关系

- 已知$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathbf{T}}$,则：

  $$\mathbf{A}\mathbf{A}^{\mathbf{T}} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathbf{T}}\mathbf{V}\boldsymbol{\Sigma}^{\mathbf{T}}\mathbf{U}^{\mathbf{T}} = \mathbf{U}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\mathbf{T}}\mathbf{U}^{\mathbf{T}} = \mathbf{U}\boldsymbol{\Lambda}_{\mathbf{L}}\mathbf{U}^{\mathbf{T}}$$

  其中

  $$\Lambda_L = \begin{bmatrix} \Sigma_1^2 & 0 \\ 0 & 0 \end{bmatrix}$$

所以可以得出：U是$\mathbf{A}\mathbf{A}^{\mathbf{T}}$的特征矩阵，$\sigma_1^2, \ldots, \sigma_\mathrm{r}^2, 0 \ldots 0$是特征值。

因此：

$$\sigma_\mathrm{k} = \sqrt{\lambda_\mathrm{k}\left(\mathrm{A}\mathrm{A}^{\top}\right)} = \sqrt{\lambda_\mathrm{k}\left(\mathrm{A}^{\top}\mathrm{A}\right)}$$

所以：A矩阵的奇异值是$\mathbf{A}\mathbf{A}^{\mathbf{T}}$矩阵特征值开根号。

同理：$\mathbf{V}$是$\mathbf{A}^{\mathbf{T}}\mathbf{A}$的特征矩阵。$\sigma_1^2, \ldots, \sigma_\mathrm{r}^2, 0 \ldots 0$是特征值。

说明:

- $\mathbf{A}\mathbf{A}^{\mathbf{T}}$是$m \times m$矩阵，有r个正特征值，为$\sigma_1^2, \ldots, \sigma_{\mathrm{r}}^2$,这些特征值开根号，即为矩阵A的r个奇异值。然后有m-r个0奇异值。
- $\mathbf{A}^{\mathbf{T}}\mathbf{A}$是$n \times n$矩阵，有r个正特征值，为$\sigma_1^2, \ldots, \sigma_{\mathrm{r}}^2$,这些特征值开根号，即为矩阵A的r个奇异值。然后有n-r个0奇异值。
- 所以$\mathbf{A}\mathbf{A}^{\mathbf{T}}$和$\mathbf{A}^{\mathbf{T}}\mathbf{A}$正特征值是相同的，区别在于0特征值的个数。

四：SVD和子空间的关系

4.1 和列空间的关系

- 列空间：$\mathrm{C}(\mathbf{A}) = \{\mathbf{y} | \mathbf{y} = \mathbf{A}\mathbf{x}\}$

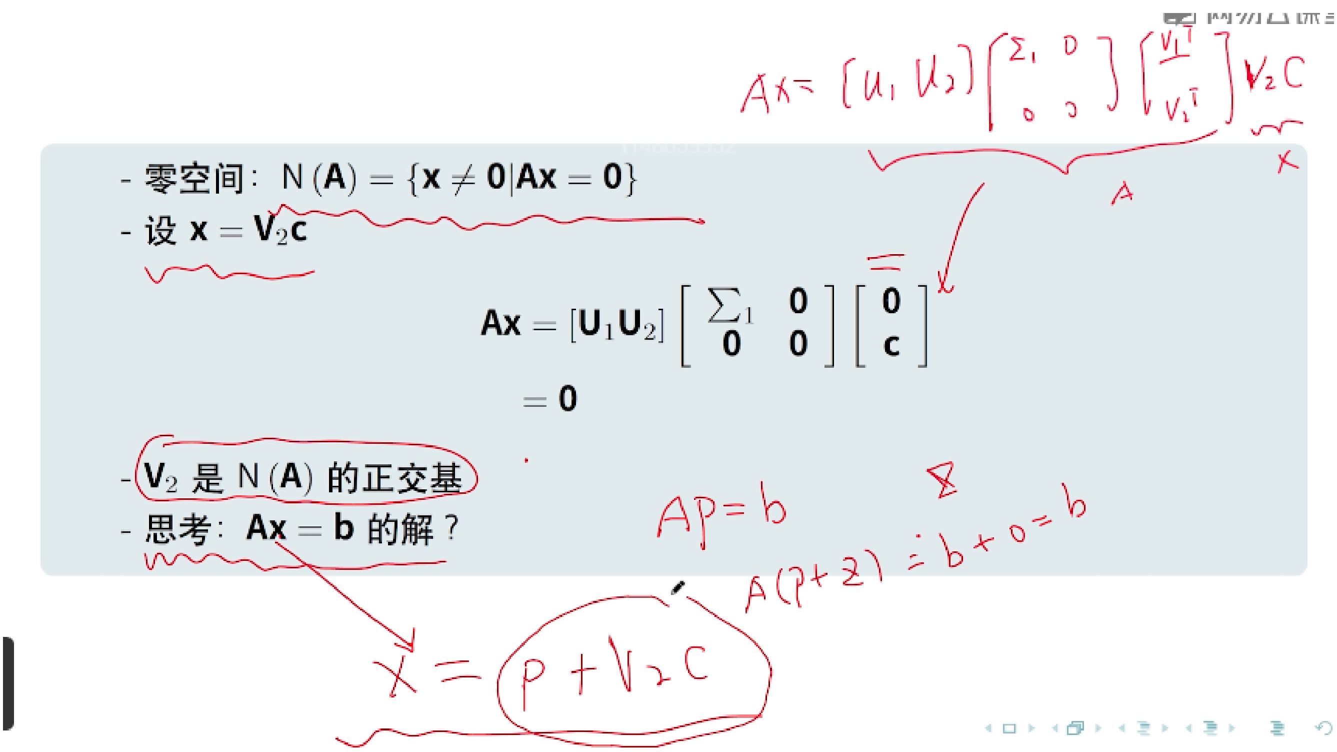
$$\begin{aligned}\mathbf{A}\mathbf{x} &= [\mathbf{U}_1\mathbf{U}_2]\begin{bmatrix}\Sigma_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{V}_1^T \\ \mathbf{V}_2^T\end{bmatrix}\mathbf{x} \\ &= [\mathbf{U}_1\mathbf{U}_2]\begin{bmatrix}\Sigma_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{c}_1 \\ \mathbf{c}_2\end{bmatrix} \\ &= [\mathbf{U}_1\mathbf{U}_2]\begin{bmatrix}\Sigma_1\mathbf{c}_1 \\ \mathbf{0}\end{bmatrix} \\ &= \mathbf{U}_1(\Sigma_1\mathbf{c}_1)\end{aligned}$$

- $\mathrm{C}(\mathbf{A}) = \mathrm{C}(\mathbf{U}_1)$

说明:

- $C(A)$是列空间，是A中列向量的线性组合。
- 因为x是任意向量，所以把$\mathbf{V}_1^{\mathbf{T}}\mathbf{x}$定义为$\mathbf{c}_1$,$\mathbf{V}_2^{\mathbf{T}}\mathbf{x}$定义为$\mathbf{c}_2$,$\mathbf{c}_1$和$\mathbf{c}_2$都是任意的列向量。
- 最后一步中：Σ_1是$r \times r$的对角矩阵,所以$\mathbf{\Sigma}_1\mathbf{c}_1$是任意的列向量。

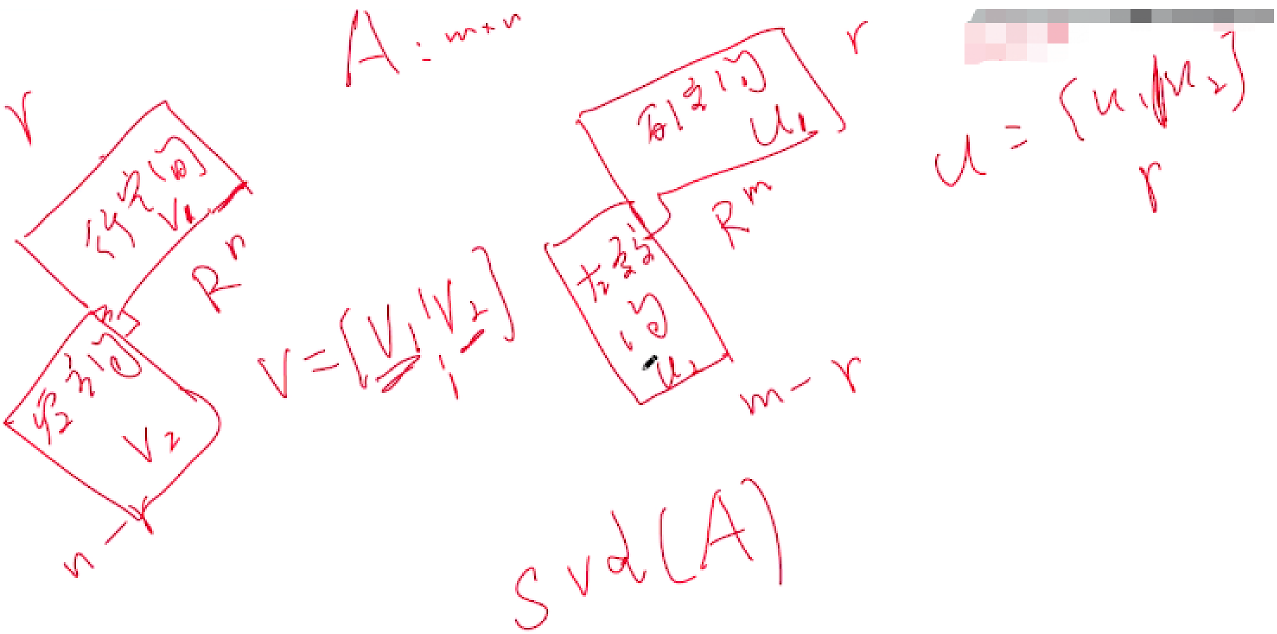
4.2 和零空间的关系



说明：

- c为一个任意的列向量。
- 因为$\mathbf{V}$是正交矩阵，所以$\mathbf{V_1^T}$中的行向量和$\mathbf{V_2}$中的列向量都是正交的，相乘得0.所以$\mathbf{V_2^T V_2} = \mathbf{I}$
- $\mathbf{x} = \mathbf{V_2 c}$说明x可以由$V_2$中得列进行线性组合而来。
- $\mathbf{Ax} = \mathbf{b}$的解为特解（p）加通解（z)。其中通解为$\mathbf{V_2 c}$

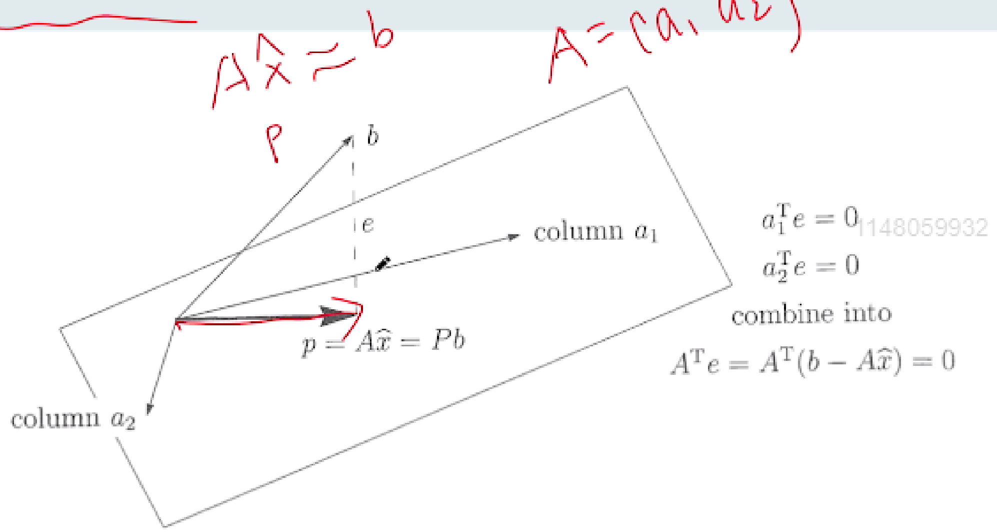
4.3 再看四个子空间



五：其它知识

5.1 投影

> **投影角度看 Ax ≠ b**

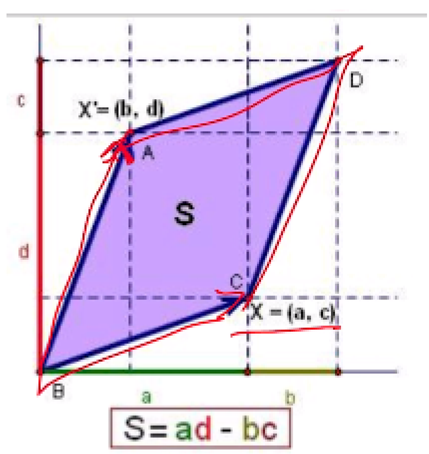


$min||\mathbf{Ax}-\mathbf{b}||_2^2$，本质上是找到一个x，使得在A的列空间上的向量（Ax）与b向量之间的模长最小。由投影的观点可以发现，当模长最小时，这个**最优的x^*恰好使得b向量在列空间上的投影与Ax^*相等**。所以可知

$$A^T(b-A\hat{x})=0$$
$$A^Tb=A^TA\hat{x}$$
$$\hat{x}=(A^TA)^{-1}A^Tb$$

在优化部分，求解该最优值，我们采用梯度等于0的方法，也是求得上述结果。因此可以与线性回归求解系数进行对比理解。

5.2 行列式的几何意义



$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad-bc$$

行列式为行列式中两个列向量围成的平行四边形的面积。对应到三维空间，就是一个平行六面体的面积（平行六面体可以理解为一个长方体歪了一下）

六： SVD实际应用

- 给定矩阵 $\mathbf{A} \in \mathbf{R}^{m\times n}$，其秩为 r，需找一个矩阵 $\mathbf{B} \in \mathbf{R}^{m\times n}$，其秩为 $k < r$，使其能够最接近 $\mathbf{A}$
- 若 $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T = \sum_{i=1}^{p}\sigma_i\mathbf{u}_i\mathbf{v}_i^T$，定义

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T = \sum_{i=1}^{k}\sigma_i\mathbf{u}_i\mathbf{v}_i^T \tag{14}$$

- 应用：Principal Component Analysis(PCA)，维数减少，数据压缩等等

因为A的秩为r，所以A有r个正奇异值。对这r个奇异值按照从大到小的顺序进行排列，取出前面比较大的k个奇异值，按照上图中的公式14进行计算，既可以得到降维后的矩阵。

所以步骤如下：

1. 对原矩阵进行奇异值分解
2. 进行奇异值排序
3. 选择较大的奇异值计算变换之后的矩阵。

SVD和PCA的比较：

- 在PCA中，我们首先对原矩阵求协方差矩阵，然后对协方差矩阵进行特征分解，求得特征值和对应的特征向量，然后选择较大的特征值对应的特征向量组成变换矩阵，进行变换。
- 在SVD中，我们直接对原矩阵进行奇异值分解，然后选择较大的奇异值及其对应的奇异向量，直接计算转换之后的矩阵。
- 在实际中，一定会用奇异值分解，因为从数值计算的角度来看，奇异值分解更加稳定。