奇异值分解

1.奇异值分解的定义与性质

1.1 奇异值分解的定义：

定义（奇异值分解）：

矩阵的奇异值分解是指，任何一个非零$m \times n$实数矩阵A都可以表示为以下三个矩阵的乘积的形式，即矩阵的因子分解：

$$A = U\Sigma V^T \tag{1}$$

其中，U是m阶正交矩阵，V是n阶正交矩阵，Σ是由降序排列的非负对角元素组成的$m \times n$对角矩阵。满足：

$$UU^T = I \tag{2}$$

$$VV^T = I \tag{3}$$

$$\Sigma = diag(\sigma_1, \sigma_2, \ldots, \sigma_p) \tag{4}$$

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0 \tag{5}$$

$$p = \min(m, n) \tag{6}$$

$U\Sigma V^T$称为矩阵A的奇异值分解，σ_i称为A的奇异值，U的列向量称为左奇异向量，V的列向量称为右奇异向量。

1.2 奇异值分解的计算：

给定$m \times n$阶矩阵A，可以根据以下步骤计算出其奇异值分解：

1. 计算出n阶对称矩阵$W = A^TA$

   求出W的特征值λ_i并且从大到小排列：

   $$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0 \tag{7}$$

   根据特征值分别求出特征向量并单位正交化，得到特征向量：

   $$v_i, Wv_i = \lambda_i v_i, i = 1, 2, \ldots, n \tag{8}$$

2. 求正交矩阵V：

   若特征值中非零个数为r，即：

   $$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r > 0, \lambda_{r+1} = \lambda_{r+2} = \ldots = \lambda_n = 0 \tag{9}$$

   令：

   $$V_1 = [v_1 \;\; v_2 \;\; \ldots \; v_r],\, V_2 = [v_{r+1} \;\; v_{r+2} \ldots v_n] \tag{10}$$

   $$V = [V_1 \quad V_2] \tag{11}$$

   则有：

   $$A = AVV^T = AV_1V_1^T + AV_2V_2^T = AV_1V_1^T \tag{12}$$

3.

4. 求$m \times n$矩阵Σ:

   令$\sigma_i = \sqrt{\lambda_i}$，则有:

$$\sigma_1 \geq \sigma \geq \ldots \geq \sigma_r > 0, \sigma_{r+1} = \sigma_{r+2} = \ldots = \sigma_n = 0 \tag{13}$$

   令:

$$\Sigma_1 = diag(\sigma_1, \sigma_2, \ldots, \sigma_n) \tag{14}$$

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \tag{15}$$

5. 求正交矩阵U:

   令:

$$u_j = \frac{1}{\sigma_j} A v_j, \quad j = 1, 2, \ldots, r \tag{16}$$

$$U_1 = [u_1 \;\; u_2 \; \ldots \; u_r] \tag{17}$$

   则:

$$AV_1 = U_1 \Sigma_1 \tag{18}$$

   求A^T的零空间$N(A^T)$的一组标准正交基$\{u_{r+1}, u_{r+2}, \ldots, u_m\}$，令:

$$U_2 = [u_{r+1} \;\; u_{r+2} \; \ldots \; u_m] \tag{19}$$

   则:

$$U = [U_1 \quad U_2] \tag{20}$$

5. 得到奇异值分解:

$$A = U \Sigma V^T \tag{21}$$

1.3 紧奇异值分解与截断奇异值分解

上面的过程得到的奇异值分解称为完全奇异值分解:

1. ***紧奇异值分解:***

$$A = U_r \Sigma_r V_r \tag{22}$$

其中，U_r是由U的前r列构成的$m \times r$矩阵，Σ_r是由Σ前r行和r列构成的$r \times r$对角矩阵，V_r是由V前r列构成的$n \times r$矩阵。

2. ***截断奇异值分解:***

$$A = U_k \Sigma_k V_k, \quad 0 < k < r \tag{23}$$

其中，U_k是由U的前k列构成的$m \times k$矩阵，Σ_k是由Σ前k行和k列构成的$k \times k$对角矩阵，V_k是由V前k列构成的$n \times k$矩阵。

1.4 几何解释

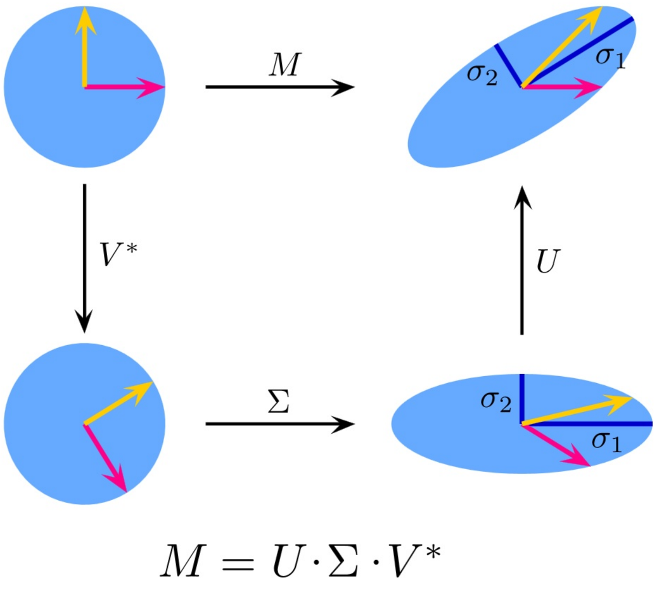
如果将$m \times n$矩阵A看成是一个n维空间R^n到m维空间R^m的一个线性变换，即：

$$x \in R^n, Ax \in R^m \tag{24}$$

奇异值分解的几何意义就是，任何一个线性变换都可以分解为三个简单的变换：

1. 一个左边系的旋转或反射：$x \to V^T x$
2. 一个坐标轴的缩放变换：$V^T x \to \Sigma V^T x$
3. 另一个坐标系的旋转或反射变换：$\Sigma V^T x \to U \Sigma V^T x$

如下图示：



2.奇异值分解与矩阵近似

奇异值分解也是一种矩阵近似的方法，这个近似在弗罗贝尼乌斯范数意义下的近似，矩阵的弗罗贝尼乌斯范数是向量L_2范数的推广，对应着机器学习中的平方损失函数。

2.1 弗罗贝尼乌斯范数

定义（弗罗贝尼乌斯范数）：

设矩阵$A \in R^{m \times n}$，$A = [a_{ij}]_{m \times n}$，定义矩阵$A$的弗罗贝尼乌斯范数为：

$$||A||_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} (a_{ij})^2 \right)^{\frac{1}{2}} \tag{25}$$

引理：

设矩阵$A \in R^{m \times n}$，其奇异值分解为$U \Sigma V^T$，其中$\Sigma = diag(\sigma_1, \sigma_2, \ldots, \sigma_n)$，则：

$$||A||_F = (\sigma_1^2 + \sigma_2^2 + \ \ldots \ + \sigma_n^2)^{\frac{1}{2}} \tag{26}$$

2.2 矩阵的最优近似

奇异值分解是在平方损失意义下，对矩阵的最优近似，即数据压缩

定理：

设矩阵$A \in R^{m \times n}$，矩阵的秩$rank(A) = r$，并设M为$R^{m \times n}$中所有秩不超过k的矩阵集合，$0 < k < r$，则存在一个秩为k的矩阵$X \in M$，使得：

$$||A - X||_F = \min_{S \in M} ||A - S||_F \tag{27}$$

称矩阵X为矩阵A在弗罗贝尼乌斯范数意义下的最优近似。

定理：

设矩阵$A \in R^{m \times n}$，有奇异值分解$A = U\Sigma V^T$，矩阵的秩$rank(A) = r$，并设M为$R^{m \times n}$中所有秩不超过k的矩阵集合，$0 < k < r$，若一个秩为k的矩阵$X \in M$满足：

$$||A - X||_F = \min_{S \in M} ||A - S||_F \tag{28}$$

则有：

$$||A - X||_F = (\sigma_{k+1}^2 + \sigma_{k+2}^2 + \ \dots \ + \sigma_n^2)^{\frac{1}{2}} \tag{29}$$

特别的，若$A' = U\Sigma' V^T$，其中：

$$\Sigma' = \begin{bmatrix} \Sigma_k & 0 \\ 0 & 0 \end{bmatrix} \tag{30}$$

则有：

$$||A - A'||_F = (\sigma_{k+1}^2 + \sigma_{k+2}^2 + \ \dots \ + \sigma_n^2)^{\frac{1}{2}} = \min_{S \in M} ||A - S||_F \tag{31}$$

所以，紧奇异值分解是在弗罗贝范数意义下对矩阵的无损压缩，截断奇异值分解是有损失的压缩。截断奇异值分解得到的矩阵的秩为k，通常远小于原始矩阵的秩r，所以是由低秩矩阵实现了对原始矩阵的压缩。