字典更新与K-SVD

凯鲁嘎吉 - 博客园 http://www.cnblogs.com/kailugaji/

1．矩阵的奇异值分解 (Singular Value Decomposition, SVD)

对于每一个秩为 r 的矩阵 $A \in C^{m\times n}$，存在正交矩阵 $U_{m\times m}$, $V_{n\times n}$ 与一个对角阵 $\Sigma_{r\times r}= \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r)$，如下

$$A = U\left[\begin{array}{c:c} \Sigma_r & O \\ \hdashline O & O \end{array}\right]V^{-1} = U\left[\begin{array}{c:c} \Sigma_r & O \\ \hdashline O & O \end{array}\right]V^{H}$$

$$\underbrace{\qquad}_{n\text{-}r}$$

这种分解方式称为矩阵 A 的奇异值分解（SVD）.

$$\overset{U}{\begin{bmatrix} -.40 & .916 \\ .916 & .40 \end{bmatrix}} \times \overset{\Sigma}{\begin{bmatrix} 5.39 & 0 \\ 0 & 3.154 \end{bmatrix}} \times \overset{V^T}{\begin{bmatrix} -.05 & .999 \\ .999 & .05 \end{bmatrix}} = \overset{A}{\begin{bmatrix} 3 & -2 \\ 1 & 5 \end{bmatrix}}$$

$$\overset{U}{\begin{bmatrix} -.39 & -.92 \\ -.92 & .39 \end{bmatrix}} \times \overset{\Sigma}{\begin{bmatrix} 9.51 & 0 & 0 \\ 0 & .77 & 0 \end{bmatrix}} \times \overset{V^T}{\begin{bmatrix} -.42 & -.57 & -.70 \\ .81 & .11 & -.58 \\ .41 & -.82 & .41 \end{bmatrix}} = \overset{A}{\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}}$$

令 $A\in C^{n\times n}$，如果 $A^HA=AA^H$ (可交换)，则 A 被称为正规矩阵(normal matrix)，其中 A^H 是矩阵 A 的共轭转置(conjugate transpose).

若 $A^HA=AA^H=E$，则 A 被称为酉矩阵(unitary matrix)，即它的共轭转置也是它的逆矩阵。

若 $A^H=A$，则 A 被称为 H 矩阵(Hermitian matrix).

例：求矩阵 A 的共轭转置 A^H

$$A=\begin{bmatrix}1+i & -i & 0\\ 2 & 3-2i & i\end{bmatrix}$$

解：我们有复共轭

$$\overline{A}=\begin{bmatrix}1-i & i & 0\\ 2 & 3+2i & -i\end{bmatrix}$$

（实部与原实部相等，虚部大小相等，符号相反。）

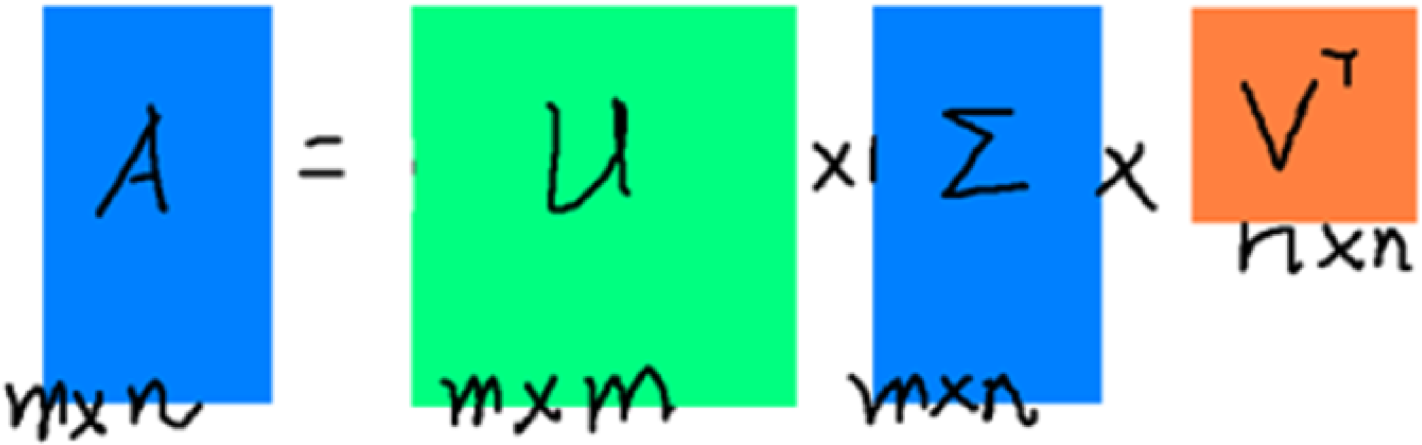
因此共轭转置是

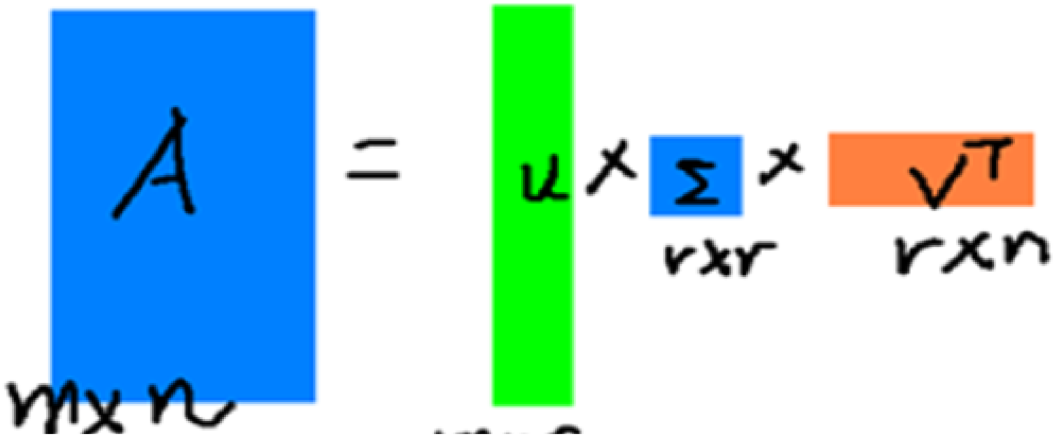
$$A^H=\overline{A}^T=\begin{bmatrix}1-i & 2\\ i & 3+2i\\ 0 & -i\end{bmatrix}$$

$$\sigma_i = \sqrt{\lambda_i}\,, \quad i = 1,2,\ldots,r.$$

A 可以分解为:

$$A = U\begin{bmatrix} \Sigma_r & O \\ O & O \end{bmatrix}V^{-1} = U\begin{bmatrix} \Sigma_r & O \\ O & O \end{bmatrix}V^{H}$$





奇异值分解有广泛的用途，例如对于低秩矩阵近似(low-rank matrix approximation)问题，给定一个秩为 r 的矩阵 $\mathbf{A}$，欲求其最优 k 秩近似矩阵 $\widetilde{\mathbf{A}}$，$k \leqslant r$，该问题可形式化为

$$
\begin{aligned}
\min_{\widetilde{\mathbf{A}} \in \mathbb{R}^{m \times n}} \quad & \|\mathbf{A} - \widetilde{\mathbf{A}}\|_F \\
\text{s.t.} \quad & \operatorname{rank}(\widetilde{\mathbf{A}}) = k \ .
\end{aligned}
\tag{A.34}
$$

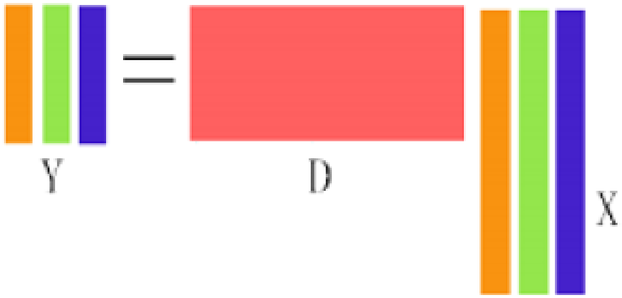
奇异值分解提供了上述问题的解析解：对矩阵 $\mathbf{A}$ 进行奇异值分解后，将矩阵 $\mathbf{\Sigma}$ 中的 $r-k$ 个最小的奇异值置零获得矩阵 $\mathbf{\Sigma}_k$，即仅保留最大的 k 个奇异值，则

$$
\mathbf{A}_k = \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{V}_k^{\mathrm{T}} \tag{A.35}
$$

就是式(A.34)的最优解，其中 $\mathbf{U}_k$ 和 $\mathbf{V}_k$ 分别是式(A.33)中的前 k 列组成的矩阵。这个结果称为 Eckart-Young-Mirsky 定理。

2．字典更新方法

设矩阵 Y 为样本集，由 N 个样本组成，每个样本由 n 个特征表示，即 Y 的大小为(n * N)。所谓稀疏表示，就是找到一组向量基，将此组向量基进行线性组合类表示矩阵(样本集)。即 Y=DX，其中 D 为字典，大小为(n*K)，X 为系数矩阵，大小为(K*N)，K 为字典的个数。



固定字典
Pre-constructed dictionary

DCT字典，Haar字典，小波字典，Gabor字典等。
优点：计算量低，使用简单。
缺点：不能保证表达的稀疏程度。只适用于部分类型的图像。

学习字典
Learned dictionary

从学习的角度建立字典。
优点：适用于任何类型的图像。
缺点：计算复杂度高。局限于低维度信号(小patch)。

超完备字典 Overcomplete Dictionary

字典的原子数量大于特征维度。

主要目标

寻找最佳的字典 D，同时使 X 系数矩阵达到稀疏最大。系数矩阵中，0 元素越多，越稀疏，即目标是用更少的原子线性组合来逼近原始矩阵。我这里系数矩阵指的是待优化的稀疏矩阵。

假设随机初始化 D，已知 Y，用 OMP(正交匹配追踪算法)等方法求出稀疏系数 X，以下方法用于求解未知变量 D。

OMP 算法具体步骤：

输　入：字典 $D=\{d_r\}r\in[1,L]$,信号 x,稀疏度 K；

初始化：残差 $r_0=x$,支撑索引集 $\Lambda_0=\phi,k=1$；

过　程：在第 k 次迭代循环，运行步骤 1-4：

步骤 1：通过运算得出支撑索引：$\lambda_k=\arg\max\limits_{i=1,\ldots,N}|\langle r_{k-1},d_i\rangle|$；

步骤 2：引入信号支撑集 $\Lambda_k=\Lambda_{k-1}\cup\{\lambda_k\}$；

步骤 3：更新残差 $r_k=g-D_{\lambda_k}\left(D_{\lambda_k}^T D_{\lambda_k}\right)^{-1}D_{\lambda_k}^T g$；

步骤 4：$k=k+1$，判断是否满足迭代终止条件 $k=L$,若满足,则停止迭代；若不满足，则返回步骤 1。

输　出：支撑索引集 $\Lambda_k=\Lambda_{k-1}$，稀疏系数 $b=D_{\lambda_k}\left(D_{\lambda_k}^T D_{\lambda_k}\right)^{-1}D_{\lambda_k}^T g$。

2.1 最优方向法 (Method of Optimal Directions, MOD)

$$
\begin{aligned}
&\min \quad \| Y - DX \|_F^2 \\
&s.t. \ \sum_i \| x_i \|_0 \le T_0
\end{aligned}
\tag{1}
$$

MOD 是早期的基于样本学习的字典学习算法. 设目标函数中 X 已知，信号的误差定义如下：

$$
\| E \|_F^2 = \| Y - DX \|_F^2 \tag{2}
$$

MOD 算法更新字典的策略就是实现表征误差最小化，所以公式两端对未知变量 D 求偏导，会推导出 $(Y - DX)X^T = 0$

整个字典的更新过程如下：

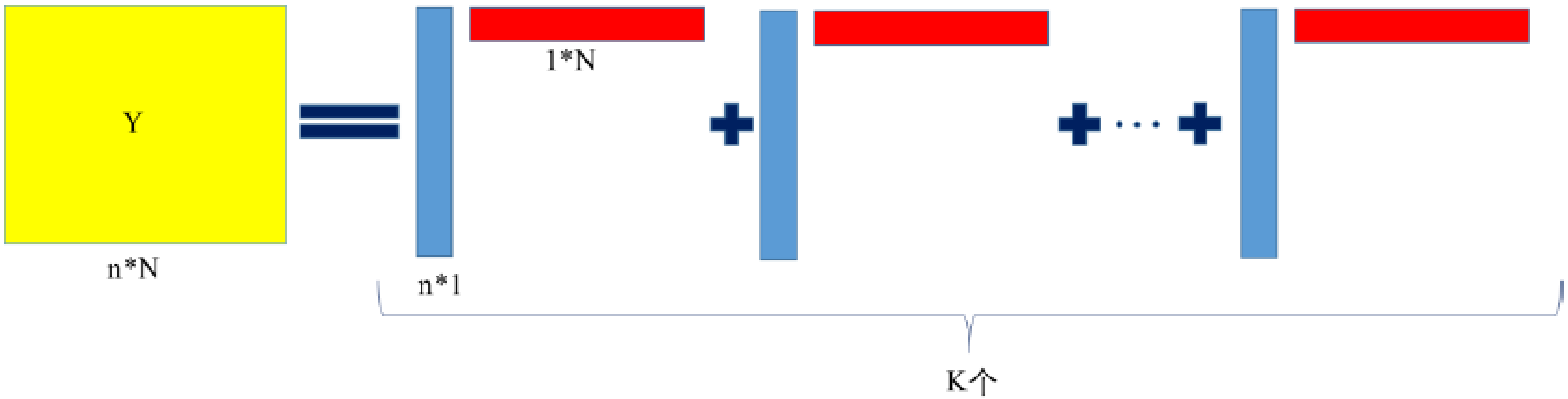
$$
D^{(n+1)} = Y(X^{(n)})^T \cdot (X^{(n)}(X^{(n)})^T)^{-1} \tag{3}
$$

一般 MOD 算法需要几十次迭代才能收敛，运算中需要对矩阵求逆，造成计算量过大。

2.2 标准正交基联合(Unions of Orthonormal Bases, UOB)

考虑一个由正交基组成的字典 D，对 D 进行按列划分，$D=[\mathrm{D}_1,\mathrm{D}_2,\ldots,\mathrm{D}_K]$，稀疏系数 X 进行按行划分，$X=\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_K \end{bmatrix}$

目标就是 Y=DX，即



UOB 算法的思想是按顺序一列一列地更新标准正交基 D_i。

$$\min \ \|Y-DX\|_F^2=\|(Y-\sum_{i\neq k}D_iX_i)-D_kX_k\|_F^2=\|E_k-D_kX_k\|_F^2 \\ s.t.\ D^TD=I. \tag{4}$$

该问题的增广拉格朗日函数为：

$$L(D,\mu)=\|Y-DX\|_F^2+Tr[\mu(D^TD-I)] \tag{5}$$

其中 u 是 N*N 的拉格朗日乘子矩阵，I 为单位阵。

令 $Z=YX^T, Q(\mu)=XX^T+(\mu+\mu^T)/2$，则式(7)可写为 $D^{-1}=Q(\mu)Z^{-1}$ 是一个正交阵，即

$$Q(\mu)Z^{-1}(Z^{-1})^T Q^T(\mu)=I \tag{8}$$

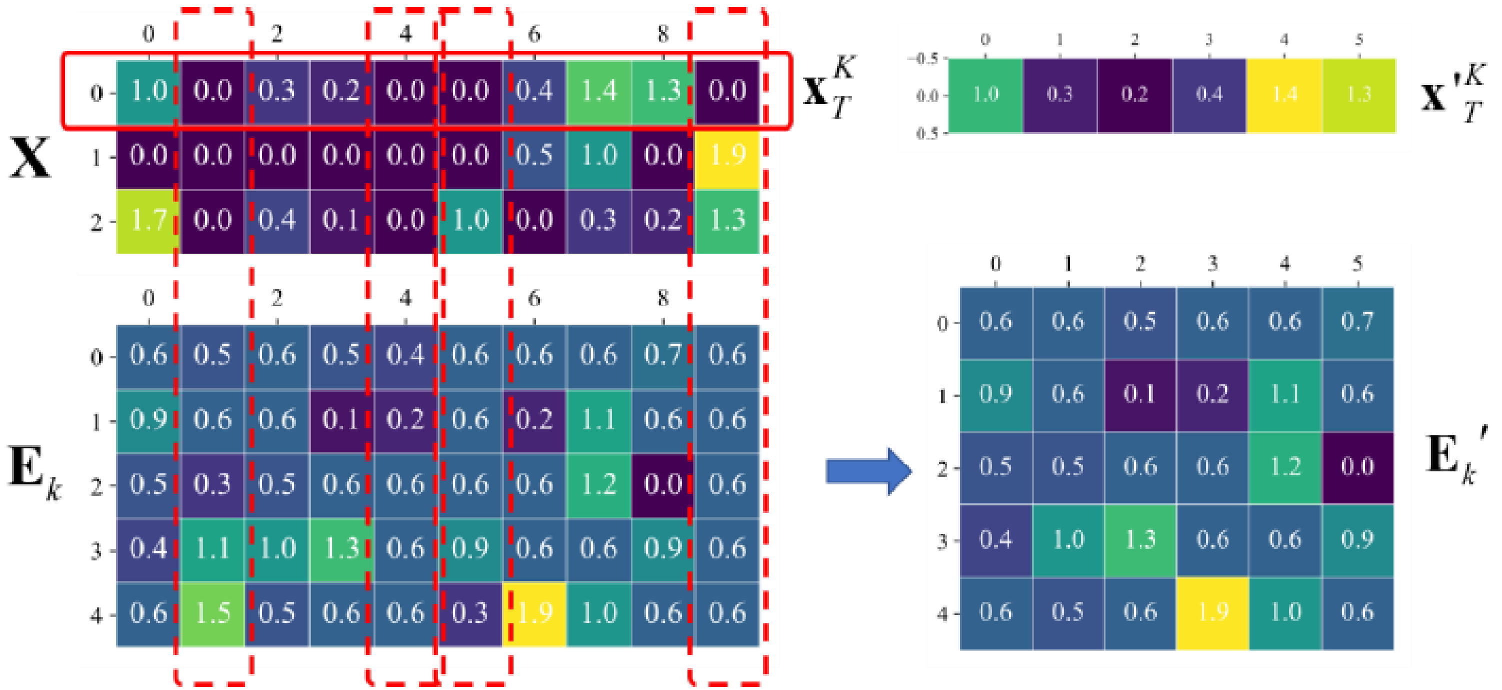
令 $Z=USV^T$ 是 Z 的奇异值分解，U, V 是正交阵，S 是对角阵。式(8)写为 $(Q(\mu)VS^{-1})(S^{-1}V^T Q^T(\mu))=I$，即下列矩阵是正定的：

$$W(\mu)=Q(\mu)VS^{-1} \tag{9}$$

注意到 $Q(\mu)=W(\mu)SV^T$ 一定对称，所以 $W(\mu)=V\Lambda$，其中 Λ 是对角线元素为 ± 1 的对角阵。$D=U\Lambda V^T$ 有 2^N 个候选解，但能使 L 最小化的唯一解为 $\Lambda=I$，此时 $D=UV^T$。即使 Z 的 SVD 分解不唯一，但 UV^T 一定是唯一的。

2.3 K-SVD

K-SVD 目标函数仍为(4)，目标是使 E_k 尽可能接近 $D_k X_k$。由于初始化 D 是超完备字典，必定存在冗余，首先去除 E_k 中对应 X_k 为 0 项对应的部分，简化计算，同时优化后的系数 X 中为 0 的部分仍是 0，保证稀疏性。



简化后的 X_k 记作 X_k^R，相应地 E_k 记作 E_k^R。此时目标函数(4)已经转化为

$$\min \ \| E_k^R - D_k X_k^R \|_F^2 \tag{10}$$

$D_k X_k^R$ 的秩为 1，直接分解 $E_k^R = U\Delta V^T$，找 SVD 分解后秩为 1 且最接近 $D_k X_k^R$ 的矩阵。由于矩阵经过奇异值分解后，奇异值的重要顺序是从上到下排列的，即靠前的都是比较重要的项，甚至有的矩阵的奇异值第一个远大于剩余的奇异值，则

K-SVD 步骤如下：

1. 在已知稀疏矩阵 X 以及数据样本 $\{y_i\}_{i=1}^{N}$ 的情况下，给出要解决的问题：

$$\min_{D}\{\|Y-DX\|_F^2\} \quad s.t \quad \forall i, \|x_i\|_0 \leqslant T_0$$

2. 给出初始字典 $D^{(0)} \in R^{n\times K}$，其中的列向量都是 l^2 范数下的标准形式。给定 $J = 1$ 。

3. 对 $D^{(J-1)}$ 中的每列 $k = 1,2,....K$ 进行迭代：

(1) 通过 $E_k = Y - \sum_{j\neq k} d_j x_T^j$ 来计算误差；

(2) 由 E_k 通过 ω_k 的限定得到 E_R^k ；

(3) 对 E_R^k 进行奇异值分解得到 $E_R^k = U\Delta V^T$，令 $\tilde{d}_k$ 是 U 的第一列，则 $\tilde{d}_k$ 是 d_k 更新的结果。同时，用 V 的第一列和 $\Delta(1,1)$ 的乘积对 x_R^k 进行更新。

最后，$J = J + 1$ 继续重复迭代过程，直到满足停止条件。

3. 参考文献

K-means & K-SVD原理

最优方向法（MOD）

K-SVD：一种用于稀疏表示的过完备字典设计算法 ||论文翻译&解读&代码实现

K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation

Learning Unions of Orthonormal Bases with Thresholded Singular Value Decomposition

机器学习中的数学(5)-强大的矩阵奇异值分解(SVD)及其应用

字典学习（Dictionary Learning, KSVD）详解

周志华-机器学习