

# 特征值和特征向量

## 特征值和SVD分解

对于一个线性变换 $\underline{A}$ ,  $\underline{A}\alpha = \lambda\alpha$ , 则称 $\alpha$ 为 $\underline{A}$ 的特征向量,  $\lambda$ 为 $\underline{A}$ 的特征值。 $\underline{A}$ 在给定基下的矩阵为 $A_n$ 。而对于一般的形如 $A_{m \times n}$ 的矩阵, 没有上述特征值和特征向量。

而且退一步, 对于域 $K$ 上的线性空间 $V$ , 对于一个 $V$ 上线性变换 $\underline{A}$ , 我们并不是总是能够找到一个(正交)基使得 $\underline{A}$ 在这组基下的矩阵能成为对角阵。(检查smith标准型的最后一项)。

扩展上面两个问题, 对于一般的一个线性空间 $V^m$ 到线性空间 $V^n$ 的线性变换 $\underline{A}$ :

$$\underline{A} : \underline{A}\alpha = \beta, \alpha \in V^n, \beta \in V^m$$

能否在 $V^m, V^n$ 中选择合适的正交基使得 $\underline{A}$ 的矩阵表示尽量简单呢?

取 $V^n$ 的两组正交基 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}, \{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n\}$

取 $V^m$ 的两组正交基 $\{\beta_1, \beta_2, \dots, \beta_m\}, \{\xi_1, \xi_2, \dots, \xi_m\}$

$$\text{使得 } \underline{A}(\alpha_1, \alpha_2, \dots, \alpha_n) = (\beta_1, \beta_2, \dots, \beta_m)A$$

此时 $\underline{A}\alpha_1, \underline{A}\alpha_2, \dots, \underline{A}\alpha_n$ 不一定正交, 那么能不能适当调整 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ , 使得变换后的结果可以扩张成 $V^m$ 的一组正交基呢?

$$\text{设 } (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n) = (\alpha_1, \alpha_2, \dots, \alpha_n)P, (\xi_1, \xi_2, \dots, \xi_m) = (\beta_1, \beta_2, \dots, \beta_m)Q$$

$$P^T P = I, Q^T Q = I$$

则对 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 的调整可以看作乘 $P$ 的结果。则

$$\underline{A}(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n) = \underline{A}(\alpha_1, \alpha_2, \dots, \alpha_n)P = (\beta_1, \beta_2, \dots, \beta_m)AP = (\xi_1, \xi_2, \dots, \xi_m)Q^T AP$$

$$\text{令 } A' = Q^T AP = (w_1, w_2, \dots, w_n)$$

如果 $\underline{A}\varepsilon_1, \underline{A}\varepsilon_2, \dots, \underline{A}\varepsilon_n$ 是彼此正交(含0)的, 则只需要 $w_1, w_2, \dots, w_n$ 彼此正交, 如果要求变换后的 $\underline{A}\varepsilon_1, \underline{A}\varepsilon_2, \dots, \underline{A}\varepsilon_n$ 刚好和 $\xi_1, \xi_2, \dots, \xi_m$ 中的元素共线(也就是从一组正交基到另一组正交基), 那么要求 $A'$ 只能有伸缩变换, 即 $A'$ 的所有非对角线元素为0。

那么问题来了: 是否存在 $Q, P$ 使得 $A'$ 存在?

对任给矩阵 $A_{m \times n}$ , 存在正交矩阵 $P_n, Q_m$ 使得 $A' = \begin{pmatrix} A_1 & 0 \\ 0 & 0 \end{pmatrix}$  其中 $A_1 = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_r\}, \lambda_i^2$ 是 $A^T A$ 的所有非零特征值。

大致想法: 设 $r(A_{m \times n}) = r$ ,  $Ax = 0$ 的解空间维度为 $n - r$ , 我们取解空间的正交基并扩充为 $V^n$ 的正交基, 记为 $\Omega_1$ , 同理在 $V^m$ 中构造 $\Omega_2$ , 则 $\Omega_2 A \Omega_1 = \begin{pmatrix} B & 0 \\ 0 & 0 \end{pmatrix}$ , 其中 $B$ 可逆。

$B^T B$ 一定是对称正定矩阵，一定存在 $S$ 为正定对称，使得 $S^2 = B^T B$ (半正定对称分解),正定对称矩阵一定可以正交相似与特征值对角阵，组合上面的结论，那么问题就完结了。

综上：

$$\text{任意矩阵 } A_{m \times n}, \text{ 存在矩阵 } Q_m, P_n \text{ 使得 } A_{m \times n} = Q_m^T \begin{pmatrix} \lambda_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & 0 & \cdots & \lambda_r & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix} P_n$$

$$\begin{pmatrix} \lambda_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & 0 & \cdots & \lambda_r & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix} = Q_m A P_n^T$$

$$\text{记 } Q_m = (u_1, u_2, \dots, u_m), P_n = (v_1, v_2, \dots, v_n)$$

$$A = \lambda_1 u_1 v_1^T + \lambda_2 u_2 v_2^T + \dots + \lambda_r u_r v_r^T$$

这实际构造出了一种逼近矩阵 $A$ 的序列。其中 $\lambda_i$ 是变换前后 $A \varepsilon_i, \varepsilon_i$ 的模之比。

## PCA

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix}$$

原理：一次观测可是看作是在一个  $m$  维的空间选取一个对应点,则矩阵 $X = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n\}$ ,现在想降维这个数据样本。即希望在少于  $m$  的维度近可能好的描述(样本信息损失最少) 原来的样本信息。这就是一个要在低维空间里最佳逼近高维空间信息的过程,有点类似于最小二乘法。具体的想法：我们记原本的空间为  $V$ , 我们需要构造的子空间为  $W$ , 假设在  $W$  内的样本点记为  $\{\vec{x}_{i_1}, \vec{x}_{i_2}, \vec{x}_{i_3}, \dots, \vec{x}_{i_r}\}$ , 不在  $W$  中的点记为  $\{\vec{x}_{o_1}, \vec{x}_{o_2}, \vec{x}_{o_3}, \dots, \vec{x}_{o_{n-r}}\}$ 。当然在  $W$  内的点完全没有任何信息被损失掉,而在  $W$  外的则不然,为了尽量挽救这些损失的信息,一个很正常的想法就是,在  $W$  内建立这些在  $W$  外的样本点的最佳逼近来尽量弥补损失--即  $\{\vec{x}_{o_1}, \vec{x}_{o_2}, \vec{x}_{o_3}, \dots, \vec{x}_{o_{n-r}}\}$  在  $W$  内投影代替原观测。一个很直接的问题就是怎样评估这个子空间选择的好与坏呢? 我们参照最小二乘法引入一个度量函数  $S = S(W) = \sum_{i=1}^{n-r} \|\vec{x}_{o_i} - Proj_W(\vec{x}_{o_i})\|$ , 令  $S = S_{min}$  的  $W$  为最优子空间。如果我们从一维开始构造这个  $W$ , 取  $W$  的基为  $\vec{u}_1$  为单位向量,  $S(W) = S(u_1) = \sum_{i=1}^n \|\vec{x}_i \times \vec{u}_1\|$

$$\min(\sum_{i=1}^n \|\vec{x}_i \times u_1\|) \Leftrightarrow \max(\sum_{i=1}^n \|\vec{x}_i \cdot \vec{u}_1\|)$$

$$S' = \sum_{i=1}^n \|\vec{x}_i \cdot \vec{u}_1\| = \sum_{i=1}^n \vec{x}_i \vec{u}_1^T \vec{x}_i \vec{u}_1^T = \sum_{i=1}^n \vec{u}_1^T \vec{x}_i^T \vec{x}_i \vec{u}_1^T = \vec{u}_1^T \sum_{i=1}^n (\vec{x}_i^T \vec{x}_i) \vec{u}_1^T = \vec{u}_1^T X X^T \vec{u}_1$$

$$S' = \overrightarrow{u_1^T} X X^T \overrightarrow{u_1} - \lambda I \overrightarrow{u_1^T} \overrightarrow{u_1} = \overrightarrow{u_1^T} (X X^T - \lambda I) \overrightarrow{u_1}$$

$$\frac{\partial S'}{\partial \overrightarrow{u_1}} = 2(X X^T - \lambda I) \overrightarrow{u_1} = 0 \Rightarrow X X^T \overrightarrow{u_1} = \lambda \overrightarrow{u_1}$$

$$\frac{\partial^2 S'}{\partial^2 \overrightarrow{u_1}} = 2(X X^T - \lambda I)$$

$$S'_{max} = \max\{\lambda_1, \lambda_2, \dots, \lambda_m\}$$

解为  $X X^T$  的最大特征根对应的特征向量 ( $X X^T - \lambda I$ , 特征值均小于等于 0, 半负定)。

当  $W$  为  $m-1$  维的时候, 取正交单位基  $u_1, u_2, \dots, u_{m-1}$

$$S(W) = \sum_{i=1}^n \|x_i \cdot (u_1 \times u_2 \times \dots \times u_{m-1})\|$$

$$w = u_1 \times u_2 \times \dots \times u_{m-1}$$

$$S(W) = \sum_{i=1}^n \|x_i \cdot w\| = \lambda$$

同上  $w$  一定是  $X X^T$  的特征向量, 使得  $S(W)$  取得最小值,  $w$  为  $\lambda_m$  对应的特征向量。这时候  $u_1, u_2, \dots, u_{m-1}$  与  $\lambda_1, \lambda_2, \dots, \lambda_{m-1}$  对应特征向量组等价。

当  $W$  为  $m-2$  维的时候, 取正交单位基  $u_1, u_2, \dots, u_{m-2}$ , 扩充  $w_1, w_2$  为正交基

$$w_0 = u_1 \times u_2 \times \dots \times u_{m-2}$$

$$\text{取 } w_0 \times w_1 = w_2, w_2 \times w_0 = w_1$$

$$S(W) = \sum_{i=1}^n (\|x_i \cdot w_1\| + \|x_i \cdot w_2\|)$$

显然舍弃  $\lambda_{m-1}, \lambda_m$  对应的信息。

这样我们能逐渐收缩到一维情况。

在这一过程中, 我们也可以直接利用正交矩阵的特征向量彼此垂直, 即以单位特征向量构成基, 依次生成对应维度的逼近子空间。

在  $SVD$  的观点下:  $X X^T$  (可对角化的) 这种线性变换本身非常好, 它存在一组基, 使得在这种基在变换后, 方向不发生改变 (没有旋转和镜面对称), 只是在基的原方向上存在拉伸变化。此时分解的对角阵为特征值对角阵,  $P, Q$  互为转置,  $P$  为特征向量列矩阵。这时候

$\lambda_i v_i v_i^T v_i = \lambda_i v_i$  且只有这么一个特征向量。所以  $\lambda_i v_i v_i^T$  是刻画  $A$  矩阵  $v_i$  特征向量的分量, 而且  $\lambda_i$  的大小反映了这种分量特征在总体中的变化的剧烈程度! (例如方差大小)

所以进一步可以将  $A_{m \times n}$  分解为  $A_{m \times n} = V_{m \times m} Z_{m \times n}$

用  $V$  的列向量反映空间场特征,  $Z$  的对应行向量反映这种场特征在时间上的变化。

## lapack\_blas(EDV,SDV)

Type of	Function and storage scheme	Single precision	Double precision
---------	-----------------------------	------------------	------------------

Type of	Function and storage scheme	Single precision	Double precision		
problem		real	complex	real	complex
SEP	simple driver	SSYEV	CHEEV	DSYEV	ZHEEV
	divide and conquer driver	SSYEVD	CHEEVD	DSYEVD	ZHEEVD
	expert driver	SSYEVX	CHEEVX	DSYEVX	ZHEEVX
	RRR driver	SSYEV	CHEEV	DSYEV	ZHEEV
	simple driver (packed storage)	SSPEV	CHPEV	DSPEV	ZHPEV
	divide and conquer driver	SSPEVD	CHPEVD	DSPEVD	ZHPEVD
	(packed storage)				
	expert driver (packed storage)	SSPEVX	CHPEVX	DSPEVX	ZHPEVX
	simple driver (band matrix)	SSBEV	CHBEV	DSBEV	ZHBEV
	divide and conquer driver	SSBEVD	CHBEVD	DSBEVD	ZHBEVD
	(band matrix)				
	expert driver (band matrix)	SSBEVX	CHBEVX	DSBEVX	ZHBEVX
	simple driver (tridiagonal matrix)	SSTEVD		DSTEVD	
	divide and conquer driver	SSTEVD		DSTEVD	
	(tridiagonal matrix)				
	expert driver (tridiagonal matrix)	SSTEVD		DSTEVD	

Type of	Function and storage scheme	Single precision	Double precision		
	RRR driver (tridiagonal matrix)	SSTEVR		DSTEVR	
NEP	simple driver for Schur factorization	SGEES	CGEES	DGEES	ZGEES
	expert driver for Schur factorization	SGEESX	CGEESX	DGEESX	ZGEESX
	simple driver for eigenvalues/vectors	SGEEV	CGEEV	DGEEV	ZGEEV
	expert driver for eigenvalues/vectors	SGEEVX	CGEEVX	DGEEVX	ZGEEVX