奇异值分解

Singular Value Decomposition

内容提要

问题引入

问题引入

- 方阵：特征分解。
  一个 n 阶可对角化方阵 A 可以分解为

$$A = P\Lambda P^{-1},$$

其中 $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ 是一个对角阵，对角元 λ_i 为 A 的特征值，P 由相应的特征向量 $\mathbf{v}_i$ 组成。

问题引入

问题引入

- 方阵：特征分解。
  一个 n 阶可对角化方阵 A 可以分解为

$$A = P\Lambda P^{-1},$$

  其中 $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ 是一个对角阵，对角元 λ_i 为 A 的特征值，P 由相应的特征向量 $\mathbf{v}_i$ 组成。

- 问题：对于一个一般的 $m \times n$ 的实矩阵，如何将其“对角化”?

奇异值分解定理

奇异值分解定理，Theorem of Singular Value Decomposition

设 M 是一个 $m \times n$ 的实矩阵，则 M 有如下的所谓的奇异值分解 (Singular Value Decomposition, SVD)：

$$M = U\Sigma V^T = U \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \end{pmatrix}_{m\times n} V^T,$$

奇异值分解定理

奇异值分解定理，Theorem of Singular Value Decomposition

设 M 是一个 $m \times n$ 的实矩阵，则 M 有如下的所谓的奇异值分解（Singular Value Decomposition, SVD）：

$$M = U\Sigma V^T = U \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \end{pmatrix}_{m\times n} V^T,$$

- $r = r(M)$，$\sigma_1 \geqslant \cdots \geqslant \sigma_r > 0$ 被称为奇异值（Singular Value）
- U, V 分别为 m 阶，n 阶正交矩阵，他们的前 r 列向量分别被称为左、右奇异向量（Left-,Right-Singular Vector）。

奇异值分解定理证明

关键点

奇异值分解定理证明

关键点

- 特征分解的特殊情况：当 A 是一个实对称方阵的时候，它正交相似于对角阵：

$$A = P\Lambda P^T,$$

其中 P 为正交阵。

奇异值分解定理证明

关键点

- 特征分解的特殊情况：当 A 是一个实对称方阵的时候，它正交相似于对角阵：

$$A = P\Lambda P^T,$$

  其中 P 为正交阵。

- 若 M 为 $m \times n$ 的实矩阵，则 MM^T, M^TM 分别为 m 阶与 n 阶实对称方阵。

奇异值分解定理证明

引理

MM^T 与 M^TM 的特征值非负，且二者的正特征值之集相同。

奇异值分解定理证明

引理

MM^T 与 M^TM 的特征值非负，且二者的正特征值之集相同。

证明

任取 MM^T 的一个特征值 λ 以及相应的一个特征向量 $\mathbf{x}$，即有 $MM^T\mathbf{x} = \lambda\mathbf{x}$。那么 $\mathbf{x}^TMM^T\mathbf{x} = \mathbf{x}^T\lambda\mathbf{x}$，即

$$\|M^T\mathbf{x}\|^2 = \lambda\|\mathbf{x}\|^2. \tag{1}$$

所以 λ 非负。同理 M^TM 的特征值也都是非负实数。

奇异值分解定理证明

引理

MM^T 与 M^TM 的特征值非负，且二者的正特征值之集相同。

证明

任取 MM^T 的一个特征值 λ 以及相应的一个特征向量 $\mathbf{x}$，即有 $MM^T\mathbf{x} = \lambda\mathbf{x}$。那么 $\mathbf{x}^TMM^T\mathbf{x} = \mathbf{x}^T\lambda\mathbf{x}$，即

$$\|M^T\mathbf{x}\|^2 = \lambda\|\mathbf{x}\|^2. \tag{1}$$

所以 λ 非负。同理 M^TM 的特征值也都是非负实数。
设 $\lambda > 0$, 还是由 $MM^T\mathbf{x} = \lambda\mathbf{x}$，有

$$M^TMM^T\mathbf{x} = \lambda M^T\mathbf{x}, \tag{2}$$

又由(1)式知 $M^T\mathbf{x}$ 非零向量，故 λ 也是 M^TM 的正特征值。 $\square$

奇异值分解定理证明

奇异值分解定理的证明

把 MM^T 与 M^TM 的正特征值记为 $\sigma_1^2 \geqslant \cdots \geqslant \sigma_r^2 > 0$, 其中 $\sigma_i > 0$. 设 M^TM 通过 n 阶正交阵 $V = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ 对角化:

$$M^TM = V \begin{pmatrix} \sigma_1^2 & & & \\ & \ddots & & \\ & & \sigma_r^2 & \\ & & & 0 \end{pmatrix}_n V^T.$$

有 $M^TM\mathbf{v}_i = \sigma_i^2\mathbf{v}_i$, 令 $\mathbf{u}_i = \dfrac{M\mathbf{v}_i}{\sigma_i}$, 那么可以证明:

奇异值分解定理证明

奇异值分解定理的证明

把 MM^T 与 M^TM 的正特征值记为 $\sigma_1^2 \geqslant \cdots \geqslant \sigma_r^2 > 0$, 其中 $\sigma_i > 0$. 设 M^TM 通过 n 阶正交阵 $V = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ 对角化:

$$M^TM = V \begin{pmatrix} \sigma_1^2 & & & \\ & \ddots & & \\ & & \sigma_r^2 & \\ & & & 0 \end{pmatrix}_n V^T.$$

有 $M^TM\mathbf{v}_i = \sigma_i^2 \mathbf{v}_i$, 令 $\mathbf{u}_i = \dfrac{M\mathbf{v}_i}{\sigma_i}$, 那么可以证明:

1. $\mathbf{u}_1, \ldots, \mathbf{u}_r$ 是 m 阶实对称阵 MM^T 的单位正交的特征向量;
2. $M\mathbf{v}_i = \sigma_i \mathbf{u}_i$, $M^T\mathbf{u}_i = \sigma_i \mathbf{v}_i$.

SVD

问题引入
奇异值分解定理
定理证明
例子与应用

奇异值分解定理的证明

于是有

$$M(\mathbf{v}_1, \ldots, \mathbf{v}_r) = (\mathbf{u}_1, \ldots, \mathbf{u}_r) \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix}.$$

SVD

问题引入

奇异值分解定理

定理证明

例子与应用

奇异值分解定理的证明

于是有

$$M(\mathbf{v}_1,\ldots,\mathbf{v}_r)=(\mathbf{u}_1,\ldots,\mathbf{u}_r)\begin{pmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_r\end{pmatrix}.$$

把 $\{\mathbf{u}_1,\ldots,\mathbf{u}_r\}$ 扩充为 $\mathbb{R}^m$ 的一组单位正交基 $\{\mathbf{u}_1,\ldots,\mathbf{u}_r,\mathbf{u}_{r+1},\ldots,\mathbf{u}_m\}$，那么我们就得到了 SVD 定理中的形式:

$$M=(\mathbf{u}_1,\ldots,\mathbf{u}_m)\begin{pmatrix}\sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0\end{pmatrix}_{m\times n}(\mathbf{v}_1,\ldots,\mathbf{v}_n)^T.$$

奇异值分解定理

奇异值分解的一种更紧凑的形式

将 $M = (\mathbf{u}_1, \ldots, \mathbf{u}_m) \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \end{pmatrix}_{m\times n} (\mathbf{v}_1, \ldots, \mathbf{v}_n)^T$ 写成分块形式

$$M = (U_1, U_2)\begin{pmatrix} \Sigma_0 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = U_1\Sigma_0 V_1,$$

其中

$$U_1 = (\mathbf{u}_1, \ldots, \mathbf{u}_r), \quad U_2 = (\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m);$$
$$V_1 = (\mathbf{v}_1, \ldots, \mathbf{v}_r)^T, \quad V_2 = (\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n)^T;$$
$$\Sigma_0 = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix}$$

奇异值分解的例子

计算矩阵 $M = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ 的奇异值分解。

奇异值分解例子

奇异值分解的例子

计算矩阵 $M = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ 的奇异值分解。

解：有

$$M^T M = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

于是有奇异值 $\sigma_1 = \sigma_2 = 1$. 我们可以选取

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

从而有

$$\mathbf{u}_1 = M\mathbf{v}_1 = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \quad \mathbf{u}_2 = M\mathbf{v}_2 = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$$

奇异值分解的例子（续）

于是 $M = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ 的奇异值分解可以是

$$M = U\Sigma V^T = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \cdot I_2 \cdot I_2.$$

奇异值分解例子

奇异值分解的例子（续）

于是 $M = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ 的奇异值分解可以是

$$M = U\Sigma V^T = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \cdot I_2 \cdot I_2.$$

我们也可以另选

$$\mathbf{v}_1 = \begin{pmatrix} \cos\alpha \\ \sin\alpha \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} -\sin\alpha \\ \cos\alpha \end{pmatrix},$$

于是，相应的奇异值分解为

$$M = U\Sigma V^T = \begin{pmatrix} \cos(\theta+\alpha) & -\sin(\theta+\alpha) \\ \sin(\theta+\alpha) & \cos(\theta+\alpha) \end{pmatrix} \cdot I_2 \cdot \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}$$

奇异值分解应用

奇异值分解的应用：图像压缩

利用 SVD，我们可以把一个秩为 r 的 $m \times n$ 矩阵 M 写作

$$M = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T .$$

考虑 M 是一幅 $m \times n$ 个像素的图片的情况。尽管图片一般接近满秩，但有效秩很低。也就是说，存在一个相对于 r 很小的 k，使得 $\sigma_{k+1}, \ldots, \sigma_r$ 非常小，接近于 0，因此用

$$M_k := \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T$$

可以很好地近似 M，同时只需要存储 $k(m+n+1)$ 个数据，而不是原始的 mn 个数据。于是所需存储的数据大大减少，图像得到了压缩。

奇异值分解的应用

SVD 的应用还包括

- 计算 Moore-Penrose 伪逆，进而求解最小二乘法问题;
- 数据集的主成分分析（Principal Components Analysis, PCA）;
- PageRank
- Eigenface
- ……

奇异值分解的应用

SVD 的应用还包括

- 计算 Moore-Penrose 伪逆，进而求解最小二乘法问题;
- 数据集的主成分分析（Principal Components Analysis, PCA）;
- PageRank
- Eigenface
- ……

还需要强调的是，计算机中 SVD 的实现，并不是按我们证明 SVD 定理中的步骤来的，而是用另外的快速的算法。（类似的还有线性代数中其他很多的计算）
这里就不做讨论了。