线性代数进阶

邓明格

七月在线

mingge_deng@brown.edu

March 31, 2017

主要内容

对称方阵

对称方阵

- 定义：$\mathbf{H} \in \mathcal{R}^{n,n}$， $h_{ij} = h_{ji}$ or $\mathbf{H}^T = \mathbf{H}$。
- 协方差矩阵：$\mathbf{H} = (\mathbf{X} - \mathbf{X_0})(\mathbf{X} - \mathbf{X_0})^\mathbf{T}$，$\mathbf{X}$ 为列向量或者 $n \times n$ 方阵。
- 海森矩阵 (牛顿法):

$$\mathbf{H}\left(f(\beta)\right) = \begin{bmatrix} \frac{\partial^2 f}{\partial \beta_1 \partial \beta_2} & \cdots & \frac{\partial^2 f}{\partial \beta_1 \partial \beta_n} \\ \ldots & \ldots & \ldots \\ \frac{\partial^2 f}{\partial \beta_n \partial \beta_1} & \cdots & \frac{\partial^2 f}{\partial \beta_n \partial \beta_n} \end{bmatrix} \tag{1}$$

- 内积矩阵

正定方阵

正定方阵

- 定义：$\mathbf{H} \in \mathcal{R}^{n,n}$ 为对称方阵，$\forall \mathbf{x} \neq \mathbf{0}$，满足 $\mathbf{x}^T\mathbf{H}\mathbf{x} > 0$，则 $\mathbf{H}$ 为正定方阵，< 0 为负定，≥ 0 为半正定方阵。
- 协方差矩阵
- 海森矩阵（如果 $f(\beta)$ 为凸函数）
- 内积矩阵

其他特殊矩阵

- 正交矩阵 $\mathbf{H} \in \mathcal{R}^{n,n}, \mathbf{H}^T\mathbf{H} = \mathbf{H}\mathbf{H}^T = \mathbf{I}$。
- 三角矩阵
- 对角矩阵
- 稀疏矩阵

相似变换

- 定义：如果方阵 $\mathbf{A}$ 与 $\tilde{\mathbf{A}}$ 满足 $\tilde{\mathbf{A}} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$，那么方阵 $\mathbf{A}$ 与 $\tilde{\mathbf{A}}$ 互为相似矩阵。
- 几何意义：同一个线性变换在不同的基下的表达形式。

- 线性空间 $\mathbf{V}$ 可以由基 α 或者基 $\tilde{\alpha}$ 来描述。两组基满足 $\tilde{\alpha} = \alpha \cdot \mathbf{P}$，那么同一个向量在不同基下的坐标满足 $\tilde{\mathbf{x}} = \mathbf{P}^{-1}\mathbf{x}$。
- 线性变换 $\mathbf{T} : \mathbf{V} \to \mathbf{V}$:
  - 线性空间 $\mathbf{V}$ 由基 α 描述，线性变换可以表示为 $\mathbf{x} = \mathbf{A}\mathbf{y}$
  - 线性空间 $\mathbf{V}$ 由基 $\tilde{\alpha}$ 描述，线性变换可以表示为 $\tilde{\mathbf{x}} = \tilde{\mathbf{A}}\tilde{\mathbf{y}}$

$$\mathbf{P}^{-1}\mathbf{x} = \tilde{\mathbf{A}}\mathbf{P}^{-1}\mathbf{y} \Longrightarrow \mathbf{x} = \mathbf{P}\tilde{\mathbf{A}}\mathbf{P}^{-1}\mathbf{y} \Longrightarrow \mathbf{A} = \mathbf{P}\tilde{\mathbf{A}}\mathbf{P}^{-1} \tag{2}$$

相似变换实例

线性变换 $\mathbf{T}:\mathcal{R}^{\mathbf{2}} \to \mathcal{R}^{\mathbf{2}}$，沿 $y = x$ 方向拉伸 2 倍

- 选择基 $\alpha_1 = (1,0), \alpha_2 = (0,1)$ 描述，$\mathbf{A} = \begin{bmatrix} 3/2 & 1/2 \\ 1/2 & 3/2 \end{bmatrix}$
- 选择基 $\alpha_1 = (1,1), \alpha_2 = (1,\ 1)$ 描述，$\tilde{\mathbf{A}} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$

相似变换不变量：行列式

$\mathbf{A}$ 与 $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ 描述同一个线性变换，体积变化率不变

$$\begin{aligned} \det(\mathbf{P}^{-1}\mathbf{A}\mathbf{P}) &= \det(\mathbf{P}^{-1})\det(\mathbf{A})\det(\mathbf{P}) \\ &= \det(\mathbf{P}^{-1})\det(\mathbf{P})\det(\mathbf{A}) \\ &= \det(\mathbf{A}) \end{aligned} \tag{3}$$

提示：$\det(\mathbf{A}\mathbf{B}) = \det(\mathbf{B}\mathbf{A}) = \det(\mathbf{A})\det(\mathbf{B})$

相似变换不变量：迹

迹

- 定义：迹是矩阵所有对角元的和：$tr(\mathbf{A}) = \sum_i a_{ii}$
- 迹是矩阵特征值的总和。（作业，提示：特征多项式韦达定理）
- $tr(\mathbf{AB}) = tr(\mathbf{BA})$。（作业，提示：定义）
- 几何意义：$\exp(tr(\mathbf{A}))$ 是线性变换 $\exp(\mathbf{A})$ 的体积变化率).

$$\exp(tr(\mathbf{A})) = \det(\exp(\mathbf{A})) \tag{4}$$

$$e^{\mathbf{A}} = \mathbf{I} + \mathbf{A} + \mathbf{A}^2/2 + \ldots \tag{5}$$

$\exp(\mathbf{A})$ 与 $\exp(\mathbf{P}^{-1}\mathbf{AP})$ 描述同一个线性变换，体积变化率不变

$$tr(\mathbf{P}^{-1}\mathbf{AP}) = tr(\mathbf{APP}^{-1}) = tr(\mathbf{A} \cdot \mathbf{I}) = tr(\mathbf{A}) \tag{6}$$

相似变换不变量：特征值

特征值

特征值是特征方程 $\det(\mathbf{A}-\lambda\mathbf{I})=0$ 的根，如果 $\det(\mathbf{A}-\lambda\mathbf{I})=0$，那么 $\det(\mathbf{P}^{-1}(\mathbf{A}-\lambda\mathbf{I})\mathbf{P})=0$，于是 $\det(\mathbf{P}^{-1}\mathbf{A}\mathbf{P}-\lambda\mathbf{I})=0$

注意：特征值是最重要的相似不变量，利用这个相似不变量可以得出其他的不变量。

内积

内积

$\mathbf{V}$ 是线性空间，内积用来表达线性空间的空间度量，是正定、对称以及双线性形式, : $\mathbf{V} \times \mathbf{V} \rightarrow \mathcal{R}$

- $G(x,x) \geq 0 \quad G(x,x) = 0 \quad iff \quad x = 0,$
- $G(x,y) = G(y,x)$
- $G(x+y,z) = G(x,z) + G(y,z)$
- 欧氏空间标准内积 $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x^T y} = \sum_{\mathbf{k=1}}^{\mathbf{n}} \mathbf{x_k y_k}$

如果 $\alpha\{\alpha_1, ..., \alpha_k\}$ 是空间 $\mathbf{V}$ 的一组基，内积一般可以用一个对称矩阵 H_α 来表示

$$h_{ij} = G(\alpha_i, \alpha_j) = G(\alpha_j, \alpha_i) = h_{ji} \tag{7}$$

任意两个向量 $\mathbf{v} = \alpha \cdot x$，$\mathbf{w} = \alpha \cdot y$，其中 x, y 为基 α 下的坐标，那么

$$G(\mathbf{v}, \mathbf{w}) = x^T H_\alpha y \tag{8}$$

相合变换

相合变换

- 如果两个对称方阵 $\mathbf{A}$ 与 $\tilde{\mathbf{A}}$ 满足 $\tilde{\mathbf{A}} = \mathbf{P}^T\mathbf{A}\mathbf{P}$. 那么这两个方阵就互为相合矩阵。
- 相似矩阵的几何意义是同一个内积结构在不同基下的表示形式。

线性空间 $\mathbf{V}$ 可以由基 α 或者基 $\tilde{\alpha}$ 来描述。两组基满足 $\alpha = \tilde{\alpha} \cdot \mathbf{P}$ ，那么任意一向量 $\mathbf{v} = \alpha \cdot x = \tilde{\alpha}\mathbf{P} \cdot x = \tilde{\alpha} \cdot \tilde{x}$。

$$\begin{aligned} G(\mathbf{v}, \mathbf{w}) = x^T H_\alpha y = \tilde{x}^T \tilde{H}_\alpha \tilde{y} = x^T \mathbf{P}^T \tilde{H}_\alpha \mathbf{P} y \\ \Longrightarrow \mathbf{H} = \mathbf{P}^T \tilde{\mathbf{H}} \mathbf{P} \end{aligned} \tag{9}$$

相合变换不变量

相合变换不变量

- 矩阵的正定性（正定，负定）
- 矩阵的正负特征值的个数（Signature）
- 相合变换下矩阵保持对称性

正交相似变换

- 正交相似变换同时满足相似与相合变换的条件，也就是说它同时保持了矩阵的相似与相合不变量。
- 如果两个对称方阵方阵 $\mathbf{A}$ 与 $\tilde{\mathbf{A}}$ 满足 $\tilde{\mathbf{A}} = \mathbf{P}^T\mathbf{A}\mathbf{P}$，且 $\mathbf{P}$ 是正交矩阵 $\mathbf{P}^T = \mathbf{P}^{-1}$. 那么 $\mathbf{A}$ 与 $\tilde{\mathbf{A}}$ 就互为正交相似.

任何一个对称矩阵 $\mathbf{A}$ 都可以正交相似于一个对角矩阵 $\mathbf{D}$，或者说总存在一个正交矩阵 $\mathbf{P}$，使得 $\mathbf{A} = \mathbf{P}^T\mathbf{D}\mathbf{P}$, 且对角矩阵 $\mathbf{D}$ 由 $\mathbf{A}$ 的特征值构成。

正交相似实例及几何意义

对称方阵 $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, 求及正交相似标准型:

- 求特征值 $\det(\mathbf{A} - \lambda \mathbf{I}) = 0$, 得到 $\lambda_1 = 1, \lambda_2 = 3$。
- 求特征向量 $\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$, 得到 $\mathbf{v}_1^T = (1/\sqrt{2}, 1/\sqrt{2})$, $\mathbf{v}_2^T = (-1/\sqrt{2}, 1/\sqrt{2})$。
- $\mathbf{A} = \mathbf{P}^T\mathbf{D}\mathbf{P}$, $\mathbf{D} = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$, $\mathbf{P} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$

几何意义：方阵 $\mathbf{A}$ 对应的线性变换：在 $\mathbf{v}_1$ 方向放大三倍，$\mathbf{v}_2$ 方向保持不变。

LU 分解与 cholesky 分解

LU 分解

- 定义：$\mathbf{A} \in \mathcal{R}^{n\times n}$, 存在唯一的单位下三角矩阵 $\mathbf{L}$ 和非奇异上三角矩阵 $\mathbf{U}$, 使得 $\mathbf{A} = \mathbf{LU}$, 充要条件是 $\mathbf{A}$ 的所有顺序主子矩阵 $\mathbf{A}_k = \mathbf{A}(1:k, 1:k)$ 都非奇异。
- 方法：高斯消去法 ($\mathcal{O}(n^3)$)，(如何求解线性方程组)

cholesky 分解

- 定义：$\mathbf{A} \in \mathcal{R}^{n\times n}$ 对称正定, 则存在唯一的对角线元素为正的下三角矩阵 $\mathbf{L}$, 使得 $\mathbf{A} = \mathbf{LL}^T$。
- 联系正交相似标准型：$\mathbf{A} = \mathbf{P}^T\mathbf{DP}$，如果 $\mathbf{A}$ 对称正定, 特征值都为正数，所以 $\mathbf{L} = \mathbf{P}^T\mathbf{D}^{1/2}\mathbf{P}$。

QR 分解

- 定义：$\mathbf{A} \in \mathcal{R}^{n \times n}$，可以被分解为 $\mathbf{A} = \mathbf{QR}$，其中 $\mathbf{Q}$ 是正交矩阵，而 $\mathbf{R}$ 是上三角矩阵。如果 $\mathbf{A}$，分解唯一。
- Gram-Schmidt 正交化.

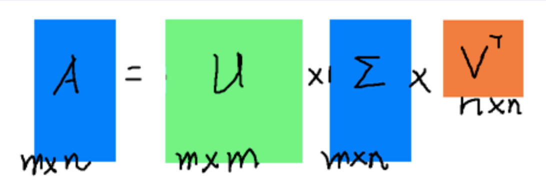
$$
\begin{aligned}
\boldsymbol{\beta}_1 &= \boldsymbol{v}_1, & \boldsymbol{\eta}_1 &= \frac{\boldsymbol{\beta}_1}{\|\boldsymbol{\beta}_1\|} \\
\boldsymbol{\beta}_2 &= \boldsymbol{v}_2 - \langle \boldsymbol{v}_2, \boldsymbol{\eta}_1 \rangle \boldsymbol{\eta}_1, & \boldsymbol{\eta}_2 &= \frac{\boldsymbol{\beta}_2}{\|\boldsymbol{\beta}_2\|} \\
\boldsymbol{\beta}_3 &= \boldsymbol{v}_3 - \langle \boldsymbol{v}_3, \boldsymbol{\eta}_1 \rangle \boldsymbol{\eta}_1 - \langle \boldsymbol{v}_3, \boldsymbol{\eta}_2 \rangle \boldsymbol{\eta}_2, & \boldsymbol{\eta}_3 &= \frac{\boldsymbol{\beta}_3}{\|\boldsymbol{\beta}_3\|} \\
&\vdots & &\vdots \\
\boldsymbol{\beta}_n &= \boldsymbol{v}_n - \sum_{i=1}^{n-1} \langle \boldsymbol{v}_n, \boldsymbol{\eta}_i \rangle \boldsymbol{\eta}_i, & \boldsymbol{\eta}_n &= \frac{\boldsymbol{\beta}_n}{\|\boldsymbol{\beta}_n\|}
\end{aligned}
$$

这样就得到$\mathrm{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$上的一组正交基$\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n\}$，以及相应的标准正交基$\{\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_n\}$

SVD 分解

- 定义：对于任何一个矩阵 $\mathbf{A} \in \mathcal{R}^{m\times n}$, 存在正交矩阵 $\mathbf{U}^{m\times m}, \mathbf{V}^{n\times n}$. 使得 $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$, 其中 $\Sigma^{m\times n}$ 是一个对角矩阵, 且

$$\Sigma = \begin{bmatrix} \tilde{\Sigma} & 0_{r,n-r} \\ 0_{m-r,r} & 0_{m-r,n-r} \end{bmatrix}, \quad \tilde{\Sigma} = diag(\sigma_1, ..., \sigma_r) \tag{10}$$



SVD 分解与特征值的联系：（作业）

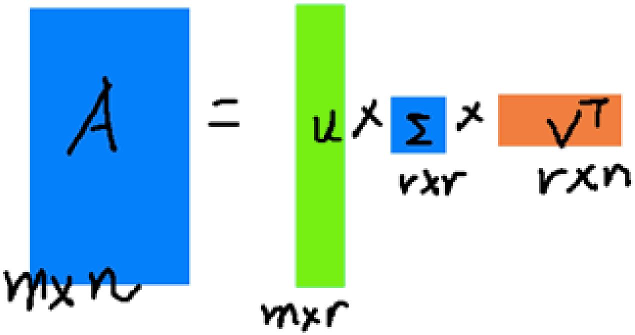
- $\mathbf{U}^{m\times m}$ 为 $\mathbf{AA}^T$ 的特征向量。
- $\mathbf{V}^{n\times n}$ 为 $\mathbf{A}^T\mathbf{A}$ 的特征向量。
- $\tilde{\Sigma}$ 为 $\mathbf{AA}^T$ 或者 $\mathbf{A}^T\mathbf{A}$ 的非零特征值的平方根，称为奇异值。

伪逆的 SVD 表达：(作业)

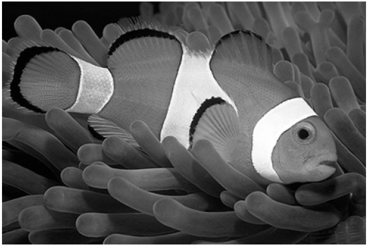
低阶矩阵近似

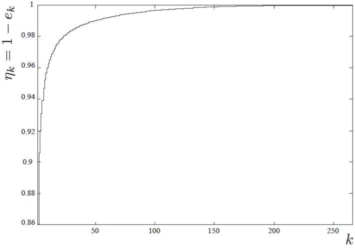
部分奇异值分解: 用少部分大的奇异值来近似描述矩阵

$$\mathbf{A} \approx \mathbf{U}_{m,k}\Sigma_{k,k}\mathbf{V}_{k,n}^{T} \tag{11}$$



低阶矩阵近似实例





k = 9 包含 96% 图片的方差。

对于任何一个矩阵 $\mathbf{A} \in \mathcal{R}^{m\times n}$, 并且 $r = rank(\mathbf{A}) > 0$, 如何用低阶的矩阵近似

$$\begin{aligned} & min_{\mathbf{A}_k \in \mathcal{R}^{m\times n}} ||\mathbf{A} - \mathbf{A}_k||_F^2 \\ & s.t. : rank(\mathbf{A}_k) = k < r \end{aligned} \tag{12}$$

最优解:

$$\begin{aligned} \mathbf{A} &= \mathbf{U}\Sigma\mathbf{V}^T = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T \\ \Longrightarrow \mathbf{A}_k &= \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^T \end{aligned} \tag{13}$$

PCA

PCA: 最常用的线性降维方法，它的目标是通过某种线性投影，将高维的数据映射到低维的空间中表示，并期望在所投影的维度上数据的方差最大。

- 数据点：$\mathbf{x}_i \in \mathcal{R}^n, i = 1, ..., m$，归一化，使中心为零，并且方差为 1。
- 寻找空间中的一个方向 $\mathbf{z} \in \mathcal{R}^n, ||z||_2 = 1$，使得所有数据在 $\mathbf{z}$ 上的投影方差和最大（保持最多信息）。
- 数据点在 $\mathbf{z}$ 方向上的投影为 $\alpha_i = \mathbf{x}_i^T\mathbf{z}$
- 数据投影方差和：$\sum_{i=1}^{m} \alpha_i^2 = \sum_{i=1}^{m} \mathbf{z}^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{z}$

PCA

PCA 问题:

$$\max_{\mathbf{z}\in\mathcal{R}^n} \mathbf{z}^T(\mathbf{x}\mathbf{x}^T)\mathbf{z} \quad s.t. \quad ||z||_2 = 1 \tag{14}$$

由 SVD 求解 PCA 问题（作业）

我们可以得到 $\mathbf{z} = \mathbf{u}_1$, 方差和为 σ_1^2。

推荐问题 PCA 分析

谢谢大家！！

下节课主要内容：

- 微积分回顾
- 牛顿法
- 概率论回顾
- 极大似然估计