CS131 Lecture03： 线性代数初级 Part2

by：斯坦福大学计算机科学系

github: https://github.com/zhaoxiongjun/CS131_notes_zh-CN （包含中英文版课件及相关课程视频）

1 向量和矩阵回顾

1.1 向量

列向量v ∈ $\mathbb{R}^{n*1}$ ：

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

行向量 v^T ∈ $\mathbb{R}^{1*n}$ ：（T 表示转置操作）

$$\mathbf{v}^T = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$$

范式：

$$||x||_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$$

更多关于向量和矩阵范式解释：

https://www.zhihu.com/question/20473040/answer/102907063

1.2 矩阵

一个 m*n 的矩阵$A \in \mathbb{R}^{m*n}$：

如果 m=n，则 A 是一个方阵

$$\mathbf{A}=\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & & & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}$$

1.2.1 矩阵的应用

灰度图像在每个像素有一个数值，被存储在一个 m*n 的矩阵。彩色图像每个像素有 3 个数值——红色，绿色，蓝色（即对应三通道）

2 矩阵的初等变换

矩阵可以通过乘法以有用的方式转换向量。最简单的应用是通过缩放，或者将一个缩放矩阵的对角线上有标量乘以向量。

我们也可以用矩阵来旋转向量。当我们把一个矩阵和一个向量相乘;得到的 x 坐标是原始向量点乘第一行。

为了将一个向量逆时针旋转一个角度θ：

$x' = cos\theta x - sin\theta y$ and

$y' = cos\theta y + sin\theta x$

therefore, we multiply it by the matrix

$$M=\begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix}$$

这相当于：P’ = RP (P 和 P’ 指向量，R 是旋转的变换矩阵 M)

我们同样可以通过矩阵乘法转换一个点。比如，p’ = R2R1Sp。这个转换是从右往左，一层接一层。在我们的例子中，这个等于（R2(R1(Sp)))。

为了翻译向量，我们必须实现一个有点粗糙的解决方案，在向量的最末端添加一个“1”。这样，使用这些“齐次坐标”，我们也可以转换向量。乘法运算的结果是将矩阵最右边的列加到各自的坐标上。均匀矩阵的底行将有[0 0 1]以确保元素被正确添加，并且结果向量的底行有“1”。

按照惯例，在齐次坐标系下，在进行矩阵乘法后，将结果除以最后一个坐标。

$$\begin{bmatrix} x \\ y \\ 7 \end{bmatrix}$$

$$\begin{bmatrix} x/7 \\ y/7 \\ 1 \end{bmatrix}$$

因此，为了得到 P（x；y）—> P’ =（SxX；SyY）的结果，我们必须首先 P=（x；y）—>（x；y；1）然后 P’ =（SxX；SyY）—>（SxX；SyY；1）这样我们才能进行矩阵乘法 S*P。但是，我们必须注意缩放后平移与平后缩放是不同的。换句话说，T*S*P ≠ S*T*P

任何旋转矩阵 R 都属于正规矩阵的范畴，它满足以下性质。例如，$\mathrm{R}R^T = I$ 和 det（R）=1

旋转矩阵的行总是是相互垂直的（a.k.a.正交）单位向量；这使得它能够满足上面提到的一些独特特性。

3 逆矩阵

给一个矩阵 A，它的逆矩阵A^{-1}定义如下：

$$AA^{-1} = A^{-1}A = I$$

I 表示有同样尺寸的单位矩阵

一个逆矩阵例子：

$$\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{bmatrix}$$

矩阵不一定有逆矩阵。如果存在A^{-1}，则称 A 为可逆或非奇异。对于可逆矩阵，一些有用的恒等式是：

- $(A^{-1})^{-1} = A$
- $(AB)^{-1} = B^{-1}A^{-1}$
- $A^{-T} \triangleq (A^T)^{-1} = (A^{-1})^T$

3.1 伪逆矩阵

在线性代数问题中，经常需要解 X 的 AX=B 方程。你需要计算A^{-1}，然后两边相乘得到 X=A^{-1}B。在 python 中，这个命令是：np.linalg.inv（A）*B。然而，对于大型浮点矩阵，计算逆矩阵可能非常昂贵，而且可能不准确。对于 A，甚至都不可能存在一个逆函数。我们应该怎么做?

4 矩阵的秩

- 转换矩阵 A 的秩告诉您它将一个矩阵转换为多少维。
- Col-rank（A）=A 的线性无关列向量的最大数目。
- Row-rank（A）=A 的最大线性无关行向量数。列秩始终等于行秩。
- 对于转换矩阵，秩表示输出的维数。例如，如果 A 的秩为 1，则转换

P’ =AP 指向一条直线。

- 满秩矩阵-如果 m*m,则秩为 m
- 奇异矩阵-如果 m*m 矩阵的秩小于 m，因为至少有一个维度正在收缩。（无法判断结果的输入值）–>对于非方形矩阵，不存在逆矩阵。

5 特征向量（SVD）

5.1 定义

线性变换 A 的特征向量 x 是一个非零向量，当 A 作用于它时，它不会改变它的方向。将 A 作用于特征向量，用一个称为特征值的标量来缩放特征向量。

下面的方程描述了特征值和特征向量之间的关系：

$$A\mathbf{x} = \lambda\mathbf{x}, \quad \mathbf{x} \neq \mathbf{0}$$

5.2 求特征向量和特征值

如果我们要求 A 的特征向量，我们可以用如下方法求解：

$$\begin{gathered} A\mathbf{x} = \lambda\mathbf{x}, \quad \mathbf{x} \neq \mathbf{0} \\ A\mathbf{x} = (\lambda I\mathbf{x}), \quad \mathbf{x} \neq \mathbf{0} \\ (\lambda I - A)\mathbf{x} = \mathbf{0}, \quad \mathbf{x} \neq \mathbf{0} \end{gathered}$$

因为 x 是非零向量，因为我们可以将上式变为：

$$|\lambda I - A| = 0$$

解这个方程给出了 A 的特征值，这些特征值可以代入原方程，找到相应的特征向量。

5.3 性质

- 矩阵 A 的迹线等于其特征值之和：

$$tr(A) = \sum_{i=1}^{n} \lambda_i$$

- 矩阵 A 的行列式等于特征值的乘积：

$$|A| = \prod_{i=1}^{n} \lambda_i$$

- 矩阵 A 的秩等于 A 的非零特征向量的数量。
- 一个对角矩阵 D 的特征值就等于其对角线的值

5.4 谱理论

5.4.1 定义

- 特征对是特征值及其相关特征向量的对。
- 和λ关联的特征空间是向量空间，其中：

$$(A - \lambda I) = 0$$

- 矩阵 A 的谱是所有特征值的集合：

$$\sigma(A) = \{\lambda \in \mathbb{C} : \lambda I - A \text{ is singular}\}$$ C 是所有特征值的空间

- A 的谱半径是其最大值特征值的大小：

$$\rho(A) = max\{|\lambda_1|, \dots, |\lambda_n|\}$$

5.4.2 定理：谱半径

Spectral radius is bounded by the infinity norm of a matrix:

$$\rho(A) = \lim_{k \to \infty} ||A^k||^{1/k}$$

Proof:

$$|\lambda|^k ||\mathbf{v}|| = ||\lambda|^k \mathbf{v}|| = ||A^k \mathbf{v}||$$

By the Cauchy–Schwarz inequality ($||\mathbf{uv}|| \leq ||\mathbf{u}|| \cdot ||\mathbf{v}||$):

$$|\lambda|^k ||\mathbf{v}|| \leq ||A^k|| \cdot ||\mathbf{v}||$$

Since $\mathbf{v} \neq 0$:

$$|\lambda|^k \leq ||A^k||$$

And we thus arrive at:

$$\rho(A) = \lim_{k \to \infty} ||A^k||^{1/k}$$

可参考：https://blog.csdn.net/lanchunhui/article/details/50619305

5.5 对角化

一个 n*n 的矩阵 A 有 n 个线性无关特征向量时，它是可对角化的。大多数方阵是可对角化的。

- 正规矩阵是可对角化的需要满足：

$$A^* A = A A^*$$

A*是 A 的共轭矩阵（共轭复数）

- 矩阵有 n 个不同的特征值是可对角化的

引理：与不同特征值相关的特征向量是线性无关的。

要使矩阵 A 对角化，需要考虑其特征值和特征向量。我们可以构造矩阵 D 和 V，其中 D 是 A 特征值的对角矩阵，V 是相应特征向量的矩阵：

$$D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

$$V = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix}$$

因为我们知道：AV = VD

所以我们可以把 A 对角化，通过：A = VDV^{-1}

如果所有特征值都是唯一的，那么 v 是正交的。因为正交矩阵的逆矩阵是它的转置，所以我们可以把对角化写成：A = VDV^T

5.6 对称矩阵

如果 A 是对称的，那么它的所有特征值都是实数，并且它的特征向量是正交的。

回顾上述对角化方程，我们可以通过以下方式对 A 进行对角化：A = VDV^T

利用上面的关系，我们可以写出下面的关系式：

给定 y = $V^T x$ ：有

$$x^T A x = x^T V D V^T x = y^T D y = \sum_{i=1}^{n} \lambda_i y_i^2$$

因此，如果我们想作下面的最大化：

$$max_{x \in \mathbb{R}^n} (x^T A x) \quad \text{subject to } ||x||_2^2 = 1$$

然后通过求 A 的最大特征值对应的特征向量，得到最大 X。

5.7 应用

特征值和特征向量的一些应用：

- 页面排名
- 薛定谔方程
- 主成分分析（PCA）

6 矩阵演算

6.1 梯度

如果函数f: $\mathbb{R}^{m*n} \to R$ ，把矩阵 A 作为输入，然后返回一个实数值，那么 f 的梯度如下：

$$\nabla_A f(A) \in \mathbb{R}^{\mathrm{mxn}} = \begin{bmatrix} \frac{\partial f(A)}{\partial A_{11}} & \frac{\partial f(A)}{\partial A_{12}} & \cdots & \frac{\partial f(A)}{\partial A_{1n}} \\ \frac{\partial f(A)}{\partial A_{21}} & \frac{\partial f(A)}{\partial A_{22}} & \cdots & \frac{\partial f(A)}{\partial A_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(A)}{\partial A_{m1}} & \frac{\partial f(A)}{\partial A_{m2}} & \cdots & \frac{\partial f(A)}{\partial A_{mn}} \end{bmatrix}$$

每个矩阵元素都是：

$$\nabla_A f(A)_{ij} = \frac{\partial f(A)}{\partial A_{ij}}$$

梯度 f(A)的大小总是和 A 的大小一样，因此，A 是一个向量 x 的话：

$$\nabla_x f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

6.2 梯度的性质

- $\nabla_x (f(x) + g(x)) = \nabla_x f(x) + \nabla_x g(x)$
- For $t \in \mathbb{R}$, $\nabla_x (t f(x)) = t \nabla_x f(x)$

6.3 海赛矩阵

关于 x 的 Hessian 矩阵可以写成如下式子或 H。它是偏导数的 n*n 矩阵：

$$\nabla_x^2 f(x) \in \mathbb{R}^{\mathrm{nxn}} = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

需要注意的是，hessian 是向量梯度的每个元素的梯度。例如，hessian 的第一列是 $\frac{\partial f(x)}{\partial x_1}$ 的梯度。

6.4 海赛矩阵：性质

施瓦兹定理：只要二阶导数存在且是连续的，偏导数的阶就不重要。

因此，hessian 始终是对称的：

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \frac{\partial^2 f(x)}{\partial x_j \partial x_i}$$

6.5 计算样例

都是演算公式，看原版即可！