特征值与特征向量

1. 对方矩阵 A，如果有 $A\mathbf{v}=\lambda\mathbf{v}$，则称 λ 是方阵 A 的特征值，$\mathbf{v}$ 是方阵 A 的特征向量。
特征向量和特征值的几何意义：
  1) 方向不变：线性变换（矩阵 A 表示）作用在特征向量，不改变向量的方向
  2) 缩放：线性变换（矩阵 A 表示）作用在特征向量只是对该向量的缩放

2. 特征分解：将矩阵分解为一组特征值和特征向量的乘积，只有可以对角化的矩阵才能进行特征分解。将特征向量组成矩阵 $V=(\mathbf{v_1},...,\mathbf{v_n})$，将所有特征值组成对应特征向量的对角矩阵 $\Lambda=diag(\lambda_1,...,\lambda_n)$，则

$$\mathbf{A}=\mathbf{V}\Lambda\mathbf{V}^{-1}$$

3. 杂
  1) 特征值都大于 0，称为正定矩阵
  2) 特征值都大于等于 0，称为半正定矩阵

对称矩阵

对称矩阵 A 满足 $A^T=A$，每个元素都为实数的矩阵称为实对称矩阵：
  1) 任意 $\mathbf{A}$ 是**实对称矩阵**，一定可以特征分解为

$$\mathbf{A}=\mathbf{Q}\Lambda\mathbf{Q}^{\mathbf{T}}$$

  其中 $\mathbf{Q}$ 是正交矩阵，由矩阵 $\mathbf{A}$ 的特征向量组成 $\{v_i\}_{i=1}^N$
  2) 实对称矩阵的所有特征值都是实数

矩阵转置性质

基本性质：

$$(A^T)^T=A \tag{1}$$
$$(A+B)^T=A^T+B^T \tag{2}$$
$$(cA)^T=cA^T \tag{3}$$
$$(AB)^T=B^TA^T \tag{4}$$
$$det(A^T)=det(A) \tag{5}$$
$$tr(A^T)=tr(A) \tag{6}$$
$$eigenValue(A)=eigenValue(A^T) \tag{7}$$

反对陈矩阵 (Skew-symmetric matrix) A:

$$A^T=-A \tag{8}$$

矩阵的逆的性质

$$AA^{-1}=I \tag{9}$$
$$(AB)^{-1}=(BA)^{-1} \tag{10}$$
$$(A^T)^{-1}=(A^{-1})^T, \text{当矩阵 A 可逆} \tag{11}$$

正交矩阵

满足 $Q^TQ=I$ 的矩阵 Q 称为正交矩阵。正交矩阵性质：

$$Q^T=Q^{-1} \tag{12}$$
$$det(Q)=+/-1 \tag{13}$$
$$\|Q\mathbf{v}\|=\|\mathbf{v}\|, \text{正交矩阵不改变向量长度} \tag{14}$$

1) 正交矩阵保持向量的夹角不变，旋转矩阵就是正交矩阵！
2) 正交矩阵的乘积依然是正交矩阵
3) 列向量和行向量都是正交的单位向量

相似矩阵

n 阶方矩阵 A 和 B 被称为相似（记作 $A\sim B$），如果存在一个可逆矩阵 P，使得 $B=P^{-1}AP$，被称为**相似变换**。
几何意义：A 和 B 是表示的是同一个线性变换，但是相对于不同的基。
相似矩阵 A 和 B 矩阵具有相同的特征值，但是并不一定有相同的特征向量。

$$eigenValue(A)=eigenValue(B) \tag{15}$$
$$det(A)=det(B) \tag{16}$$
$$tr(A)=tr(B) \tag{17}$$
$$rank(A)=rank(B) \tag{18}$$
$$A\sim B, B\sim C\rightarrow A\sim C \tag{19}$$

线性变换 Linear Map

线性变换 T: $V\rightarrow W$ 满足以下两个条件：
  1) 对所有的 $u,v\in V$，有 $T(u+v)=T(u)+T(v)$
  2) 对所有的 $v\in V$ 和所有的标量 c，有 $T(cv)=cT(v)$

即线性变换满足加法和标量乘法。除法呢？
透视投影矩阵涉及到除法操作（求像素齐次坐标时包含除法）不是一个线性变换！可以通过上面的线性变换的性质，例如加法性质自行证明。

奇异性 Singular

奇异性在数学中是指某种性质或状态异常或特殊。奇异矩阵会在线性变换时减少一个维度使得变换不可逆了，这很“异常”。
- 奇异矩阵是行列式为 0 的矩阵，奇异矩阵是不可逆矩阵，奇异矩阵的行或列线性相关，奇异矩阵至少有一个特征值为 0
- 奇异值分解 (Sigular Value Decomposition,SVD) 对于一个 $m\times n$ 的矩阵 A，奇异值分解可以表示为：

$$A=\mathbf{U}\Sigma\mathbf{V}^T$$

  Σ 对角矩阵上的值称为奇异值。

对角矩阵 Diag

对角矩阵 (D): 除了对角线都为 0 的方阵，记作 $D = diag(d_1, d_2, ..., d_n)$。对角矩阵有很多有优良的性质，计算上非常便利:

$$D^{-1} = diag(\frac{1}{d_1}, \frac{1}{d_2}, ..., \frac{1}{d_n}) \tag{20}$$
$$det(D) = d_1 \cdot d_2 \cdot ...d_n \tag{21}$$
$$tr(D) = d_1 + d_2 + ... + d_n \tag{22}$$
$$D^k = diag(d_1^k, d_2^k, ..., d_n^k) \tag{23}$$
$$D_k \cdot D_m = diag(d_{k1}d_{m1}, d_{k2}d_{m2}, ..., d_{kn}d_{mn}) \tag{24}$$
$$D_k + D_m = diag(d_{k1} + d_{m1}, d_{k2} + d_{m2}, ..., d_{kn} + d_{mn}) \tag{25}$$

矩阵的相似对角化：一个矩阵 A 与某个对角矩阵 Λ 相似，即存在可逆矩阵 P，使得

$$\Lambda = P^{-1}AP$$

则称矩阵 A 可以相似对角化, 称 Λ 为 A 的对角化形式, P 为对角化矩阵 (或特征向量矩阵)。对角化矩阵的主对角线元素的值全部为 A 的特征值。
相似对角化的条件: **并不是所有矩阵都可以相似对角化,** 条件是:

1) 矩阵 $A \in \mathbb{R}^{n\times n}$ 有 n 个线性无关的特征向量
2) 特征值的代数重数等于其几何重数：1) 代数重数是特征值在特征多项式中出现的次数，2) 几何重数是特征值对应的线性无关特征向量的个数（即特征空间的维数）

KL Divergence

KL 散度用来度量两个分布之间的距离，**不符合交换律，即交换两个分布后的距离不相等**

$$D_{KL}(P_\theta(x)\|Q_\phi(x)) = \sum_{x\in X} P_\theta(x) \cdot log(\frac{P_\theta(x)}{Q_\phi(x)}) \tag{26}$$

Bayes' Theorem

贝叶斯定理理解: 给定相关的证据来更新我们对某一个假设的信念 (先验).

$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)} \tag{27}$$

上面公式中, H 是我们的假设, E 是我们的证据 (数据), $P(H)$ 是我们对假设的信念，比如说是对假设的先天的看法, 我们带着假设采集了很多新的证据 ($P(E|H)$), 由此更新了我们新的认知。$P(H)$ 称为先验，$P(H|E)$ 称为后验，$P(E|H)$ 称为似然。

联合概率、条件概率、边缘概率

联合概率：$p(x, y)$ for $p(X = x, Y = y)$
条件概率 (X conditioned on Y): $p(x|y) = \frac{p(x,y)}{p(y)}$
边缘概率: $p(y_i) = \sum_{x\in X} p(x, y_i)$