3 矩阵和线性代数

① 奇异值分解 (Singular Value Decomposition) ⟵ 整理：见⑮ 的性质4：①②

对 m×n 阶实矩阵 $A_{m\times n}$，存在一个分解使得

单位正交方阵，即 $U^TU=I$，$V^TV=I$

$$A_{m\times n} = \boxed{U_{m\times m}}\ \Sigma_{m\times n}\ \boxed{V^T_{n\times n}}$$

$$\begin{bmatrix} 1&0&0&0&2\\ 0&0&3&0&0\\ 0&0&0&0&0\\ 0&4&0&0&0 \end{bmatrix} = \begin{bmatrix} 0&0&1&0\\ 0&1&0&0\\ 0&0&0&-1\\ 1&0&0&0 \end{bmatrix} \begin{bmatrix} 4&0&0&0&0\\ 0&3&0&0&0\\ 0&0&\sqrt5&0&0\\ 0&0&0&0&0 \end{bmatrix} \begin{bmatrix} 0&1&0&0&0\\ 0&0&1&0&0\\ \sqrt{0.2}&0&0&0&\sqrt{0.8}\\ 0&0&0&1&0\\ -\sqrt{0.8}&0&0&0&\sqrt{0.2} \end{bmatrix}$$

$\sigma_3$ 的左奇异向量　　矩阵A的奇异值　　$\sigma_3$ 的右奇异向量 → 突出，奇特，非凡

PCA（主成因分析）本质是求几个主方向，这些方向方差最大，

$$\begin{bmatrix} \sigma_1&&&\\ &\sigma_2&&\\ &&\sigma_3&\\ &&&\sigma_4 \end{bmatrix}$$

$\sigma_i$ 是特征值 $\lambda_i$ 开根号得到的

截取前k个系数，使前k个系数加和大于90%

② 方阵行列式：1×1方阵，2×2方阵，3×3方阵，……

$|A| = a_{11}$　　$|A| = a_{11}a_{22} - a_{12}a_{21}$　　$|A| = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}$

$|A_{n\times n}|$：主对角线（↘↘↘）元素乘积 减去次对角线元素（↙↙↙）乘积，

也等于任意一行/列 元素与其对应的代数余子式乘积之和

$$\forall\ 1\le j\le n,\quad |A| = \sum_{i=1}^{n} a_{ij}(-1)^{i+j}M_{ij}$$

$$\forall\ 1\le i\le n,\quad |A| = \sum_{j=1}^{n} a_{ij}(-1)^{i+j}M_{ij}$$

代数余子式 $M_{ij}$：在 $A_{n\times n}$ 中，划掉 $a_{ij}$ 所在的第i行和第j列余下元素组成的矩阵 ~~$M_{ij}$~~ 得到的行列式 $M_{ij}$ 乘以 $(-1)^{i+j}$ 记为 $A_{n\times n}$ 关于 $a_{ij}$ 的代数余子式 $A_{ij}$

③ 伴随矩阵 $A^*$：由原矩阵A各元素代数余子式组成的矩阵

$$A^* = \begin{bmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{bmatrix} \qquad A_{ij} = (-1)^{i+j} M_{ij}$$

$$A \cdot A^* = \begin{bmatrix} |A| & & & 0 \\ & |A| & & \\ & & \ddots & \\ 0 & & & |A| \end{bmatrix} = |A| \cdot I \Rightarrow A^{-1} = \frac{1}{|A|} A^*$$

④ 范德蒙德行列式

用数学归纳证明

⇓

$$D_n = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{vmatrix} = \prod_{i,j\,(n\geq i>j\geq 1)} (x_i - x_j) \neq 0$$

性质：对任意 $i, j$，如果满足 $\forall i,j \quad x_i \neq x_j$，则 $D_n$ 可逆

用途：$D_n$ 可逆 ⟺ 能得到 $D_n^{-1}$（逆矩阵）⟺ 以 $D_n$ 做为方程组参数，则参数可解 ⟺ 用范德蒙德行列式构造一个 $n$ 阶多项式，得到一条曲线，让曲线穿过 $n+1$ 个已知点，并对多项式参数求解 ⇐ 拉格朗日插值法，

⑤ 拟合 V.S 插值

100个点，用99阶多项式来拟合，保证曲线经过每个点 ⟶ 插值

100个点，用20阶多项式来拟合，保证各个点近似落在曲线上 ⟶ 拟合

⑥ 克莱姆法则（略）

⑦ 状态转移概率矩阵，马尔可夫，与平稳分布

*马尔可夫模型：假定状态转移概率只依赖前一个时刻的状态，不需要更早的状态，它是一个简单的贝叶斯网络

*例子：$\pi^{(n+1)} = \pi^{(n)} \cdot P$

$$
(\pi_1^{n+1}, \pi_2^{n+1}, \pi_3^{n+1}) = (\pi_1^{n}, \pi_2^{n}, \pi_3^{n}) \begin{bmatrix} 0.65, & 0.28, & 0.07 \\ 0.15, & 0.67, & 0.18 \\ 0.12, & 0.36, & 0.52 \end{bmatrix}
$$

子代 上层 中层 下层　　父代 上层 中层 下层(阶级)

状态转移概率

不断迭代最终的 $\pi(\pi_1, \pi_2, \pi_3)$ 稳定在 [0.286, 0.488, 0.225]，它只与状态转移概率矩阵中有关，与初始值无关

↓ 对于非周期马尔可夫随机过程

而这个稳定后的 $\pi$，是 $P$ 最大的特征值对应的特征向量（在迭代过程中，这个特征值不断增大，直到为1，其它特征值不断减小，直到为0）

*平稳分布：状态转移概率矩阵的特性，$\pi$ 最终稳定

⑧ 矩阵乘法，加速(略)(看代码)

⑨ 矩阵和向量的乘法：$\vec{y} = A_{m\times n} \cdot \vec{x}$

一个

表示从n维坐标空间到另一个m维坐标空间的线性变换

$A_{n\times n}$ 可以实现向量

平移，旋转，平移+旋转，伸缩，…

⑩ 矩阵的k阶子式：

$A_{m\times n}$ 矩阵中任取k行，k列（共有 $C_m^k C_n^k$ 种取法），得到子矩阵的行列式

⑪ 矩阵的秩（理解成倒数的推广）：矩阵 $A_{m\times n}$ 有一个 $r$ 阶子式不为0，同时所有 $r+1$ 阶子式都为0，则称 $r$ 为 $A_{m\times n}$ 的秩

$D$

则 1) $D$：$A_{m\times n}$ 的最高阶非零子式

$R(A_{m\times n}) = r$：记为 $A_{m\times n}$ 的秩

也与方程组求解有关

特性：

√ $n\times n$ 的可逆矩阵，秩为 $n$

√ 可逆矩阵又称为满秩矩阵

√ 矩阵的秩 = 它行(列)向量组的秩

⑫ 秩与线性方程组求解：

对于 $n$ 元线性方程组 $A\vec{x}=\vec{b}$

<a> 无解 $\Leftrightarrow R(A) < R(A,b)$

<b> 有唯一解 $\Leftrightarrow R(A) = R(A,b) = n$

<c> 有无限多解 $\Leftrightarrow R(A) = R(A,b) < n$

推论

<d> $A\vec{x}=0$ 有非零解 $\Leftrightarrow R(A) < n$

<e> $A\vec{x}=b$ 有解 $\Leftrightarrow R(A) = R(A,b)$

⑬ $C = A \cdot B$ 向量组 $C$ 能由向量组 $B$ 线性表示

$\Leftrightarrow$ 矩阵 $C=(c_1,c_2,\cdots c_m)$ 的秩等于矩阵 $(A,B)=(a_1,a_2\cdots a_m,b_1,b_2\cdots b_m)$ 的秩，即 $R(A)=R(A,B)$

观测信号 原始信号

⑭ 正交阵：$\boxed{A^TA=I}$ $\Leftrightarrow$ $\boxed{A\text{的行/列向量都是单位向量，且两两正交}}$

正交变换：$A\vec{x}$ 其中 $A$ 是正交阵，$\vec{x}$ 是向量，正交变换不改变向量长度

⑮ 特征值和特征向量：$A_{n\times n}\vec{x}=\lambda\vec{x}$ 若这样的 $\lambda$ 和 $\vec{x}$ 存在

$n$ 维非 0 向量

就认为 $\lambda$ 是 $A_{n\times n}$ 的一个特征值，$\vec{x}$ 是 $\lambda$ 对应的特征向量

用途举例：白化 (whiting)

（求特征值方法见 PPT）

性质 1：设 $A_{n\times n}$ 特征值为 $\lambda_1,\lambda_2,\cdots,\lambda_n$，则

① $\lambda_1+\lambda_2+\cdots+\lambda_n = a_{11}+a_{22}+\cdots+a_{nn}$ 叫作 $A_{n\times n}$ 的迹

② $\lambda_1\cdot\lambda_2\cdots\lambda_n = |A|$

用于 SVD

性质 2：① $\lambda$ 是 $A$ 的特征值 $\Rightarrow \lambda^2$ 是 $A^2$ 的特征值

② $A$ 可逆时，$\lambda^{-1}$ 是 $A^{-1}$ 的特征值

性质 3：① 不同特征值对应的特征向量线性无关

性质 4：① 实对称阵的不同特征值的特征向量正交

② 若 $A_{n\times n}$ 是对称阵，则必有正交阵 $P$，使得 $P^{-1}AP = P^TAP = \Lambda$

特征向量组成

$P^TAP=\Lambda=\begin{bmatrix}\lambda_1 & & \\ & \lambda_2 & \\ & & \ddots \\ & & & \lambda_n\end{bmatrix}$

$PP^TAPP^T = P\Lambda P^T$

$IAI = P\Lambda P^T$

$A = P\Lambda P^T$

特征值

正交阵定义⑭

$A$ 与 $\Lambda$ 互为合同矩阵

协方差矩阵，二次型矩阵，无向图邻接矩阵

用于降维，PCA，SVD

⑯ 正定阵：正数在n阶的推广

对于n阶方阵A，若∀n阶向量x，都有 $\vec{x}^T A \vec{x} > 0$，则 A 是正定阵。

$\geq 0$ 半正定阵

$< 0$ 负定阵

$\leq 0$ 半负定阵

推论 性质

对∀$A_{m\times n}$，$A^TA$ 一定是半正定阵

$(A^TA)^T = A^TA$

$\vec{x}^T(A^TA)\vec{x} = (\vec{x}^TA^T)(A\vec{x}) = (A\vec{x})^T(A\vec{x}) \geq 0$

→ 用于线性回归

判定：对称阵A为正定阵 ⟺ A的特征值都为正 ⟺ A的顺序主子式大于0

⑰ 基，正交基，标准正交基

基：如果存在一组向量 $e_1, e_2, \cdots e_k$ 使得∀列向量 $a_i$ 都能用它们线性表示，就说 $e_1, e_2, \cdots e_k$ 是矩阵的基。

正交基：如果 $e_1, e_2, \cdots e_k$ 还是两两正交的

⑱ QR分解 → 求解矩阵的逆，特征值 → 用于梯度下降计算 → 线性回归

LFM隐特征模型

⑲ 向量的导数：$\vec{y}_{m\times1} = A_{m\times n}\vec{x}_{n\times1}$

$$\vec{y} = A\cdot\vec{x} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = A_{m\times n}\vec{x}_{n\times1}$$

定义 $\boxed{\dfrac{\partial \vec{y}}{\partial \vec{x}} = \dfrac{\partial A\vec{x}}{\partial \vec{x}} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & & a_{m2} \\ \vdots & \vdots & \ddots & \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix} = A^T}$

一阶 $\dfrac{dy}{dx} = \dfrac{dax}{dx} = a$

n阶 $\dfrac{\partial \vec{y}}{\partial \vec{x}} = \dfrac{\partial A\vec{x}}{\partial \vec{x}} = A^T$

⑳ 向量的偏导公式 ⇒ 用于线性回归

$$\frac{\partial A\vec{x}}{\partial \vec{x}} = A^T \quad (如定义)$$

$$\frac{\partial A\vec{x}}{\partial \vec{x}^T} = \frac{\partial(\vec{x}^T A^T)^T}{\partial \vec{x}^T} = (A^T)^T = A$$

$$\frac{\partial \vec{x}^T A}{\partial \vec{x}} = \frac{\partial (A^T\vec{x})^T}{\partial \vec{x}} = \left(\frac{\partial A^T\vec{x}}{\partial \vec{x}}\right)^T = (A^T)^T = A.$$

㉑ 标量对向量的导数

若 $y = (\vec{x}_{n\times1})^T A_{n\times n}\vec{x}_{n\times1}$ 则 $\dfrac{\partial y}{\partial \vec{x}_{n\times1}} = \dfrac{\partial(\vec{x}_{n\times1}\cdot A_{n\times n}\vec{x}_{n\times1})}{\partial \vec{x}_{n\times1}} = (A^T + A)\vec{x}_{n\times1}$

当 $A^T = A$ (对称阵)时：$\dfrac{\partial y}{\partial \vec{x}_{n\times1}} = 2A\vec{x}_{n\times1}$

㉒ 标量对方阵的导数.

$$\frac{\partial |A|}{\partial a_{ij}} = \frac{\partial \sum_{i=1}^{n} a_{ij}(-1)^{i+j}M_{ij}}{\partial a_{ij}} = (-1)^{i+j}M_{ij} = A^*_{ji}$$ 即 $a_{ji}$ 的伴随矩阵.

$$\frac{\partial |A|}{\partial A} = (A^*)^T = |A|(A^{-1})^T$$

$$\uparrow$$

$$A\cdot A^* = |A|I.$$

㉓ FFT 快速傅立叶变换(略)

# 数据清洗

① 用pandas读数据集，根据字符串编辑距离对拼错单词做模糊替换

② 手机用户流失率分析：字段缺失，年龄等，详见⑦

③ PCA主成因分析，从 $n$ 个特征中选出最影响力的 $k<n$ 个特征

详解见后面

$$(\lambda_1+\cdots+\lambda_k) \geq 90\% (\lambda_1+\cdots+\lambda_k+\cdots+\lambda_n)$$

④ PCA主成因分析（选出影响力最大的k个特征）→ Poly（用多项来对PCA选出来的特征进行组合）→ 再用 LR/RF/SVM… 分类

二阶模型 ⇒ 三阶模型

⑤ 异常值清洗

⇒ 用随机森林 6.5.polution.py

⑥ one-hot编码.

| | | | | |
|---|---|---|---|---|
| 价格 | 0001 (very high) | 0010 (high) | 0100 (low) | 1000 (very low) |
| 车门数量 | 00001 (2门) | 00010 (3门) | 00100 (4门) | 01000 (5门或更多) |
| 载人数目 | 0001 (2人) | 0010 (4人) | 0100 (6人) | 1000 (更多) |

☆ 这样做的原因是，不希望原始特征之间产生关联，例如28度酒×2不代表56度酒。

☆ LR经常需要one-hot；RF，决策树逻辑上不需要，实际中经常也会。

⑦ 手机用户流失率分析

☆ 有些数据需要从字符串转成ID

☆ 有些数据需要根据给定的分界值进行分组

☆ 有些数据需要取对数

☆ 有些数据需要简单分成若干等份

☆ 有些数据需要做one-hot编码

⑧ PCA主成分分析：特征降维，例如把 $f_1, f_2, f_3, f_4$ 四个特征映射到 $u_1, u_2$ 两个特征。

<a> 怎样的 $u_1$ 最好？所有样本 $x^{(i)}=(f_1^{(i)}, f_2^{(i)}, f_3^{(i)}, f_4^{(i)})$ 映射到 $u_1=(f_1^{(u_1)}, f_2^{(u_1)}, f_3^{(u_1)}, f_4^{(u_1)})$ 后，都集在 $u_1$ 附近（到 $u_1$ 距离最小，方差最大，$u_1$ 区分样本能力最强）

Good

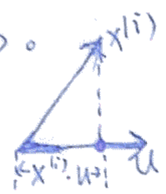
Bad.

<b> 向量投影，当 $u$ 是单位向量（即 $u^Tu=1$）时，$x^{(i)}\cdot u$ 即为 $x^{(i)}$ 在 $u$ 上的投影。

分

行向量 列向量

点乘




同向时点乘最大

垂直时点乘为零

<c> 求所有样本在 $u$ 上的投影，求投影值的方差

$$\begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(n)} \end{bmatrix}\cdot u = \begin{bmatrix} f_1^{(1)} & f_2^{(1)} & f_3^{(1)} & f_4^{(1)} \\ f_1^{(2)} & f_2^{(2)} & f_3^{(2)} & f_4^{(2)} \\ & & \vdots & \\ f_1^{(n)} & f_2^{(n)} & f_3^{(n)} & f_4^{(n)} \end{bmatrix} \begin{bmatrix} f_1^{(u_1)} \\ f_2^{(u_1)} \\ f_3^{(u_1)} \\ f_4^{(u_1)} \end{bmatrix} = \begin{bmatrix} z^{(1)} \\ z^{(2)} \\ z^{(3)} \\ \vdots \\ z^{(n)} \end{bmatrix}$$

feature

$u_1$ $Z$

求 $z^{(1)} \dots z^{(n)}$ 方差，方差越大 $u_1$ 越合适

更进一步（如本页顶部的图）可通过平移 $u_1$，来让投影均值 $\frac{1}{n} z^{(i)}$ 为零（即 z 去中心化）

方差 $Var(z)=\frac{1}{n}\sum_{i=1}^{n}(z^{(i)}-0)^2=\frac{1}{n}\sum_{i=1}^{n}(z^{(i)})^2=\frac{1}{n}z^Tz=\frac{1}{n}(xu)^T(xu)$

常数

<d> 转为 $u^T\cdot u=1$ 约束下求 $(xu)^T(xu)$ 极值问题，用拉格朗日法，

（单位列向量的定义）

$L(u,\lambda)=(xu)^T(xu)+\lambda(u^Tu-1)=u^T(x^Tx)u+\lambda u^Tu-\lambda$.

$\frac{\partial L(u,\lambda)}{\partial u}=2x^Txu+2\lambda u$ ← $\boxed{线代㉑：A^T=A 时 \frac{\partial(x^TAx)}{\partial x}=2Ax}$

令偏导为零：$2x^Txu=2(-\lambda)u=2\lambda' u$ $\boxed{线代⑮：特征向量定义}$

4x4已知矩阵 $x^Tx$的 u 也是 $x^Tx$ 特征向量

求解该方程，得到的 $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ 就是特征值，对应的 $u_1, u_2, u_3, u_4$

因为 $x^Tx$ 是实对称阵，因此一定有解，且 $u_1, u_2, u_3, u_4$ 正交 $\boxed{线代⑰性质4}$

<f> 按 $\lambda_{(1)}\geq\lambda_{(2)}\geq\lambda_{(3)}\geq\lambda_{(4)}$ 排序，按总和的90%截断，来从 $u_1, u_2, u_3, u_4$ 可用作线性空间坐标轴