Mathematics_for_Machine_Learning(1)——线性代数

笔记本:	UAV+DL		
创建时间:	2018/10/15 14:04	**更新时间:**	2018/10/17 17:13
作者:	zyhhhh1993@gmail.com		

1、 矢量的性质：a)矢量相加仍是矢量；b)矢量乘以标量得到相同类型矢量。
举例：几何矢量；多项式；音频信号；n维实数集中的元素
矢量空间和它的性质是机器学习的基础
矩阵可以被认为是矢量的集合

2、线性方程组
定义：符合以下形式，其中a,b是常数，x是未知数，每个满足下面方程组的n维数组（x1……xn）称为解

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

对于一个实值线性方程组系统，有无解、唯一解、无穷解三种情况
线性方程组解的几何解释：未知数为2维时，相当于求平面上两条直线的关系：两线平行为无解；两线交叉为唯一解；两线重合为无穷多解

3、矩阵
3.1定义：一个(m,n)的矩阵A表示由m×n个aij组成的元组，顺序按照m行n列排成矩形，形式如下：

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad a_{ij} \in \mathbb{R}.$$

行向量是只有一行的矩阵；列向量是只有一列的矩阵
3.2 矩阵的加法：对应位置上的元素相加

$$\boldsymbol{A} + \boldsymbol{B} := \begin{bmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{bmatrix} \in \mathbb{R}^{m \times n}.$$

3.3矩阵的乘法：A的第i行和B的第j列元素一一对应相乘然后相加，此操作为点积。
注：矩阵相乘的条件是相邻维度一样
3.4单位阵：矩阵为方阵（行和列相等），且除了对角线元素全为1之外，其它元素均为0

$$\boldsymbol{I}_n = \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{n\times n}$$

3.5矩阵的性质：

a）结合律

$$\forall \boldsymbol{A} \in \mathbb{R}^{m\times n}, \boldsymbol{B} \in \mathbb{R}^{n\times p}, \boldsymbol{C} \in \mathbb{R}^{p\times q} : (\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$$

b）分配律

$$\forall \boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m\times n}, \boldsymbol{C}, \boldsymbol{D} \in \mathbb{R}^{n\times p} : (\boldsymbol{A}+\boldsymbol{B})\boldsymbol{C} = \boldsymbol{AC} + \boldsymbol{BC}$$
$$\boldsymbol{A}(\boldsymbol{C}+\boldsymbol{D}) = \boldsymbol{AC} + \boldsymbol{AD}$$

c）中性元素

$$\forall \boldsymbol{A} \in \mathbb{R}^{m\times n} : \boldsymbol{I}_m\boldsymbol{A} = \boldsymbol{A}\boldsymbol{I}_n = \boldsymbol{A} \quad \boldsymbol{I}_m \neq \boldsymbol{I}_n \text{ for } m \neq n$$

3.6矩阵的逆：对于Rnxn中的矩阵A和矩阵B有AB=In，那B就是A的逆，记作A-1。如果转置存在，则A被称作正定/可逆/非奇。可以根据矩阵的行列式是否为0来判别矩阵是否可逆

3.7矩阵的转置：对于mxn的矩阵A和nxm的矩阵B，如果有aij=bji，那么B是A的转置，记作AT

3.8逆和转置的性质：

$$\boldsymbol{AA}^{-1} = \boldsymbol{I} = \boldsymbol{A}^{-1}\boldsymbol{A}$$
$$(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$$
$$(\boldsymbol{A}+\boldsymbol{B})^{-1} \neq \boldsymbol{A}^{-1} + \boldsymbol{B}^{-1}$$
$$(\boldsymbol{A}^\top)^\top = \boldsymbol{A}$$
$$(\boldsymbol{A}+\boldsymbol{B})^\top = \boldsymbol{A}^\top + \boldsymbol{B}^\top$$
$$(\boldsymbol{AB})^\top = \boldsymbol{B}^\top\boldsymbol{A}^\top$$

如果A是可逆的，那么A的转置也是可逆的 $(\boldsymbol{A}^{-1})^\top = (\boldsymbol{A}^\top)^{-1} =: \boldsymbol{A}^{-\top}$；如果A和A的转置相等，表示A是对称的，这必须要求A是方阵

3.9矩阵和标量的乘法：矩阵乘以一个标量等于其中每个元素都乘以此标量

性质：a）分配律：

$$(\lambda+\psi)\boldsymbol{C} = \lambda\boldsymbol{C} + \psi\boldsymbol{C}, \quad \boldsymbol{C} \in \mathbb{R}^{m\times n}$$
$$\lambda(\boldsymbol{B}+\boldsymbol{C}) = \lambda\boldsymbol{B} + \lambda\boldsymbol{C}, \quad \boldsymbol{B}, \boldsymbol{C} \in \mathbb{R}^{m\times n}$$

b）结合律：

$$(\lambda\psi)\boldsymbol{C} = \lambda(\psi\boldsymbol{C}), \quad \boldsymbol{C} \in \mathbb{R}^{m\times n}$$
$$\lambda(\boldsymbol{BC}) = (\lambda\boldsymbol{B})\boldsymbol{C} = \boldsymbol{B}(\lambda\boldsymbol{C}) = (\boldsymbol{BC})\lambda, \quad \boldsymbol{B} \in \mathbb{R}^{m\times n}, \boldsymbol{C} \in \mathbb{R}^{n\times k}$$

c）此外还允许标量值的移动：

$(\lambda \boldsymbol{C})^\top = \boldsymbol{C}^\top \lambda^\top = \boldsymbol{C}^\top \lambda = \lambda \boldsymbol{C}^\top$ since $\lambda = \lambda^\top$ for all $\lambda \in \mathbb{R}$

3.10 线性方程组可以写成紧凑表达式Ax=b

4、解线性方程组

4.1特解和通解

三步求解：a）找到一个Ax=b的特解；b）找到Ax=0的所有解；c）结合上两步结果得到通解

通解和特解都不是唯一的

对于一般线性方程组，通过基本变换进行高斯消元能化成简单的形式

4.2基本变换

基本变换操作包括：a）交换矩阵两行；b）某一行乘以一个非零常数；c）两行相加

主元和阶梯结构：矩阵某一行的前导系数（从左往右第一个非零数）称作主元，并且严格位于高于此行的主元右边，所以会产生阶梯的结构。

4.3行阶梯矩阵：只包括0的所有行都在矩阵底部；至少有一个非零数的行总是在只有0的行之上；对于有非零数的行，从左数起第一个非零数（称为前导系数或者主元）总是严格位于其上面行的主元右边。

4.4 基本变量和自由变量

在行阶梯矩阵中，相对于主元的变量称为基本变量，其余的是自由变量。

4.5 获得特解

用主元列来表示矩阵右边部分。b=求和（λipi）。如果从最右边的主元列开始到左边，确定λi最容易

4.6行最简形式：处于行阶梯形式中；每个主元都为1；主元是这一列中唯一的非零数

将矩阵化成行最简形式，再用主元列去表示非主元列，有几列非主元列就表示AX=b的所有解里有几组不同基。

4.7 “-1技巧”

将行最简形式用[0...0 -1 0...0]补充成n x n 的方阵，那么其中包含-1的主元列就是Ax=0的解空间，之后会称为内核或者null空间

4.8 计算矩阵的逆

将矩阵A扩展成[A|In],通过基本变换成[In|A-1]就能得到A的逆

4.9 解线性方程组的算法

a）可以直接求A的逆，但是A必须是方阵且可逆

b）摩尔彭若斯广义逆(ATA)-1AT来求解，对应于最小范数的最小二乘解

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \iff \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^\top \boldsymbol{b} \iff \boldsymbol{x} = (\boldsymbol{A}^\top \boldsymbol{A})^{-1}\boldsymbol{A}^\top \boldsymbol{b}$$

c）高斯消除，但是对于变量过多的情况计算量成三次方增加

d）通过固定的迭代方法间接处理，如Richardson method Jacobi method Gauß-Seidel method successive over-relaxation method Krylov subspace methods conjugate gradients generalized minimal residual biconjugate gradients

迭代方法的主要思想是通过一步一步的迭代来使残差逐渐收敛到解。

5、向量空间

群的定义：由一个集合和一个二元运算组成，且符合下面四个性质：

1. Closure of $\mathcal{G}$ under $\otimes$: $\forall x, y \in \mathcal{G} : x \otimes y \in \mathcal{G}$
2. *Associativity:* $\forall x, y, z \in \mathcal{G} : (x \otimes y) \otimes z = x \otimes (y \otimes z)$
3. *Neutral element:* $\exists e \in \mathcal{G}\, \forall x \in \mathcal{G} : x \otimes e = x$ and $e \otimes x = x$
4. *Inverse element:* $\forall x \in \mathcal{G}\, \exists y \in \mathcal{G} : x \otimes y = e$ and $y \otimes x = e$.

如果任意两个集合中的元素满足交换性，则称为阿贝尔群