Data Mining and Matrices

02 – Linear Algebra Refresher

Rainer Gemulla, Pauli Miettinen

April 18, 2013

Vectors

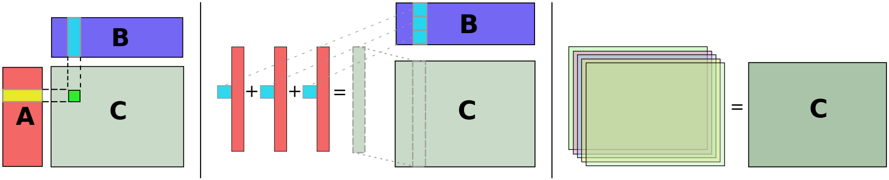
- A **vector** is
  - a 1D array of numbers
  - a geometric entity with magnitude and direction
  - a matrix with exactly one row or column $\Rightarrow$ row and column vectors
- A **transpose** $\mathbf{a}^T$ transposes a row vector into a column vector and vice versa
- The **norm** of vector defines its magnitude
  - Euclidean or L_2: $\|\mathbf{a}\| = \|\mathbf{a}\|_2 = \left(\sum_{i=1}^{n} a_i^2\right)^{1/2}$
  - General L_p $(1 \leq p \leq \infty)$: $\|\mathbf{a}\|_p = \left(\sum_{i=1}^{n} a_i^p\right)^{1/p}$
- A **dot product** of two vectors of same dimension is $\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^{n} a_i b_i$
  - Also known as **scalar product** or **inner product**
  - Alternative notations: $\langle \mathbf{a}, \mathbf{b} \rangle$, $\mathbf{a}^T\mathbf{b}$ (for column vectors), $\mathbf{a}\mathbf{b}^T$ (for row vectors)
- In Euclidean space we can define $\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\|\|\mathbf{b}\| \cos\theta$
  - θ is the angle between $\mathbf{a}$ and $\mathbf{b}$
  - $\mathbf{a} \cdot \mathbf{b} = 0$ if $\theta = \frac{1}{2}\pi + k\pi$ (they are **orthogonal**)

Matrix algebra

- Matrices in $\mathbb{R}^{n \times n}$ form a ring
  - Addition, subtraction, and multiplication
  - Addition and subtraction are element-wise
  - Multiplication doesn't always have inverse (division)
  - Multiplication isn't commutative ($\mathbf{AB} \neq \mathbf{BA}$ in general)
  - The identity for the multiplication is the **identity matrix** $\mathbf{I}$ with 1s on the main diagonal and 0s elsewhere
    - $\mathbf{I}_{ij} = 1$ iff $i = j$; $\mathbf{I}_{ij} = 0$ iff $i \neq j$
- If $\mathbf{A} \in \mathbb{R}^{m \times k}$ and $\mathbf{B} \in \mathbb{R}^{k \times n}$, then $\mathbf{AB} \in \mathbb{R}^{m \times n}$ with $(\mathbf{AB})_{ij} = \sum_{\ell=1}^{k} a_{i\ell} b_{\ell j}$
  - The **inner dimension** (k) of $\mathbf{A}$ and $\mathbf{B}$ must agree
  - The dimensions of the product are the outer dimensions of $\mathbf{A}$ and $\mathbf{B}$

Intuition for Matrix Multiplication

- Element $(\mathbf{AB})_{ij}$ is the inner product of row i of $\mathbf{A}$ and column j of $\mathbf{B}$
- Row i of $\mathbf{AB}$ is the linear combination of rows of $\mathbf{B}$ with the coefficients coming from row i of $\mathbf{A}$
  - Similarly, column j is a linear combination of columns of $\mathbf{A}$
- Matrix $\mathbf{AB}$ is a sum of k matrices $\mathbf{a}_\ell \mathbf{b}_\ell^T$ obtained by multiplying ℓ-th column of $\mathbf{A}$ with ℓ-th row of $\mathbf{B}$
  - This is known as vector **outer product**



Matrices as linear mappings

- A matrix $\mathbf{M} \in \mathbb{R}^{m \times n}$ is a **linear mapping** from $\mathbb{R}^n$ to $\mathbb{R}^m$
  - If $\mathbf{x} \in \mathbb{R}^n$ then $\mathbf{y} = \mathbf{M}\mathbf{x} \in \mathbb{R}^m$ is the image of $\mathbf{x}$
  - $\mathbf{y}_i = \sum_{j=1}^{n} \mathbf{M}_{ij}\mathbf{x}_j$
- If $\mathbf{A} \in \mathbb{R}^{m \times k}$ and $\mathbf{B} \in \mathbb{R}^{k \times n}$, then $\mathbf{AB}$ is a mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$
  - Combination of $\mathbf{A}$ and $\mathbf{B}$
- Square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is **invertible** if there is matrix $\mathbf{B} \in \mathbb{R}^{n \times n}$ such that $\mathbf{AB} = \mathbf{I}$
  - Matrix $\mathbf{B}$ is the **inverse** of $\mathbf{A}$, denoted $\mathbf{A}^{-1}$
  - If $\mathbf{A}$ is invertible, then $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$
    - $\mathbf{A}\mathbf{A}^{-1}\mathbf{x} = \mathbf{A}^{-1}\mathbf{A}\mathbf{x} = \mathbf{x}$
  - Non-square matrices don't have (general) inverses but can have **left** or **right inverses**: $\mathbf{AR} = \mathbf{I}$ or $\mathbf{LA} = \mathbf{I}$
- The **transpose** of $\mathbf{M} \in \mathbb{R}^{m \times n}$ is a linear mapping $\mathbf{M}^T : \mathbb{R}^m \to \mathbb{R}^n$
  - $(\mathbf{M}^T)_{ij} = \mathbf{M}_{ji}$
  - Generally, transpose is **not** the inverse ($\mathbf{A}\mathbf{A}^T \neq \mathbf{I}$)

Matrix rank and linear independence

- A vector $\mathbf{u} \in \mathbb{R}^n$ is **linearly dependent** on set of vectors $V = \{\mathbf{v}_i\} \subset \mathbb{R}^n$ if $\mathbf{u}$ can be expressed as a linear combination of vectors in V
  - $\mathbf{u} = \sum_i a_i \mathbf{v}_i$ for some $a_i \in \mathbb{R}$
  - Set V is linearly dependent if some $\mathbf{v}_i \in V$ is linearly dependent on $V \setminus \{\mathbf{v}_i\}$
  - If V is not linearly dependent, it is **linearly independent**
- The **column rank** of matrix $\mathbf{M}$ is the number of linearly independent columns of $\mathbf{M}$
- The **row rank** of $\mathbf{M}$ is the number of linearly independent rows of $\mathbf{M}$
- The **Schein rank** of $\mathbf{M}$ is the least integer k such that $\mathbf{M} = \mathbf{AB}$ for some $\mathbf{A} \in \mathbb{R}^{m \times k}$ and $\mathbf{B} \in \mathbb{R}^{k \times n}$
  - Equivalently, the least k such that $\mathbf{M}$ is a sum of k vector outer products
- **All these ranks are equivalent!**
  - Matrix has rank 1 iff it is an outer product of two vectors

Matrix norms

- Matrix norms measure the magnitude of the matrix
  - Magnitude of the values
  - Magnitude of the image
- **Operator norms** measure how big the image of an unit vector can be
  - For $p \geq 1$, $\|\mathbf{M}\|_p = \max\{\|\mathbf{Mx}\|_p : \|\mathbf{x}\|_p = 1\}$
- The **Frobenius norm** is the vector-L_2 norm applied to matrices
  - $\|\mathbf{M}\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} \mathbf{M}_{ij}^2\right)^{1/2}$
  - N.B. $\|\mathbf{M}\|_F \neq \|\mathbf{M}\|_2$ (but sometimes Frobenius norm is referred to as L_2 norm)

Matrices as systems of linear equations

- A matrix can hold the coefficients of a system of linear equations
  - The original use of matrices (Chinese *The Nine Chapters on the Mathematical Art*)

$$
\begin{aligned}
a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,m}x_m &= b_1 \\
a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,m}x_m &= b_2 \\
&\vdots \\
a_{n,1}x_1 + a_{n,2}x_2 + \cdots + a_{n,m}x_m &= b_n
\end{aligned}
\Leftrightarrow
\begin{pmatrix} a_{1,1} & a_{1,2}, & \cdots & a_{1,m} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,m} \end{pmatrix}
\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}
=
\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}
$$

- If the coefficient matrix $\mathbf{A}$ is invertible, the system has exact solution $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$
- If $m < n$ the system is **underdetermined** and can have infinite number of solutions
- If $m > n$ the system is **overdetermined** and (usually) does not have an exact solution
- The **least-squares** solution is the vector $\mathbf{x}$ that minimizes $\|\mathbf{Ax} - \mathbf{b}\|_2^2$
  - Linear regression

Special types of matrices

- The **diagonals** of matrix $\mathbf{M}$ go from top-left to bottom-right
  - The main diagonal contains the elements $\mathbf{M}_{i,i}$
  - The k-th upper diagonal contains the elements $\mathbf{M}_{i,(i+k)}$
  - The k-th lower diagonal contains the elements $\mathbf{M}_{(i+k),i}$
  - The **anti-diagonals** go rom top-right to bottom-left
- Matrix is **diagonal** if all its non-zero values are in a diagonal (typically main diagonal)
  - **Bi-diagonal** matrices have values in two diagonals, etc.
- Matrix $\mathbf{M}$ is **upper (right) triangular** if all of its non-zeros are in or above the main diagonal
  - **Lower (left) triangular** matrices have all non-zeros in or below main diagonal
  - Upper left and lower right triangular matrices replace diagonal with anti-diagonal
- A square matrix $\mathbf{P}$ is **permutation matrix** if each row and each column of $\mathbf{P}$ has exactly one 1 and rest are 0s
  - If $\mathbf{P}$ is a permutation matrix, $\mathbf{PM}$ is like $\mathbf{M}$ but with permuted order of rows

Orthogonal matrices

- A set $V = \{\mathbf{v}_i\} \subset \mathbb{R}^n$ is **orthogonal** if all vectors in V are mutually orthogonal
  - $\mathbf{v} \cdot \mathbf{u} = 0$ for all $\mathbf{v}, \mathbf{u} \in V$
  - If all vectors in V also have unit norm ($\|\mathbf{v}\|_2 = 1$), V is **orthonormal**
- A square matrix $\mathbf{M}$ is **orthogonal** if its columns are a set of orthonormal vector
  - Then also rows are orthonormal
  - If $\mathbf{M} \in \mathbb{R}^{n \times m}$ and $n > m$, $\mathbf{M}$ can be column-orthogonal, but its rows cannot be orthogonal
- If $\mathbf{M}$ is orthogonal, $\mathbf{M}^T = \mathbf{M}^{-1}$ (i.e. $\mathbf{M}\mathbf{M}^T = \mathbf{M}^T\mathbf{M} = \mathbf{I}_n$)
  - If $\mathbf{M}$ is only column-orthogonal ($n > m$), $\mathbf{M}^T$ is the left inverse ($\mathbf{M}^T\mathbf{M} = \mathbf{I}_m$)
  - If $\mathbf{M}$ is row-orthogonal ($n < m$), $\mathbf{M}^T$ is the right inverse ($\mathbf{M}\mathbf{M}^T = \mathbf{I}_n$)

Suggested reading

- Any (elementary) linear algebra text book
  - For example: Carl Meyer
    Matrix Analysis and Applied Linear Algebra
    Society for Industrial and Applied Mathematics, 2000
    http://www.matrixanalysis.com
- Wolfram MathWorld articles
- Wikipedia articles