Basics of Matrix Vector Algebra

CAP 5415

Basics of matrix vector algebra

- Scalar: x
- row Vector (1 x n) : $\boldsymbol{x} = [x_1 \quad x_2 \quad \cdots \quad x_n]$, Column vector (n x 1) : $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = [x_1 \quad x_2 \quad \cdots \quad x_n]^{\boldsymbol{T}}$
- Matrix (m x n): $\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$ m (rows), n (columns) $\quad \boldsymbol{A}^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}$ n (rows), m (columns)
- Symmetric Matrix: $\boldsymbol{A} = \boldsymbol{A}^T$

Basics of matrix vector algebra

Linear Independence

- $\{\boldsymbol{x}_1, \boldsymbol{x}_2 \cdots, \boldsymbol{x}_M\}$ is a set of linearly independent vectors provided
  - $a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_M\boldsymbol{x}_M = \boldsymbol{0} \Rightarrow a_1 = 0 = a_2 = a_M$

- In other words, none of the vectors can be expressed as a linear combination of the other vectors

- Let $\boldsymbol{x}_i = [x_1^i \quad x_2^i \quad \cdots \quad x_n^i]^T$, and $\boldsymbol{x}_j = [x_1^j \quad x_2^j \quad \cdots \quad x_n^j]^T$

Inner product

- $\boldsymbol{x}_i^T\boldsymbol{x}_j = \sum_k^n x_k^i\, x_k^j$ (scalar number)

Outer product $\boldsymbol{x}_i\boldsymbol{x}_j^T = \begin{bmatrix} x_1^i x_1^j & x_1^i x_2^j & \cdots & x_1^i x_n^j \\ x_2^i x_1^j & x_2^i x_2^j & \cdots & x_2^i x_2^j \\ \vdots & \vdots & \ddots & \vdots \\ x_n^i x_1^j & x_n^i x_2^j & \cdots & x_n^i x_m^j \end{bmatrix}$ (a matrix)

- $\boldsymbol{x}_i^T\boldsymbol{x}_i = \sum_k^n \left(x_k^i\right)^2$ = squared norm of $\boldsymbol{x}_i$

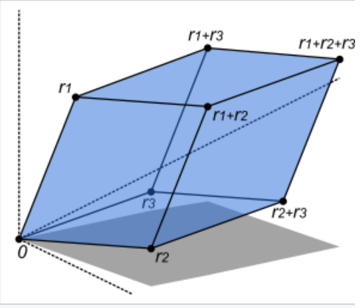
Orthonormality

- From a set of *M* linearly independent vectors, we can always construct *M* ortho-normal vectors.

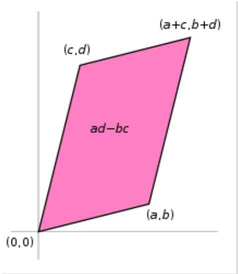
- $\boldsymbol{x}_i^T\boldsymbol{x}_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$

Basics of matrix vector algebra

- Matrix Multiplication –
  - Let $\boldsymbol{a}_i$ denote the *i-th* column of the matrix $\boldsymbol{A}$, and $\boldsymbol{b}_j$ denote the *j-th* column of the matrix $\boldsymbol{B}$.
    - $\boldsymbol{A} = [\boldsymbol{a}_1 \quad \boldsymbol{a}_2 \quad \cdots \quad \boldsymbol{a}_n]$, and $\boldsymbol{B} = [\boldsymbol{b}_1 \quad \boldsymbol{b}_2 \quad \cdots \quad \boldsymbol{b}_{\boldsymbol{m}}]$
  - The product of the two matrices is defined as
  - $\boldsymbol{A}^T\boldsymbol{B} = \begin{bmatrix} \boldsymbol{a}_1^T\boldsymbol{b}_1 & \boldsymbol{a}_1^T\boldsymbol{b}_2 & \cdots & \boldsymbol{a}_1^T\boldsymbol{b}_m \\ \boldsymbol{a}_2^T\boldsymbol{b}_1 & \boldsymbol{a}_2^T\boldsymbol{b}_2 & \cdots & \boldsymbol{a}_2^T\boldsymbol{b}_m \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{a}_n^T\boldsymbol{b}_1 & \boldsymbol{a}_n^T\boldsymbol{b}_2 & \cdots & \boldsymbol{a}_n^T\boldsymbol{b}_m \end{bmatrix}$
- Determinant of a Matrix
  - The determinant of a matrix $\boldsymbol{A}$ is denoted by $|\boldsymbol{A}| = \sum_{i=1}^{K} a_{ij}C_{ij}$ where $C_{ij}$is the cofactor of a_{ij} defined by $C_{ij} = (-1)^{i+1}|\boldsymbol{M}_{ij}|$, and $\boldsymbol{M}_{ij}$ is the minor of matrix $\boldsymbol{A}$ formed by eliminating row i and column j from $\boldsymbol{A}$. (source: http://mathworld.wolfram.com)
  - $|\boldsymbol{AB}| = |\boldsymbol{A}||\boldsymbol{B}|$



The volume of this parallelepiped is the absolute value of the determinant of the matrix formed by the rows constructed from the vectors r1, r2, and r3.



The area of the parallelogram is the absolute value of the determinant of the matrix formed by the vectors representing the parallelogram's sides.

Basics of matrix vector algebra

Matrix Inverse

- The inverse of $\boldsymbol{A}$ is denoted by $\boldsymbol{A}^{-1}$, and satisfies $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix.
  - If $|\boldsymbol{A}|$= 0 implies that $\boldsymbol{A}$ is singular or non-invertible.

Eigen-vectors and Eigen-values

- An eigenvalue and eigen vector of an n x n matrix satisfy $\boldsymbol{A}\boldsymbol{\varphi}_i = \lambda_i \boldsymbol{\varphi}_i$
  - To solve for eigen vectors and eigen values, we solve the equation $|\boldsymbol{A} - \lambda_i \boldsymbol{I}| = 0$
- The determinant of a matrix is the product of its eigen-values, i.e. $|\boldsymbol{A}| = \prod_i^n \lambda_i$
  - If $|\boldsymbol{A}| = 0$, then at least one eigen-value is 0.
- The trace of a matrix is the sum of its diagonal values, ie. $trace\{\boldsymbol{A}\} = \sum_i^n a_{ii} = \sum_{i=1}^n \lambda_i$
- Rank of the matrix is the number of non-zero eigenvalues.
- If $\boldsymbol{A}$ is symmetric, then λ_i are real
- If $\boldsymbol{A}$ is symmetric and $\lambda_i \neq \lambda_j$, then $\boldsymbol{\varphi}_i^T \boldsymbol{\varphi}_j = 0$.
  - Since the eigenvectors can be always normalized to unit norm, the eigen-vectors of a n x n symmetric matrix form a *orthonormal basis set* for an *n-dimensional vector space*
- Condition Number of a matrix = $\frac{\lambda_{max}}{\boldsymbol{\lambda_{min}}}$

Basics of matrix vector algebra

- Positive Definite Matrix
  - $\boldsymbol{A}$ is positive definite implies $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \geq 0$, for all $\boldsymbol{x}$ with the equality holding iff $\boldsymbol{x} = 0$
  - All eigen-values are positive, i.e. $\lambda_i > 0$
    - Positive semi-definite $\lambda_i \geq 0$
    - Negative definite $\lambda_i < 0$
    - Negative semi-definite $\lambda_i \leq 0$

Vector Calculus

- Let $f(\boldsymbol{x})$ be a scalar function of n variables $x_1 \quad x_2 \quad \cdots \quad x_n$ (which are elements of the vector $\boldsymbol{x}$)
- $\nabla_{\boldsymbol{x}} f = \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \frac{\partial f(\boldsymbol{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\boldsymbol{x})}{\partial x_n} \end{bmatrix}$ n x 1 vector
- Now consider a vector function $\boldsymbol{f}(\boldsymbol{x}) = [f_1(\boldsymbol{x}) \quad f_2(\boldsymbol{x}) \quad \cdots \quad f_n(\boldsymbol{x})]^{\boldsymbol{T}}$ of size m x 1
- $\nabla_{\boldsymbol{x}} \boldsymbol{f} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$ m x n matrix

Basics of matrix vector algebra

- Some useful results
  - $\frac{d}{d\boldsymbol{x}}(\boldsymbol{y}^T\boldsymbol{x}) = \frac{d}{d\boldsymbol{x}}(\boldsymbol{x}^T\boldsymbol{y}) = \boldsymbol{y}$
  - $\frac{d}{d\boldsymbol{x}}(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{A}$
  - $\frac{d}{d\boldsymbol{x}}(\boldsymbol{y}^T\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{A}^T\boldsymbol{y}$
  - $\frac{d}{d\boldsymbol{x}}(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}) = (\boldsymbol{A}^T + \boldsymbol{A})\boldsymbol{x}.$
    - If $\boldsymbol{A}$ is symmetric, then $\frac{d}{d\boldsymbol{x}}(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}) = 2\boldsymbol{A}\boldsymbol{x}$

- **Least Squares Solution**
  - Consider $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, where $\boldsymbol{A}$ is a m x n matrix, $\boldsymbol{x}$ is a n x 1 vector, and $\boldsymbol{b}$ is a m x 1 vector.
  - We wish to find $\boldsymbol{x}$ given $\boldsymbol{A}$ and $\boldsymbol{b}$ when m > n (also known as overspecified system of equations)
  - Approach: Define the error vector $\boldsymbol{e} = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}$ and find $\boldsymbol{x}$ to minimize the error norm $J = \boldsymbol{e}^T\boldsymbol{e}$

Basics of matrix vector algebra

- $J = \boldsymbol{e}^T\boldsymbol{e} = (\boldsymbol{Ax} - \boldsymbol{b})^T(\boldsymbol{Ax} - \boldsymbol{b})$
  $= \boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{Ax} - 2\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{b} + \boldsymbol{b}^T\boldsymbol{b}$

- $\nabla_{\boldsymbol{x}} J = 2\boldsymbol{A}^T\boldsymbol{Ax} - 2\boldsymbol{A}^T\boldsymbol{b} = \boldsymbol{0}$

or $\boldsymbol{x} = (\boldsymbol{A}^T\boldsymbol{A})^{-\boldsymbol{1}}\boldsymbol{A}^T\boldsymbol{b}$

$(\boldsymbol{A}^T\boldsymbol{A})^{-\boldsymbol{1}}\boldsymbol{A}^T$ is the pseudo inverse of $\boldsymbol{A}$.

If $\boldsymbol{A}$ is square and invertible, then$(\boldsymbol{A}^T\boldsymbol{A})^{-\boldsymbol{1}}\boldsymbol{A}^T = \boldsymbol{A}^{-1}$

Basics of matrix vector algebra

- Suppose $\boldsymbol{A}$ is a n x n invertible matrix with eigenvalues λ_i and eigenvectors $\varphi_i, i = 1,2, \dots n$
- Assume that all eigenvalues are distinct, and $\boldsymbol{A}$ is symmetric. Then

$$\varphi_i^T \varphi_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

and

$$\boldsymbol{A}\varphi_i = \lambda_i \varphi_i \qquad 1 \leq i \leq n$$

- Modal Matrix
  - Define $\boldsymbol{\Phi} = [\varphi_1 \quad \varphi_2 \quad \cdots \quad \varphi_n]$ - n x n matrix with the eigenvectors φ_i as its columns
  - Note that $\boldsymbol{\Phi}^T \boldsymbol{\Phi} = \mathbf{I}$, and therefore $\boldsymbol{\Phi}^{-1} = \boldsymbol{\Phi}^T$. Furthermore $\boldsymbol{\Phi}\boldsymbol{\Phi}^T = \mathbf{I}$.
- $\boldsymbol{A\Phi} = [\lambda_1\varphi_1 \quad \lambda_2\varphi_2 \quad \cdots \quad \lambda_n\varphi_n] = [\varphi_1 \quad \varphi_2 \quad \cdots \quad \varphi_n] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$
- $\boldsymbol{A\Phi} = \boldsymbol{\Phi\Delta}$, where $\boldsymbol{\Delta} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$ is a diagonal matrix

Basics of matrix vector algebra

- $\boldsymbol{A\Phi} = \boldsymbol{\Phi\Delta}$
  - $\boldsymbol{A} = \boldsymbol{\Phi\Delta\Phi}^T$, and $\boldsymbol{\Delta} = \boldsymbol{\Phi}^T\boldsymbol{A\Phi}$

Therefore, a symmetric positive definite matrix is diagonalized by its the modal matrix (matrix of eigenvectors).

- **Simultaneous diagonalization**
  - Assume that $\boldsymbol{P}$ and $\boldsymbol{Q}$ are n x n real symmetric matrices, and that $\boldsymbol{P}$ is positive definite. Then we can a matrix $\boldsymbol{V}$ such that

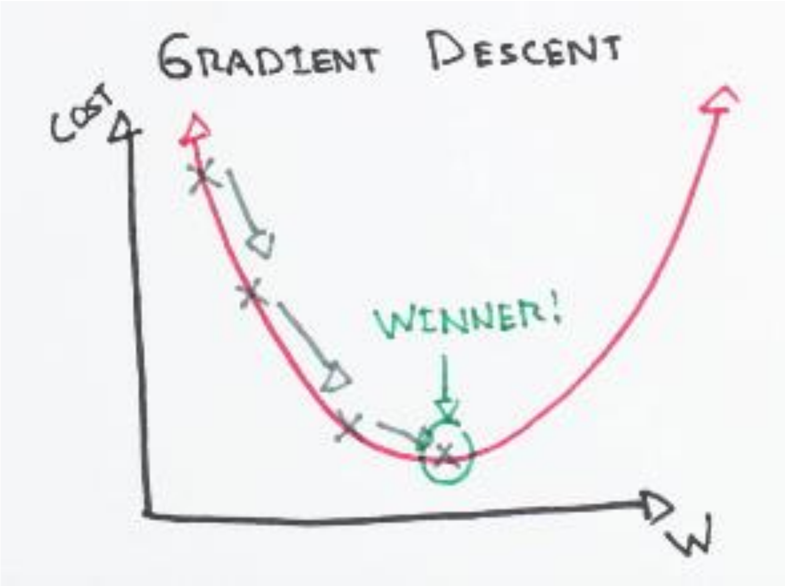
$$\boldsymbol{V}^T\boldsymbol{PV} = \mathbf{I},\ \text{and}$$

$$\boldsymbol{V}^T\boldsymbol{QV} = \boldsymbol{\Delta}\ \text{(a diagonal matrix).}$$

  - These eigen values satisfy $\boldsymbol{Qv}_{\mathrm{i}} = \lambda_i \boldsymbol{P}\ \boldsymbol{v}_{\mathrm{i}}$

Basics of matrix vector algebra

- Let $f(\boldsymbol{x})$ be a scalar function of vector $\boldsymbol{x}$. We want to find the value of $\boldsymbol{x}$ which minimizes $f(\boldsymbol{x})$.
  - Steepest Descent method is
    1. Start with an initial guess $\boldsymbol{x}_0$
    2. Compute the gradient at $\boldsymbol{x}_0$, i.e. $\frac{df(\boldsymbol{x}_0)}{d\boldsymbol{x}}$
    3. Determine $\boldsymbol{x}_k = \boldsymbol{x}_{k-1} - \mu\frac{df(\boldsymbol{x}_{k-1})}{d\boldsymbol{x}}$, where μ is the step size parameter
    4. Iterate until $\boldsymbol{x}_k$ is relatively stable



https://ml-cheatsheet.readthedocs.io/en/latest/gradient_descent.html

Iterative Solution to Least Squares problem $\boldsymbol{Ax} = \boldsymbol{b}$

- Error $\boldsymbol{e} = \boldsymbol{Ax} - \boldsymbol{b}$
- Find $\boldsymbol{x}$ such that $J(\boldsymbol{x}) = \boldsymbol{e}^T\boldsymbol{e} = (\boldsymbol{Ax} - \boldsymbol{b})^T(\boldsymbol{Ax} - \boldsymbol{b})$ is minimum
- $J(\boldsymbol{x})=\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{Ax} - 2\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{b} + \boldsymbol{b}^T\boldsymbol{b}$
- $\nabla_{\boldsymbol{x}}J = 2\boldsymbol{A}^T\boldsymbol{Ax} - 2\boldsymbol{A}^T\boldsymbol{b} = 2\boldsymbol{A}^T(\boldsymbol{Ax} - \boldsymbol{b}) = 2\boldsymbol{A}^T\boldsymbol{e}$

$$\boldsymbol{x}_k = \boldsymbol{x}_{k-1} - \mu\nabla_{\boldsymbol{x}}J = \boldsymbol{x}_{k-1} - \mu\boldsymbol{A}^T\boldsymbol{e}_{k-1}$$