Linear Algebra Cheat Sheet

Yiping Lu

January 2024

1 Projection and Least Square

Approximately solve $Ax = b$ equals to project b to col(A)

- $x = (A^\top A)^{-1}A^\top b$
- Projection Matrix $P = A(A^\top A)^{-1}A^\top$ (Projection of b to col(A) is Ax, $i.e.Pb = Ax$)
- rank($A^\top A$) = rank(A) so if A is full column rank (or columns of A is a basis of col(A)) then $A^\top A$ is invertible
- application: best linear fit

Orthogonal Vectors $\{q_1, \cdots, q_n\}_{i=1}^n$ satisfies

- $q_i^\top q_i = 1$
- $q_i^\top q_j = 0$ for $i \neq j$

$\{q_1, \cdots, q_n\}_{i=1}^n$ are orthogonal vectors then $q_1, \cdots, q_n$ are linear independent

Let $Q = [q_1, \cdots, q_n]$, then $Q^\top Q = I$. $QQ^\top$ may not be I but it's the projection matrix to Q.

Furthermore if Q is a square matrix , then we call Q is a orthonogal matrix. For orthogonal matrix we have $Q^\top Q = QQ^\top = I$ and $Q^\top = Q^{-1}$

Project b to Q is $QQ^\top b = (q_1^\top b)q_1 + (q_2^\top b)q_2 + \cdots + (q_n^\top b)q_n$. ($Q^\top b = \begin{bmatrix} q_1^\top b \\ q_2^\top b \\ \cdots \\ q_n^b \end{bmatrix}$)

G-S The Gram-Schmidt process is a method used in linear algebra for orthogonalizing a set of vectors. Given a set of linearly independent vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ in $\mathbb{R}^n$, the Gram-Schmidt process produces a set of orthogonal vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ as follows:

1. Set $\mathbf{u}_1 = \mathbf{v}_1$.

2. For $i = 2$ to k, compute

$$\mathbf{u}_i = \mathbf{v}_i - \sum_{j=1}^{i-1} \frac{\langle \mathbf{v}_i, \mathbf{u}_j \rangle}{\langle \mathbf{u}_j, \mathbf{u}_j \rangle} \mathbf{u}_j$$

where $\langle \mathbf{x}, \mathbf{y} \rangle$ denotes the dot product of vectors $\mathbf{x}$ and $\mathbf{y}$.

To normalize each vector to get an orthonormal set, compute

$$\mathbf{e}_i = \frac{\mathbf{u}_i}{\|\mathbf{u}_i\|}$$

for $i = 1$ to k, where $\|\mathbf{u}_i\|$ is the Euclidean norm of $\mathbf{u}_i$.

This process orthogonalizes the set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$, ensuring that each $\mathbf{u}_i$ is orthogonal to all previous $\mathbf{u}_j$ for $j < i$. The last step, normalization, is optional if the goal is to obtain an orthonormal set.

QR**-decompostion**: $A = QR$, Q is orthogonal (but not square), R is upper triangular

- Using G-S to calculate Q
- caluclate R by $Q^\top A$

2 Determinate

Four Basic Property

- (1) $\det([v_1, \cdots, v_i, \cdots, v_j, \cdots, v_n]) = -\det([v_1, \cdots, v_j, \cdots, v_i, \cdots, v_n])$ (switching i, j column make the determinate negative)
- (2) $\det(AB) = \det(A)\det(B)$
- (linear combination of single column)
  - (3) $\det([v_1+v_1', v_2, \cdots, v_n]) = \det([v_1, v_2, \cdots, v_n])+\det([v_1', v_2, \cdots, v_n])$
  - (4) $\det([cv_1, v_2, \cdots, v_n]) = c\det([v_1, v_2, \cdots, v_n])$

determinates of basic matrix

- $\det(I_n) = I$
- What is the determinate of a diagonal matrix?
- What is the determinate of an orthogonal matrix matrix?
- What is the determinate of a permutation matrix?
- What is the determinate of an elimination matrix?
- What is the determinate of a lower diagonal matrix?

Try to prove the following property:

- $\det([\vec{0}, v_1, v_2, \cdots, v_{n-1}]) = 0$
- $\det([c_1 v_1, c_2 v_2, \cdots, v_n]) = c_1 c_2 \det([v_1, v_2, \cdots, v_n])$
- $\det([v_1, v_1, v_2, \cdots, v_{n-1}]) = 0$ (A have two equal columns, then determiante is 0)
- $\det(cA) = c^n \det(A)$
- $\det(A) = \det(A^T)$ (*hint*:LU Decomposition)
- $\det(A^{-1}) = \frac{1}{\det(A)}$

Cofactor : $C_{ij} = (-1)^{i+j} M_{ij}$ where M_{ij} is the minor (determinate of $(n-1) \times (n-1)$ submatrix)

- Cofactor expansion
  - For rows $\det(A) = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}$
  - For columns $\det(A) = a_{1i}C_{1i} + a_{2i}C_{2i} + \cdots + a_{ni}C_{ni}$
- **Cramer Rule** Consider a system of linear equations in matrix form $Ax = b$, where:
  - A is an $n \times n$ matrix of coefficients.
  - x is a column vector of unknowns $x_1, x_2, \ldots, x_n$.
  - b is a column vector of constants.

  Cramer's Rule states that if $\det(A) \neq 0$, then the system has a unique solution. The solution for each unknown x_i is given by:

  $$x_i = \frac{\det(A_i)}{\det(A)}$$

  where:
  - $\det(A)$ is the determinant of the matrix A.
  - A_i is the matrix formed by replacing the i-th column of A with the column vector b.
- $A^{-1} = \frac{1}{\det(A)} C^\top$, where C is the matrix of cofactors
- **Hard Questions**: What is $\det(C)$? ($\frac{1}{\det(A)^{n+1}}$, why?)

3 Eigenvaule

Eigenvalue and eigenvectors means

$$Ax = \lambda x$$

$\lambda \in \mathbb{R}$ is eigen value and $x \in \mathbb{R}^n$ is the eigne vector.

- λ is the solution of $p(\lambda) = \det(A - \lambda I) = 0$
- $p(\lambda)$ is a $n-$the order polynomial (by cofactor expansion)
- $x \in \text{Nul}(A - \lambda I)$

$$p(\lambda) = (-1)^n(\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n)$$

- $n \times n$ matrix have n eigen values

  ($n \times n$ matrix have n don't means it has n eigenvectors, if it has it can be diagonalize (this means not every matrix can be diagonalize), example $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, $p(\lambda) = -\lambda^2$ it has two eigenvalue 0 but only one egienvector. All matrix can't be diagnoalized will similar to this.)

If $\lambda_1, \cdots, \lambda_n$ are distinct eigenvalues then eigenvectors $x_1, \cdots, x_n$ are linear independent which mens $X = [x_1, \cdots, x_n]$ are invertible.

Diagnoalization $A = X\Lambda X^{-1}$ where Λ is a diagnoal matrix, diagnoal elements are eigen values.

Similar Matrix $A = XBX^{-1}$

- $\text{trace}(A) = \text{trace}(B) = \lambda_1 + \cdots + \lambda_n$
- $\det(A) = \det(B) = \lambda_1 \cdots \lambda_n$
- all A's eigenvectors are $[\vec{a}_1, \vec{a}_2, \cdots, \vec{a}_n]$, B's eigenvectors are $[\vec{b}_1, \vec{b}_2, \cdots, \vec{b}_n]$, then

$$X = [\vec{a}_1, \vec{a}_2, \cdots, \vec{a}_n][\vec{b}_1, \vec{b}_2, \cdots, \vec{b}_n]^{-1}$$

Symmetric Matrix $A = A^\top$

- eigenvectors $x_1, x_2, \cdots, x_n$ are orthogonal
- $A = Q\Lambda Q^\top$, Q is orthogonal matrix (remember to normalize eigenvectors to unit vectors)
- $A = \sum_i \lambda_i \underbrace{x_i x_i^\top}_{\text{projection to } x_i}$ (we need x_i are unit vector)
- $\vec{x}^\top A \vec{x} = \sum_i \lambda_i (x_i^\top x)^2$

- Motivation: $\vec{x}^\top A\vec{x}$ for $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix}$ is a qudratic function respect to $x_1, \cdots, x_n$

- **Positive Definite Matrix**: Symmetric matrix whose eigenvalue is positive ($\lambda_i > 0$ for all i)

  - $\vec{x}^\top A\vec{x}$ is always positive ($\vec{x}^\top A\vec{x}, \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix}$ is a quadratic function respect to variable $x_1, \cdots, x_n$)
  - $\underbrace{\vec{x}^\top A\vec{x}}_{\text{quadratic function who is always positive}} = \underbrace{\sum_i \lambda_i (x_i^\top x)^2}_{\text{sum of positive square functions}}$

4 Singular Value Decomposition

$$A = \underbrace{U}_{\text{orthogonal}} \underbrace{\Sigma}_{\text{diag}} \underbrace{V^\top}_{\text{orthogonal}}$$

- $A \in \mathbb{R}^{m\times n}$
- $U \in \mathbb{R}^{m\times m}, V \in \mathbb{R}^{n\times n}$
- $\Sigma \in \mathbb{R}^{m\times n}$

$$AA^\top = U\Sigma\Sigma^\top U^\top, A^\top A = V\Sigma^\top \Sigma V^\top$$

- $U = [u_1, \cdots, u_n]$ then $u_1, \cdots, u_n$ are eigenvectors of $AA^\top$
- $V = [v_1, \cdots, v_n]$ then $v_1, \cdots, v_n$ are eigenvectors of $A^\top A$

  $$A = \sigma_1 u_1 v_1^\top + \sigma_2 u_2 v_2^\top + \cdots + \sigma_r u_r v_r^\top \quad \text{where } r = \text{rank}(A)$$

- $u_1 = \frac{1}{\sigma_1} A v_1$, $v_1 = \frac{1}{\sigma_1} A^\top u_1$

We have

$$\text{rank}(A) = \text{rank}(AA^\top) = \text{rank}(A^\top A)$$

$\mathbf{u}_1, \ldots, \mathbf{u}_r$	is an orthonormal basis for the column space
$\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m$	is an orthonormal basis for the left nullspace $\mathcal{N}(\mathbf{A}^T)$
$\mathbf{v}_1, \ldots, \mathbf{v}_r$	is an orthonormal basis for the row space
$\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n$	is an orthonormal basis for the nullspace $\mathcal{N}(\mathbf{A})$.

5 Linear Transform

- check linear transform
  - Prove: check $T(cx) = cT(x), T(x+y) = T(x) + T(y)$
  - Disprove: counter example
  - $\mathbb{P}_n$ polynomial of degree n.
- Change of Basis: `https://2prime.github.io/files/linear/recitation11_sol.pdf`