Application: least square lines

Data $= \{x_i, y_i\}_{i=1}^{N}$

Model : $y_i = \beta_1 + \beta_2 x_i = f_\theta(x_i)$, $\theta = \{\beta_1, \beta_2\}$

loss : $\frac{1}{N}\sum_{i=1}^{N} (y_i - f_\theta(x_i))^2$

$\Rightarrow$ Matrix form

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$$

$\rightarrow S(\beta) = \|\vec{y} - X\vec{\beta}\|^2$

$$\nabla_\beta (\vec{y} - X\vec{\beta})^T(\vec{y} - X\vec{\beta}) = \nabla_\beta \left(y^Ty - \beta^TX^Ty - y^TX\beta + \beta^TX^TX\beta \right)$$

$$= -2X^Ty + 2X^TX\beta$$

$$= 0$$

$$\therefore \hat{\beta} = (X^TX)^{-1}X^Ty$$

Example) $\{(2,1), (5,2), (7,3), (8,3)\}$

$$\Rightarrow X^TX = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}$$

$$X^T\vec{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

Solving $\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \hat{\beta} = \begin{bmatrix} 9 \\ 57 \end{bmatrix} \rightarrow \therefore \hat{\beta} = \begin{bmatrix} \frac{2}{7} \\ \frac{5}{14} \end{bmatrix}$ $\quad \therefore y = \frac{2}{7} + \frac{5}{14}x$

Example) $y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2$ with parameter $\{\beta_1, \beta_2, \beta_3\}$

$$\Rightarrow \underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \end{bmatrix}}_{\text{Design matrix } X} \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}}_{\text{param}} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \end{bmatrix}}_{\text{Observation vector } \vec{y}}$$

$\rightarrow X\hat{\beta} \cong \vec{y}$, $\hat{\beta} = (X^TX)^{-1}X\vec{y}$

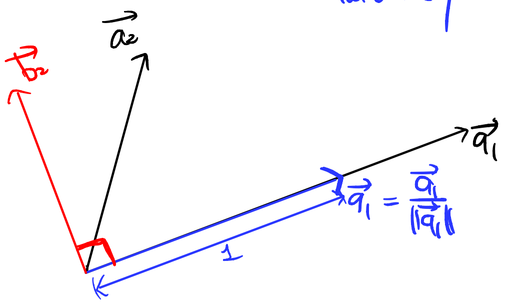
* Gram–Schmidt.

Given a basis $\{\vec{a}_1, \dots, \vec{a}_n\}$, produce an orthonormal basis $\{\vec{q}_1, \dots, \vec{q}_n\}$

- $\vec{b}_1 = \vec{a}_1$, $\vec{q}_1 = \dfrac{\vec{b}_1}{\|\vec{b}_1\|}$
- $\vec{b}_2 = \vec{a}_2 - \langle \vec{a}_2, \vec{q}_1 \rangle \vec{q}_1$, $\vec{q}_2 = \dfrac{\vec{b}_2}{\|\vec{b}_2\|}$
- $\vec{b}_3 = \vec{a}_3 - \langle \vec{a}_3, \vec{q}_1 \rangle \vec{q}_1 - \langle \vec{a}_3, \vec{q}_2 \rangle \vec{q}_2$, $\vec{q}_3 = \dfrac{\vec{b}_3}{\|\vec{b}_3\|}$

$\vdots$

- $\vec{b}_i = \vec{a}_i - \underbrace{\sum_{j=1}^{i-1} \langle \vec{a}_i, \vec{q}_j \rangle \vec{q}_j}_{\text{take away}}$, $\vec{q}_i = \dfrac{\vec{b}_i}{\|\vec{b}_i\|}$



The QR decomposition

Let A be an $m \times n$ matrix of rank n (columns are independent)

Then, we have the QR decomposition $A = QR$

where $Q \in \mathbb{R}^{m \times n}$ with orthonormal columns, and R is upper-triangular, $n \times n$, invertible.

Example) Find the QR decomposition of $A = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 0 & 5 \\ 0 & 3 & 6 \end{bmatrix}$

$\Rightarrow$ First, GS on columns of A

$\vec{b}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\vec{q}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$

$\vec{b}_2 = \begin{bmatrix} 2 \\ 0 \\ 3 \end{bmatrix} - \left\langle \begin{bmatrix} 2 \\ 0 \\ 3 \end{bmatrix}, \vec{q}_1 \right\rangle \vec{q}_1 = \begin{bmatrix} 2 \\ 0 \\ 3 \end{bmatrix} - \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}$, $\vec{q}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

$\vec{b}_3 = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} - \left\langle \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}, \vec{q}_1 \right\rangle \vec{q}_1 - \left\langle \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}, \vec{q}_2 \right\rangle \vec{q}_2 = \begin{bmatrix} 0 \\ 5 \\ 0 \end{bmatrix}$, $\vec{q}_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

Hence $Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$

To find R in $A = QR \rightarrow R = Q^T A = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 3 & 6 \\ 0 & 0 & 5 \end{bmatrix}$.

$$\therefore \underset{A}{\begin{bmatrix} 1 & 2 & 4 \\ 0 & 0 & 5 \\ 0 & 3 & 6 \end{bmatrix}} = \underset{Q}{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}} \underset{R}{\begin{bmatrix} 1 & 2 & 4 \\ 0 & 3 & 6 \\ 0 & 0 & 5 \end{bmatrix}}$$

Note that $R = Q^T A = \begin{bmatrix} \vec{q}_1^T \vec{a}_1 & \vec{q}_1^T \vec{a}_2 & q_1^T a_3 & \cdots \\ 0 & q_2^T a_2 & q_2^T a_3 & \\ 0 & 0 & q_3^T a_3 & \\ \vdots & \vdots & & \ddots \end{bmatrix}$

The QR decomposition is useful for solving LS problems.

$$\text{LS}: A^T A \hat{x} = A^T \vec{b} \iff (QR)^T (QR) \hat{x} = (QR)^T \vec{b}$$

$$\downarrow R^T \underbrace{Q^T Q}_{I} R = R^T R$$

$$\iff R^T R \hat{x} = R^T Q^T \vec{b}$$

$$\iff R \hat{x} = Q^T \vec{b} \quad \text{— } (*)$$

Since R is upper triangular, $(*)$ can be solved by back substitution.

To summarize

$\hat{x}$ is a least square solution of $A\vec{x} = \vec{b}$

$\iff R\hat{x} = Q^T \vec{b}$ (where $A = QR$)

- Determinants

Definition) The determinant is characterized by

- the normalization $\det I = 1$ — Gauss Jordan elimination.
- and how it is affected by elementary row operation.
  - (replacement) $R_1 \leftarrow R_1 + 2R_2$
    → does not change the Det.
  - (interchange) $R_1 \leftrightarrow R_2$
    → Reverse the sign
  - (scaling) $R_2 \leftarrow s \cdot R_2$
    → Multiples the Det with s

Example) compute $\begin{vmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 7 \end{vmatrix}$.

$$\Rightarrow \begin{vmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 7 \end{vmatrix} \xrightarrow{R_2 \leftarrow \frac{1}{2} R_2} 2 \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 7 \end{vmatrix} \xrightarrow{R_3 \leftarrow \frac{1}{7} R_3} 14 \underbrace{\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}}_{=1} = 14$$

Example) Compute $\begin{vmatrix} 1 & 2 & 3 \\ 0 & 2 & 4 \\ 0 & 0 & 7 \end{vmatrix}$

$$\Rightarrow \begin{vmatrix} 1 & 2 & 3 \\ 0 & 2 & 4 \\ 0 & 0 & 7 \end{vmatrix} \xrightarrow{R2 \leftarrow \frac{1}{2}R2} 2\begin{vmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 7 \end{vmatrix} \xrightarrow{R3 \leftarrow \frac{1}{7}R3} \begin{vmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{vmatrix} \cdot 14$$

$$\xrightarrow[R1 \leftarrow R1 - 3R3]{R2 \leftarrow R2 - R3} 14 \cdot \begin{vmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} \xrightarrow{R1 \leftarrow R1 - 2R2} 14 \cdot \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 14$$

* The determinant of a triangular matrix is the product of diagonal entries!

Example) Compute $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$.

$$\rightarrow \begin{vmatrix} a & b \\ c & d \end{vmatrix} \xrightarrow{R2 \leftarrow R2 - \frac{c}{a}R1} \begin{vmatrix} a & b \\ 0 & d - \frac{c}{a}b \end{vmatrix} = \text{"}\underline{ad - bc}\text{"}$$

* $\det(A) = 0 \iff A$ is not invertible.

$\det(AB) = \det(A)\det(B)$

$\det(A^{-1}) = \dfrac{1}{\det(A)}$

$\det(A^T) = \det(A)$

- Eigen vectors & Eigen values

Definition) An eigenvector of $A \in \mathbb{R}^{n \times n}$ is a nonzero $\vec{x}$ such that

$$A\vec{x} = \lambda\vec{x} \text{ for some scalar } \lambda$$

The λ is the corresponding eigenvalue.

How to solve $A\vec{x} = \lambda\vec{x}$

$\iff A\vec{x} - \lambda\vec{x} = \vec{0}$

$\iff (A - \lambda I)\vec{x} = \vec{0}$

$\iff (A - \lambda I)$ is not invertible

$\iff \underline{\det(A - \lambda I)} = 0$

└ characteristic polynomial.

Example) Find the eigenvectors and eigenvalues of A.

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}.$$

$$\Rightarrow \left| A - \lambda I \right| = \begin{vmatrix} 3-\lambda & 1 \\ 1 & 3-\lambda \end{vmatrix} = (3-\lambda)^2 - 1 = 0$$

$$\rightarrow \lambda^2 - 6\lambda + 8 = (\lambda-2)(\lambda-4) = 0, \quad \lambda_1 = 2, \quad \lambda_2 = 4.$$

i) eigenvector of $\lambda_1 = 2$

$$(A - 2I)\vec{x_1} = \vec{0} \rightarrow \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \vec{x_1} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \vec{x_1} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

· Note that all multiples of $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ also becomes eigenvectors corresponding to $\lambda_1 = 2$.

↳ eigenspace

ii) e.$\vec{v}$ of $\lambda_2 = 4$

$$(A - 4I)\vec{x_2} = \vec{0} \rightarrow \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \vec{x_2} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \vec{x_2} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

· The eigenspace of $\lambda_2 = 4$ is $\text{span}\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$.

Theorem) If $\vec{x_1}, \dots, \vec{x_m}$ are eigenvectors of A corresponding to different eigenvalues, then they are independent.

⇒ (proof by contradiction) Suppose $\vec{x_1}, \dots, \vec{x_m}$ are dependent.
By excluding some $\vec{x_i}$, we may assume that there is upto multiples of on $\vec{c}$ s.t.,

$$c_1\vec{x_1} + \dots + c_m\vec{x_m} = \vec{0} \quad \text{——} \ (*)$$

However,

$$A(c_1\vec{x_1} + \dots + c_m\vec{x_m})$$
$$= c_1\lambda_1\vec{x_1} + \dots + c_m\lambda_m\vec{x_m} = \vec{0}$$

give another $\{c_1\lambda_1, \dots, c_m\lambda_m\}$ that makes (*) true.
Hence, $\vec{x_1}, \dots, \vec{x_m}$ are independent.

★ Some meaning of eigen vectors.

· Consider a linear map $T: \mathbb{R}^n \to \mathbb{R}^n \Rightarrow A \in \mathbb{R}^{n\times n}$.
then for vector $\vec{v} \in \mathbb{R}^n$, this transformation T only scales up the vector. i.e., $T(\vec{v}) = \lambda\vec{v}$, and the direction is not changed!

$$\vec{x} = c_1\vec{v_1} + c_2\vec{v_2} \cdots$$

$$T(\vec{x}) = T(c_1\vec{v_1}) + T(c_2\vec{v_2}) + \cdots$$

$$= c_1\cdot\lambda_1\cdot\vec{v_1} + c_2\cdot\lambda_2\vec{v_2} + \cdots$$

$$c_1\overset{x_1}{\begin{bmatrix}1\\0\\0\\0\\0\end{bmatrix}} \quad c_2\overset{x_2}{\begin{bmatrix}0\\1\\0\\0\\0\end{bmatrix}} \quad c_3\overset{x_3}{\begin{bmatrix}0\\0\\1\\0\\0\end{bmatrix}} \quad c_4\overset{6}{\begin{bmatrix}3\\0\\0\\0\\0\end{bmatrix}} \quad c_5\overset{x_5}{\begin{bmatrix}5\\0\\0\\0\\0\end{bmatrix}} = \begin{bmatrix}4\\0\\0\\0\\0\end{bmatrix}$$

$$\vec{c} = \begin{bmatrix}c_1\\ \vdots\\ c_5\end{bmatrix} \rightarrow \text{basis가 안돼.}$$

$$\vec{c} = \begin{bmatrix}c_1\\ c_2\\ c_3\\ c_5\end{bmatrix} \rightarrow \text{basis가 돼.}$$