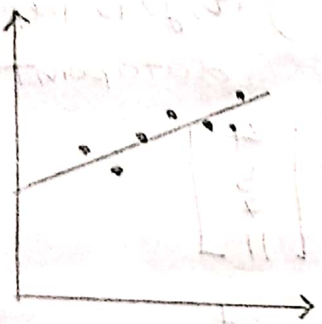
19-07-20

Lecture 14:

Least square regression line:

→ The best fit line from datapoints using linear algebra.



→ We find it by minimizing the sum of the squares of the differences of a data-point with the regression line. That is why it is called "least square" regression line.

Application:

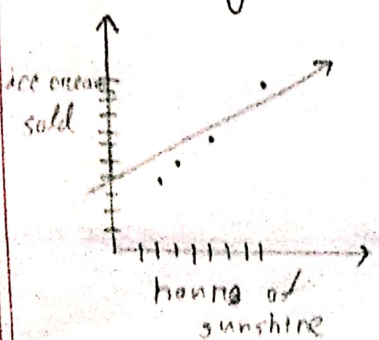
(*) Core concept of machine learning. Used to predict data. (future data).

Example:

Data-points	hours of sunshine	ice-cream sold
(2,4), (3,5), (5,7), (8,11)	2	4
	3	5
	5	7
	8	11

Suppose, 6 hours of sunshine tomorrow. How can we predict ice-cream being sold tomorrow?

→ Using least square regression line.



$Y = c + Dt$

$(2, 3, 5, 8)$

$\boxed{A^T A \hat{x} = A^T b}$ (learnt previously)

Corresponding equations:

$$\left.\begin{aligned} C+2D&=4 \\ C+3D&=5 \\ C+5D&=7 \\ C+8D&=11 \end{aligned}\right\}$$ Derived from 4 different values of t and y. (t,y) is replaced by each datapoint

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 5 \\ 1 & 8 \end{bmatrix} \text{ and } b = \begin{bmatrix} 4 \\ 5 \\ 7 \\ 11 \end{bmatrix}$$

Solving $A^TA\hat{x} = A^Tb$, we get,

$$y = 1.518x + 0.305 \text{ (approximately)}$$

For tomorrow, $x = 6$,

$$\text{So, } y = 1.1518 \times 6 + 0.305$$
$$= 9 + 0.305$$
$$y \approx 9$$

Around 9 ice-creams can be sold tomorrow.

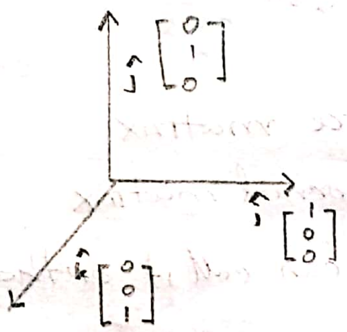
Orthogonal matrices and Gram-Schimdtt process:

A^TA would be invertible if A has $\boxed{\text{independent columns}}$

columns will be independent if they are perpendicular and unit vectors

has magnitude of 1.

Ex:



Orthonormal vectors

← These are unit vectors and perpendicular

If unit vectors and perpendicular then it is orthonormal vector

Defined as $q_1, q_2, q_3 \dots q_n$

Orthonormal basis/vectors are expressed by 'q'.

$$q_i^T q_j = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$$

property of orthonormal vectors

We put the orthrormal vectors in a matrix and get a orthonormal matrix.

$$Q = \begin{bmatrix} \uparrow & \uparrow & \uparrow & & \uparrow \\ q_1 & q_2 & q_3 & \cdots & q_n \\ \downarrow & \downarrow & \downarrow & & \downarrow \end{bmatrix}$$

If Q is a square matrix then it is called orthogonal matrix.

Ex:

$$Q_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \qquad Q_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Orthonormal Matrix   Orthogonal Matrix

Another example of orthogonal matrices are permutation matrices.

$$P = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

→ Square matrix
→ Orthonormal matrix

Hence, we can call it orthogonal matrix

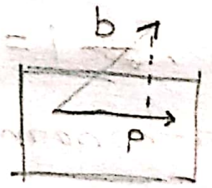
$$Q^TQ = I$$

$$\Rightarrow Q^T = Q^{-1}I$$

$$\Rightarrow \boxed{Q^T = Q^{-1}}$$

→ Transposing will result in inverse of the matrix.

Relation with projections:



Q

projection

$$P = A(A^TA)^{-1}A^T = Q(Q^TQ)^{-1}Q^T$$ [∵ Column space is created by Q]

$$\cancel{= Q(Q^TQ}$$

$$= Q(I)^{-1}Q^T$$

$$= QQ^T$$

$$= I$$ (only when Q is square matrix)

Projection $P = Pb$

$= Ib$

$\therefore P = b$

What does this mean?

→ Whenever we get a column space created by Q, it actually covers the whole space.

But why?

→ Because of the basis of Q are orthonormal vectors and they are independent. So, no vector would be excluded from the column space.

So, b will always be in the column space of Q.

Using our previous equation,

$$A^T A \hat{x} = A^T b$$

$$\Rightarrow Q^T Q \hat{x} = Q^T b$$

$$\Rightarrow I \hat{x} = Q^T b$$

$$\Rightarrow \boxed{\hat{x} = Q^T b}$$

And, $P = Q\hat{x} = A\hat{x}$

$$= \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ q_1 & q_2 & \cdots & q_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} q_1^T b \\ q_2^T b \\ q_n^T b \end{bmatrix}$$

$$= q_1(q_1^T b) + q_2(q_2^T b) + \cdots\cdots q_n(q_n^T b)$$

Gram-Schmidt process:

→ Process to create orthnormal vectors

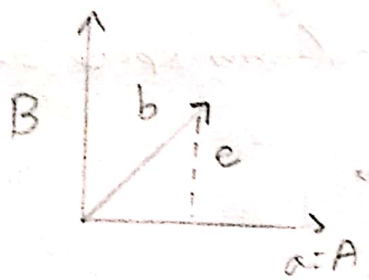
→ Any independent vectors can be transformed to orthonormal vectors using this process.

If you have 3 orthogonal vectors then

i) construct 3 orthogonal vec - A, B, C

ii) dividing by their own values

For 2 independent vectors



We know,

$$\hat{x} = \frac{a^T b}{a^T a}$$

And, $e = b - A\hat{x}$

$$= b - A \cdot \frac{A^T b}{A^T A}$$

Now ~~Bst~~, $B = b - A\frac{A^T b}{A^T A}$ [B would be same as e] ——— (i)

A and B has to be ~~or~~ ~~d~~ orthogonal or dot product is 0.

So, $A^T B = 0$ ——— (ii)

Combining (i) and (ii), we get,

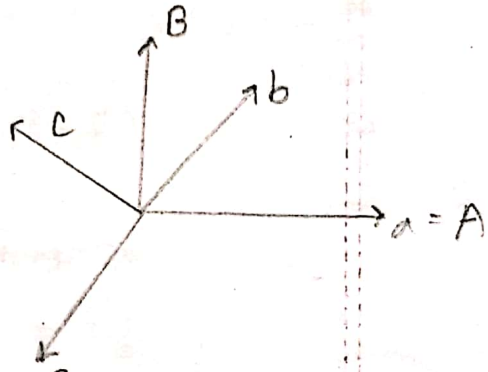
$$A^TB = A^T\left(b - \frac{A^Tb}{A^TA}A\right)$$

$$= \frac{(A^TA)(A^Tb) - (A^Tb)(A^TA)}{A^TA}$$

$$= 0$$

For the 3 dimensional case, there is another independent vector.

Then, $C = c - \frac{A^Tc}{A^TA}A - \frac{B^Tc}{B^TB}B$



	A	B	C
Orthogonal vectors	a	$b - \frac{A^Tb}{A^TA}A$	$c - \frac{A^Tc}{A^TA}A - \frac{B^Tc}{B^TB}B$
Orthonormal vectors	$\frac{A}{\lVert A\rVert}$	$\frac{B}{\lVert B\rVert}$	$\frac{C}{\lVert C\rVert}$

Problems: Given, $a = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$, $b = \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix}$, $c = \begin{bmatrix} 3 \\ -3 \\ 3 \end{bmatrix}$

Find the orthogonal matrix Q, and q_1, q_2, q_3

Ans: (i) $A = a = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$

$$B = b - \frac{A^T b}{A^T A} A$$

$$= \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix} - \frac{\begin{bmatrix} 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix}}{\begin{bmatrix} 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}} \cdot \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix} - \frac{2}{2} \cdot \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix} - \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$$

$$C = c - \frac{A^T c}{A^T A} A - \frac{B^T c}{B^T B} B$$

$$= \begin{bmatrix} 3 \\ -3 \\ 3 \end{bmatrix} - \frac{\begin{bmatrix} 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ -3 \\ 3 \end{bmatrix}}{\begin{bmatrix} 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} -$$

$$\frac{\begin{bmatrix} 1 & -1 & -2 \end{bmatrix} \begin{bmatrix} 3 \\ -3 \\ 2 \end{bmatrix}}{\begin{bmatrix} 1 & -1 & -2 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix}} \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$$

$$= \begin{bmatrix} 3 \\ -3 \\ 3 \end{bmatrix} - \frac{6}{2}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} - \frac{(-6)}{6}\begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix}$$

$$= \begin{bmatrix} 3 \\ -3 \\ 3 \end{bmatrix} - \begin{bmatrix} 3 \\ -3 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

(ii) Now, $A^T B = [1 \ -1 \ 0]\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} = 1 - 1 + 0 = 0$

$\therefore A^T B = 0$ (Checked)

(iii) Next we find orthonormal vectors

$$q_1 = \frac{A}{\|A\|} = \frac{1}{\sqrt{1^2 + (-1)^2 + 0}} \cdot \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{bmatrix}$$

$$q_2 = \frac{B}{\|B\|} = \frac{1}{\sqrt{1^2 + 1^2 + (-2)^2}} \cdot \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{bmatrix}$$

$$q_3 = \frac{C}{\|C\|} = \frac{1}{\sqrt{1^2 + 1^2 + 1^2}} \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$$

(iv) So, orthogonal matrix is

$$Q = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ -1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ 0 & -2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix}$$ (Ans.)