1.a

بردار های x_1 و x_2 را به صورت زیر تعریف می‌کنیم:

$$\forall x=\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix} \longrightarrow x_1=\begin{bmatrix}|x_1|\\ \vdots\\ |x_n|\end{bmatrix}$$

$$x_2=\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix}_{n\times 1} \xrightarrow[\text{شوارتز}]{\text{کوشی}} x_1 \cdot x_2 \leqslant \|x_1\|_2 \, \|x_2\|_2$$

$$\underbrace{\sum_{i=1}^{n}|x_i|}_{\|x\|_1} \leqslant \underbrace{\|x_1\|_2}_{\|x\|_2} \times \sqrt{n} \longrightarrow \|x\|_1 \leqslant \sqrt{n}\,\|x\|_2$$

1.b

Holder inequality: $$\sum_{i=1}^{n}|a_i||b_i| \leqslant \left(\sum_{i=1}^{n}|a_i|^r\right)^{\frac{1}{r}} \left(\sum_{i=1}^{n}|b_i|^{\frac{r}{r-1}}\right)^{1-\frac{1}{r}}$$

$$\xrightarrow{|a_i|=|x_i|^p\ ,\ |b_i|=1\ ,\ r=\frac{q}{p}} \sum_{i=1}^{n}|x_i|^p \leqslant \left(\sum_{i=1}^{n}|x_i|^q\right)^{\frac{p}{q}} \Big(\underbrace{\sum_{i=1}^{n}1}_{n}\Big)^{\frac{q-p}{q}}$$

$$\xrightarrow{\text{طرفین را به توان } \frac{1}{p} \text{ می‌رسانیم}} \underbrace{\left(\sum_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}}}_{\|x\|_p} \leqslant \underbrace{\left(\sum_{i=1}^{n}|x_i|^q\right)^{\frac{1}{q}}}_{\|x\|_q} \times \underbrace{n^{\frac{q-p}{pq}}}_{C}$$

بنابراین نشان دادیم که ثابتی وجود دارد که نابرابری برقرار باشد.

2.a

اگر قرار باشد $x_{\text{min-norm}}$ وجود داشته باشد باید AA^T وارون پذیر باشد. اثبات می‌کنیم اگر $\underbrace{\text{ماتریس } A \text{ سطری full rank باشد}}_{rank(A)=n}$ آنگاه $\underbrace{AA^T \text{ وارون پذیر است}}_{rank(AA^T)=n}$ و برعکس. کافی است اثبات کنیم رنک A با رنک AA^T برابر است تا هر 2 طرف اثبات شود.

$$\forall x \in Null\,space(A) \longrightarrow Ax=0 \longrightarrow A^TAx=0 \longrightarrow Null\,space(A) \subseteq Null\,space(A^TA)$$

$$\forall x \in Null\,space(A^TA) \longrightarrow A^TAx=0 \longrightarrow \underbrace{x^TA^TAx}_{\|Ax\|}=0 \longrightarrow Ax=0 \longrightarrow Null\,space(A^TA) \subseteq Null\,space(A)$$

$$Null\,space(A^TA) = Null\,space(A)$$

$$A^TA:\quad m = \dim(Null\,space(A^TA)) + \overbrace{\dim(range(A^TA))}^{rank(A^TA)}$$

$$A:\quad m = \dim(Null\,space(A)) + \underbrace{\dim(range(A))}_{rank(A)}$$

$$\longrightarrow rank(A) = rank(A^TA) \xrightarrow[\text{change the rank}]{\text{transposing doesn't}} rank(A) = rank(AA^T)$$

بنابراین نشان دادیم که A فول رنک سطری است اگر و تنها اگر AA^T وارون پذیر باشد، پس شرط لازم است.

2.b

ابتدا نشان می دهیم به ازای هر x' ای که جواب معادله باشد ، $x_{min-norm} - x'$ بر $x_{min-norm}$ عمود است :

$$\left(x_{min-norm} - x'\right)^T \underbrace{x_{min-norm}}_{A^T(AA^T)^{-1}b} = \Big(\underbrace{A\left(x_{min-norm} - x'\right)}_{b-b}\Big)^T (AA^T)^{-1} b = 0$$

حال x' را به صورت زیر می نویسیم :

$$\|x'\|_2^2 = \|x' - x_{min-norm} + x_{min-norm}\|_2^2 \overset{\text{عمود بودن}}{=\!=\!=} \|x' - x_{min-norm}\|_2^2 + \|x_{min-norm}\|_2^2 \geqslant \|x_{min-norm}\|_2^2$$

پس نشان دادیم نورم آن بین تمام جواب ها کمترین است

3.a

$$\begin{vmatrix} 2-\lambda & 0 & 1 \\ -1 & 2-\lambda & 3 \\ 1 & 0 & 2-\lambda \end{vmatrix} = (2-\lambda)^3 - (2-\lambda) = (2-\lambda)\left[(2-\lambda)^2 - 1\right] = 0 \longrightarrow \begin{cases} \lambda = 2 \\ \lambda = 1 \\ \lambda = 3 \end{cases}$$

$$\xrightarrow{\lambda=2} \left[\begin{array}{ccc|c} 0 & 0 & 1 & 0 \\ -1 & 0 & 3 & 0 \\ 1 & 0 & 0 & 0 \end{array}\right] \underset{\text{equivalent}}{\overset{\text{row}}{\equiv}} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \longrightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ x_2 \\ 0 \end{bmatrix} \longrightarrow \text{eigen Vector} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$\xrightarrow{\lambda=1} \left[\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ -1 & 1 & 3 & 0 \\ 1 & 0 & 1 & 0 \end{array}\right] \equiv \left[\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \longrightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -x_3 \\ -4x_3 \\ x_3 \end{bmatrix} \longrightarrow \text{eigen vector} = \begin{bmatrix} -1 \\ -4 \\ 1 \end{bmatrix}$$

$$\xrightarrow{\lambda=3} \left[\begin{array}{ccc|c} -1 & 0 & 1 & 0 \\ -1 & -1 & 3 & 0 \\ 1 & 0 & -1 & 0 \end{array}\right] \equiv \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \longrightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_3 \\ 2x_3 \\ x_3 \end{bmatrix} \longrightarrow \text{eigen vector} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 0 & -1 & 1 \\ 1 & -4 & 2 \\ 0 & 1 & 1 \end{bmatrix} \equiv \begin{bmatrix} 1 & -4 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \longrightarrow$$ 3 بردار ویژه مستقل خطی اند $\longrightarrow$

طبق درس می دانیم که اگر ماتریس $n \times n$ دارای n بردار ویژه مستقل خطی باشد ماتریس قطری پذیر است .

و داریم

$$P = \begin{bmatrix} 0 & -1 & 1 \\ 1 & -4 & 2 \\ 0 & 1 & 1 \end{bmatrix}$$

$\boxed{3.b}$

$$\begin{vmatrix} 6-\lambda & 2 & 0 \\ 2 & 3-\lambda & 0 \\ 0 & 0 & -1-\lambda \end{vmatrix} = (-1-\lambda)\left[(6-\lambda)(3-\lambda)-4\right] = 0 \longrightarrow \begin{cases} \lambda = -1 \\ \lambda = 7 \\ \lambda = 2 \end{cases}$$

$$\xrightarrow{\lambda=-1} \left[\begin{array}{ccc|c} 7 & 2 & 0 & 0 \\ 2 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \equiv \left[\begin{array}{ccc|c} 1 & \frac{2}{7} & 0 & 0 \\ 0 & \frac{24}{7} & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \longrightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ x_3 \end{bmatrix} \longrightarrow \text{eigen vector} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\xrightarrow{\lambda=7} \left[\begin{array}{ccc|c} -1 & 2 & 0 & 0 \\ 2 & -4 & 0 & 0 \\ 0 & 0 & -8 & 0 \end{array}\right] \equiv \left[\begin{array}{ccc|c} 1 & -2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \longrightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2x_2 \\ x_2 \\ 0 \end{bmatrix} \longrightarrow \text{eigen vector} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$$

$$\xrightarrow{\lambda=2} \left[\begin{array}{ccc|c} 4 & 2 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 0 & 0 & -3 & 0 \end{array}\right] \equiv \left[\begin{array}{ccc|c} 1 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \longrightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2}x_2 \\ x_2 \\ 0 \end{bmatrix} \longrightarrow \text{eigen vector} = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 2 & -1 \\ 0 & 1 & 2 \\ 1 & 0 & 0 \end{bmatrix} \equiv \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \longrightarrow$$ 3 بردار ویژه مستقل خطی اند $$\longrightarrow P = \begin{bmatrix} 0 & 2 & -1 \\ 0 & 1 & 2 \\ 1 & 0 & 0 \end{bmatrix}$$ قطری پذیر است و داریم

$\boxed{3.c}$

$$\begin{vmatrix} 2-\lambda & 0 & 1 \\ 0 & 2-\lambda & 1 \\ 1 & 1 & 1-\lambda \end{vmatrix} = (2-\lambda)\Big[(2-\lambda)(1-\lambda)-1\Big] + \Big[-(2-\lambda)\Big] \longrightarrow \begin{cases} \lambda = 2 \\ \lambda = 0 \\ \lambda = 3 \end{cases}$$

$$\xrightarrow{\lambda=2} \left[\begin{array}{ccc|c} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & -1 & 0 \end{array}\right] \equiv \left[\begin{array}{ccc|c} 1 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \longrightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -x_2 \\ x_2 \\ 0 \end{bmatrix} \longrightarrow \text{eigen vector} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

$$\xrightarrow{\lambda=0} \left[\begin{array}{ccc|c} 2 & 0 & 1 & 0 \\ 0 & 2 & 1 & 0 \\ 1 & 1 & 1 & 0 \end{array}\right] \equiv \left[\begin{array}{ccc|c} 1 & -1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \longrightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{-x^3}{2} \\ \frac{-x_3}{2} \\ x_3 \end{bmatrix} \longrightarrow \text{eigen vector} = \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}$$

$$\xrightarrow{\lambda=3} \left[\begin{array}{ccc|c} -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ 1 & 1 & -2 & 0 \end{array}\right] \equiv \left[\begin{array}{ccc|c} 1 & 1 & -2 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \longrightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_3 \\ x_3 \\ x_3 \end{bmatrix} \longrightarrow \text{eigen vector} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} -1 & -1 & 1 \\ 1 & -1 & 1 \\ 0 & 2 & 1 \end{bmatrix} \equiv \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 3 \end{bmatrix} \longrightarrow$$ 3 بردار ویژه مستقل خطی اند $$\longrightarrow P = \begin{bmatrix} -1 & -1 & 1 \\ 1 & -1 & 1 \\ 0 & 2 & 1 \end{bmatrix}$$ قطری پذیر است و داریم

4.a

طبق برهان خلف فرض می‌کنیم مقدار ویژه های متناظر با u و v یکی نباشند

$$\left.\begin{aligned} Tu &= \lambda_1 u \\ Tv &= \lambda_2 v \end{aligned}\right\} \longrightarrow T(u+v) = \lambda_1 u + \lambda_2 v$$

$$T(u+v) = \lambda_3 (u+v)$$

$$\longrightarrow \lambda_1 u + \lambda_2 v = \lambda_3 u + \lambda_3 v$$

$$\longrightarrow (\lambda_1 - \lambda_3) u + (\lambda_2 - \lambda_3) v = 0$$

می دانیم بردار ویژه های متناظر با مقادیر ویژه متفاوت از هم مستقل خطی اند

$$\longrightarrow \begin{aligned} \lambda_1 - \lambda_3 &= 0 \\ \lambda_2 - \lambda_3 &= 0 \end{aligned} \longrightarrow \lambda_1 = \lambda_2 = \lambda_3$$

پس فرض خلف باطل است و مقادیر ویژه متناظر با u, v یکسان اند

4.b

واضح است که بردار ویژه های متناظر با مقدار ویژه غیر صفر ، عضو Nullspace(T) نیستند و در range(T) به صورت $\lambda_i u_i$ حضور دارند. طبق درس می دانیم که بردار ویژه های متناظر با مقادیر ویژه متمایز نسبت به هم مستقل هستند و بنابراین در فضای رنج هم مستقل هستند زیرا تنها در مقدار ویژه‌شان ضرب شوند.

طبق برهان خلف فرض می‌کنیم $rank(T)=r$ و T حداقل $r+2$ مقدار ویژه متمایز دارد. بنابراین حداقل $r+1$ ای از این بردار ویژه ها، مقدار ویژه شان مخالف صفر است و بنابراین طبق مطالب گفته شده، حداقل $r+1$ بردار مستقل خطی در رنج داریم

$$r+1 \leq \underbrace{\dim(range(T))}_{rank(T)}$$

پس به تناقض رسیدیم و فرض خلف باطل است. حال باید نشان دهیم که $r+1$ مقدار ویژه متمایز می تواند رخ دهد

اگر یکی از آنها مقدار ویژه اش 0 باشد این حالت رخ می دهد ← T حداکثر $r+1$ مقدار ویژه متمایز دارد

4.c

پایه های $e_1 \dots e_n$ را در نظر بگیرید. داریم:

$$\begin{aligned} Te_1 &= \lambda_1 e_1 \\ &\vdots \\ Te_n &= \lambda_n e_n \end{aligned} \longrightarrow T = \begin{bmatrix} | & & | \\ \lambda_1 e_1 & \cdots & \lambda_n e_n \\ | & & | \end{bmatrix}_{n\times n} = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}_{n\times n}$$

هر برداری، بردار ویژه است پس $v' = e_1+\dots+e_n$ نیز بردار ویژه است و داریم

$$Tv' = T(e_1+\dots e_n) = \lambda'(e_1+\dots+e_n)$$

T خطی است $\longrightarrow T(e_1+\dots+e_n) = Te_1+\dots+Te_n = \lambda_1 e_1 + \dots \lambda_n e_n$

$$\longrightarrow (\lambda' - \lambda_1) e_1 + \dots + (\lambda' - \lambda_n) e_n = 0$$

پایه ها مستقل خطی اند

$$\lambda_1 = \dots = \lambda_n = \lambda' \longrightarrow T = \lambda' I$$

مضرب اسکالر

$\boxed{5.a}$

$$\frac{\partial \hat{y}_k}{\partial z_2^k} = \frac{\partial}{\partial z_2^k}\left(\frac{e^{z_2^k}}{\sum_{j=1}^{k} e^{z_2^j}}\right) = \frac{\left(e^{z_2^k}\sum_{j=1}^{k} e^{z_2^j}\right) - e^{2 z_2^k}}{\left(\sum_{j=1}^{k} e^{z_2^j}\right)^2}$$

$\boxed{5.b}$

$$\frac{\partial \hat{y}_k}{\partial z_2^{i\neq k}} = \frac{\partial}{\partial z_2^{i\neq k}}\left(\frac{e^{z_2^k}}{\sum_{j=1}^{k} e^{z_2^j}}\right) = \frac{-e^{\left(z_2^{i\neq k} + z_2^k\right)}}{\left(\sum_{j=1}^{k} e^{z_2^j}\right)^2}$$

$\boxed{5.c}$
$\boxed{5.d}$

ابتدا مواردی را تعریف می کنیم تا جواب در ادامه تمیز تر باشد

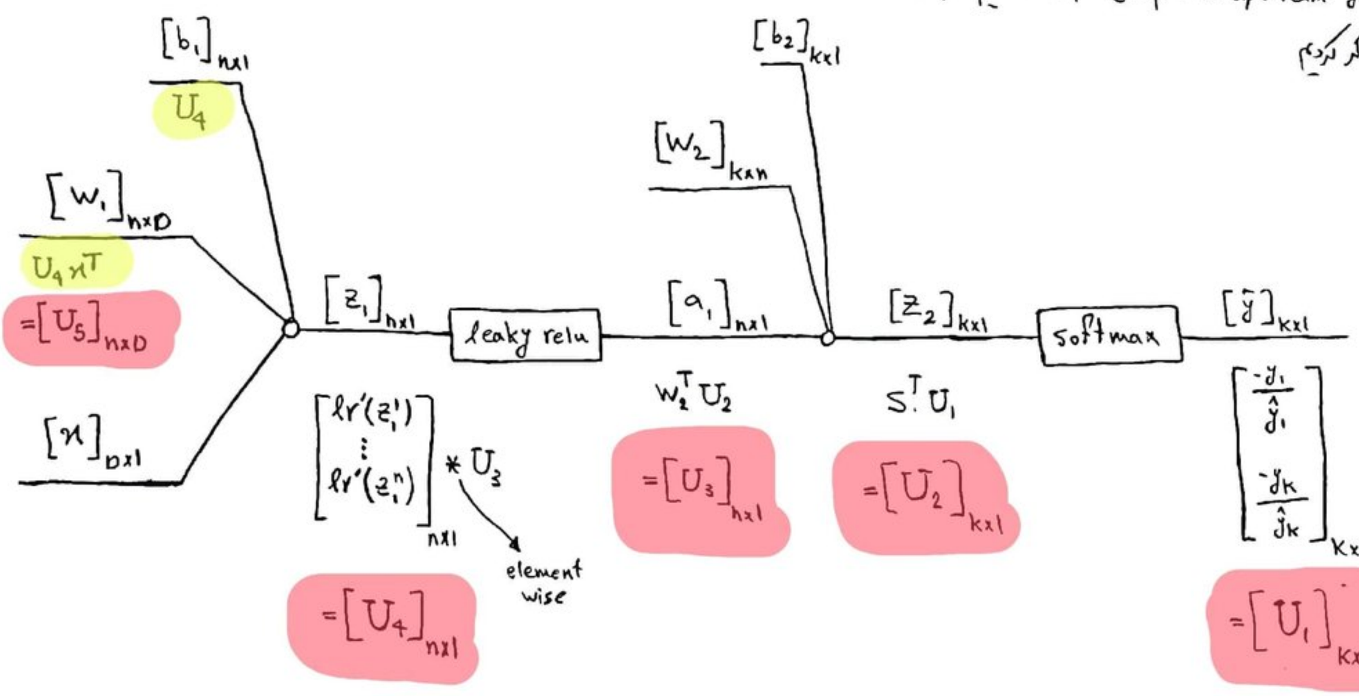
$$s(k) = \frac{\partial \hat{y}_k}{\partial z_2^k} \quad \text{بخش الف}$$

$$s(k,i) = \frac{\partial \hat{y}_k}{\partial z_2^{i\neq k}} \quad \text{بخش ب}$$

$$\longrightarrow S = \begin{bmatrix} s(1) & \cdots & s(1,k) \\ \vdots & \ddots & \vdots \\ s(k,1) & \cdots & s(k) \end{bmatrix}_{k\times k}$$

$$lr' = \begin{cases} \alpha & x \leqslant 0 \\ 1 & x > 0 \end{cases}$$

* حاصل chain rule را روی شکل نشان دادیم
* حاصل مقادیر forward را بالای یال‌ها نوشتیم
* حاصل مقادیر backward (مشتق loss نسبت به هر بخش) را زیر یال ها نوشتیم
* برای upstream gradient ها اسم‌های U_i گذاشتیم و بعد آن‌ها را نیز ذکر کردیم



اگر w_i ها را مثبت در نظر بگیریم می‌توان نوشت :

$\boxed{6.a}$

$$\sum_{i=1}^{m} w_i \left(a_i^T x - b_i\right)^2 = \sum_{i=1}^{m} \left[\sqrt{w_i}\left(a_i^T x - b_i\right)\right]^2$$

$$= \Bigg\| \underbrace{\begin{bmatrix} \sqrt{w_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{w_m} \end{bmatrix}_{m\times m}}_{D} \underbrace{\begin{bmatrix} a_1^T x - b_1 \\ \vdots \\ a_m^T x - b_m \end{bmatrix}_{m\times 1}}_{Ax-b} \Bigg\|^2$$

$\boxed{6.b}$ طبق بخش قبل واضح است که ضرب ماتریس قطری D در A باعث می‌شود هر سطر A در $\sqrt{w_i}$ ضرب شود یعنی داریم :

$$DA = \begin{bmatrix} — \sqrt{w_1}\, a_1^T — \\ — \sqrt{w_m}\, a_m^T — \end{bmatrix}_{m\times n}$$

بنابراین واضح است که در تعداد سطرهای مستقل خطی تغییری ایجاد نمی‌شود پس row rank همان مقدار قبلی می‌ماند و بنابراین col rank نیز ثابت می‌ماند و چون $\text{col rank}(A)=n$ پس $\text{col rank}(B)=n$ و ستون های B نیز مستقل خطی اند

روش 2) می‌دانیم ضرب یک ماتریس در ماتریس full rank، رنک آن را تغییری نمی‌دهد و چون D فول رنک است، پس همچنان ستون های B نیز مستقل خطی می‌مانند

$\boxed{6.c}$ $\text{جواب} = (B^TB)^{-1}B^Td = (A^TD^TDA)^{-1}A^TD^TDb$

$$D^TD = \begin{bmatrix} w_1 & & 0 \\ & \ddots & \\ 0 & & w_m \end{bmatrix} = \text{diag}(w) = W \longrightarrow \text{جواب} = (A^TWA)^{-1}A^TWb$$

$\boxed{7}$

عبارت داده شده معادل $\|Ax-(b-k)\|^2+l$ است. $[k]_{m\times 1}$ و l یک اسکالر

$$\|Ax-b+k\|^2 = \|Ax-b\|^2+\|k\|^2+ 2\underbrace{k^T(Ax-b)}_{k^TAx - k^Tb}$$

$$= \|Ax-b\|^2 + \underbrace{\underbrace{2k^TAx}_{(2A^Tk)^T}}_{c^T} + \underbrace{\|k\|^2 - 2k^Tb}_{d}$$

بنابراین به ازای مقادیر به دست آمده برای c و d می‌توان از این عبارت به جای اولی استفاده کرد. پس داریم

$$\text{جواب} = (A^TA)^{-1}A^T(b-k)$$

در برخی سوال ها با علی بابابیگ همفکری داشتم