Ans 1 : $x_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$, $x_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$, $x_3 = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}$

$$\|x_1\|_2 = \sqrt{x_{11}^2 + x_{12}^2 + x_{13}^2} = \sqrt{1^2 + 0^2 + 1^2}$$
$$= \sqrt{2}$$
$$\|x_2\|_2 = \sqrt{x_{21}^2 + x_{22}^2 + x_{23}^2} = \sqrt{1^2 + 1^2 + 0^2} = \sqrt{2}$$
$$\|x_3\|_2 = \sqrt{x_{31}^2 + x_{32}^2 + x_{33}^2} = \sqrt{0^2 + 2^2 + 1^2} = \sqrt{5}$$

$$\|x_1 - x_2\|_2 = \sqrt{(x_{11} - x_{21})^2 + (x_{12} - x_{22})^2 + (x_{13} - x_{23})^2}$$
$$= \sqrt{(1-1)^2 + (0-1)^2 + (1-0)^2} = \sqrt{2}$$
$$\|x_2 - x_3\|_2 = \sqrt{(1-0)^2 + (1-2)^2 + (1)^2} = \sqrt{3}$$
$$\|x_3 - x_1\|_2 = \sqrt{(1-0)^2 + (0-2)^2 + (1-1)^2} = \sqrt{5}$$

x_1 and x_2 are closest.

x_3 and x_1 are farthest.

$$x_1^T x_2 = x_{11}x_{21} + x_{12}x_{22} + x_{13}x_{23} = 1\times1 + 0\times1 + 1\times0$$
$$= 1.$$
$$x_2^T x_3 = 1\times0 + 2\times1 + 0\times1 = 2$$
$$x_3^T x_1 = 1\times0 + 0\times2 + 1\times1 = 1.$$

Ans 2: Two vectors x_1 ~~x_2~~ and x_2 are orthogonal to each other if

$$x_1^T x_2 = 0$$

Now, we have $x_1 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{bmatrix}$ and $x_2 = \begin{bmatrix} 1/2 \\ 1/\sqrt{2} \\ 1/2 \end{bmatrix}$

Therefore we have that

$$x_1^T x_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1/2 \\ 1/\sqrt{2} \\ 1/2 \end{bmatrix}$$

$$= \frac{1}{2} \times \frac{1}{\sqrt{2}} + 0 \times \frac{1}{\sqrt{2}} - \frac{1}{2} \times \frac{1}{\sqrt{2}}$$

$$= 0.$$

Hence, given vectors are orthogonal to each others.

Ans 3: $A = \begin{bmatrix} a_1 & a_2 & & a_n \end{bmatrix} \in \mathbb{R}^{d \times n}$

we have that $a_i^T a_j = 0$ if $i \neq j$

and $a_i^T a_j = 1$ if $i = j$.

That means, columns of A are orthogonal to each others and have unit norm.

Now, we have to show that $A^T A = I$, where I is an identity matrix.

$$A^T A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}^T \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}$$

$$= \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_n^T \end{bmatrix} \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}$$

$$= \begin{bmatrix} a_1^T a_1 & a_1^T a_2 & \cdots & a_1^T a_n \\ a_2^T a_1 & a_2^T a_2 & \cdots & a_2^T a_n \\ \vdots & \vdots & \vdots & \vdots \\ a_n^T a_1 & a_n^T a_2 & \cdots & a_n^T a_n \end{bmatrix}$$

since we have that $a_i^T a_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$

Therefore, $A^T A = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & & 1 \end{bmatrix} = I.$

<u>Ans 4.</u> We have $A \in \mathbb{R}^{n \times n}$, $u, v \in \mathbb{R}^n$ and

$u' = Au$, $v' = Av$ $\in \mathbb{R}^n$.

Now we have that

$$\|u' - v'\|_2^2 = \|Au - Av\|_2^2$$
$$= \|A(u-v)\|_2^2$$
$$= (A(u-v))^T (A(u-v))$$
$$= (u-v)^T A^T A (u-v)$$

since we are given that A is an orthonormal matrix that means $A^T A = I$

Therefore,

$$\|u' - v'\|_2^2 = (u-v)^T I (u-v)$$
$$= (u-v)^T (u-v)$$
$$= \|u-v\|_2^2$$

$$\Rightarrow \|u' - v'\|_2 = \|u - v\|_2 .$$

Ans. 5.

$f(x) = x^T Ax + x^T b$

~~since ... that~~

Let $f(x) = f_1(x) + f_2(x)$

where $f_1(x) = x^T A x$, $f_2(x) = x^T b$

$\Rightarrow \nabla f(x) = \nabla f_1(x) + \nabla f_2(x)$

In prob 7, we will see that $x^T A x = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i \cdot x_j \cdot a_{ij}$

Therefore,

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}^T$$

Now let us find $\frac{\partial f}{\partial x_k}$

$$\frac{\partial f_1}{\partial x_k} = \frac{\partial}{\partial x_k}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} x_i \cdot x_j \cdot a_{ij}\right]$$

$$= \sum_{i=1}^{n} x_i a_{ik} + \sum_{j=1}^{n} x_j \cdot a_{kj}$$

$$= x^T a_k + x^T \bar{a}_k$$

Here $\bar{a}_k$ is the k^{th} column of the matrix A^T.

$$\Rightarrow \nabla f_1 = \begin{bmatrix} x^T a_1 + x^T \bar{a}_1 \\ x^T a_2 + x^T \bar{a}_2 \\ \vdots \quad\quad \vdots \\ x^T a_n + x^T \bar{a}_n \end{bmatrix} = \begin{bmatrix} a_1^T x \\ a_2^T x \\ \vdots \\ a_n^T x \end{bmatrix} + \begin{bmatrix} \bar{a}_1^T x \\ \bar{a}_2^T x \\ \vdots \\ \bar{a}_n^T x \end{bmatrix}$$

$$= \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_n^T \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} \bar{a}_1^T \\ \bar{a}_2^T \\ \vdots \\ \bar{a}_n^T \end{bmatrix} x$$

$$= A^T x + (A^T)^T x$$

$$= A^T x + Ax = (A^T + A)x$$

$$\nabla f_2 = \begin{bmatrix} \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \end{bmatrix}^T$$

$$f_2 = x^T b = x_1 b_1 + x_2 b_2 + \cdots + x_n b_n$$

$$\Rightarrow \frac{\partial f_2}{\partial x_i} = \frac{\partial}{\partial x_i}\left[x_1 b_1 + \cdots + x_n b_n\right]$$

$$= b_i$$

$$\Rightarrow \nabla f_2 = \begin{bmatrix} b_1 & b_2 & \cdots & b_n \end{bmatrix}^T = b$$

$$\Rightarrow \nabla f = (A^T + A)x + b$$

if $A^T = A$ then

$$\boxed{\nabla f = 2Ax + b}$$

Ans 6 :

$$f(A) = \text{trace}(A),$$

$$\frac{\partial f}{\partial A} = \begin{bmatrix} \frac{\partial f}{\partial a_{11}} & \frac{\partial f}{\partial a_{12}} \\ \frac{\partial f}{\partial a_{21}} & \frac{\partial f}{\partial a_{22}} \end{bmatrix}$$

(a) $f(A) = \text{trace}(A^T A)$

$$= \text{trace}\left(\begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right)$$

$$= a_{11}^2 + a_{21}^2 + a_{12}^2 + a_{22}^2$$

Now,

$$\frac{\partial f}{\partial A} = \begin{bmatrix} 2a_{11} & 2a_{12} \\ 2a_{21} & 2a_{22} \end{bmatrix} = 2\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

$$= 2A.$$

(6) $f(A) = \text{trace}(B^T A)$

$$= \text{trace}\left(\begin{bmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right)$$

$$= a_{11}b_{11} + b_{21}a_{21} + b_{12}a_{12} + b_{22}a_{22}$$

$$\Rightarrow \frac{\partial f}{\partial A} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = B.$$

(b) $f(A) = \text{trace}(A^T B A)$

$$= \text{trace}\left(\begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix}\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} A\right)$$

$$= \text{trace}\left(\begin{bmatrix} a_{11}b_{11} + a_{21}b_{21} & a_{11}b_{12} + a_{21}b_{22} \\ a_{12}b_{11} + a_{22}b_{21} & a_{12}b_{12} + a_{22}b_{22} \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right)$$

$$= a_{11}^2 b_{11} + a_{11}a_{21}b_{21} + a_{11}a_{21}b_{12} + a_{21}a_{21}b_{22}$$
$$+ a_{12}b_{11}a_{12} + a_{12}a_{22}b_{21} + a_{22}a_{12}b_{12} + a_{22}a_{22}b_{22}$$

$$\Rightarrow \frac{\partial f}{\partial a_{11}} = 2a_{11}b_{11} + a_{21}b_{21} + a_{21}b_{12}.$$

$$\frac{\partial f}{\partial a_{12}} = 2b_{11}a_{12} + a_{22}b_{21} + \cancel{a_{22}b_{22}} + a_{22}b_{12}$$

$$\frac{\partial f}{\partial a_{21}} = a_{11}b_{21} + a_{11}b_{12} + 2a_{21}b_{22}$$

$$\frac{\partial f}{\partial a_{22}} = 2a_{22}b_{22} + a_{12}b_{21} + a_{12}b_{12}$$

$$\Rightarrow \frac{\partial f}{\partial A} = \begin{bmatrix} 2a_{11}b_{11} + b_{21}a_{21} + a_{21}b_{12} & 2a_{12}b_{11} + a_{22}b_{21} + a_{22}b_{12} \\ 2a_{21}b_{22} + a_{11}b_{21} + a_{11}b_{12} & 2a_{22}b_{22} + a_{12}b_{21} + a_{12}b_{12} \end{bmatrix}$$

$$= \begin{bmatrix} 2b_{11} & b_{12}+b_{21} \\ b_{21}+b_{12} & 2b_{22} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

$$= (B+B^T)A$$

$$= BA + B^TA$$

Ans 7. let $n = 2$

$$[x_1 \;\; x_2] \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}x_1 + a_{21}x_2 & x_1a_{12} + x_2a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= a_{11}x_1^2 + a_{21}x_2x_1 + x_1x_2a_{12} + x_2^2a_{22}$$

$$= \sum_{i=1}^{2}\sum_{j=1}^{2} x_i \cdot x_j\, a_{ij}$$

let now ~~to~~ solve it for any n.

$$x^T A x = x^T [a_1 \ a_2 \ \dots \ a_n] x$$

$$= [x^T a_1 \ x^T a_2 \ \dots \ x^T a_n] x$$

$$= \sum_{i=1}^{n} (x^T a_i) x_i$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} x_j \cdot a_{ij} \cdot x_i$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} x_i \cdot x_j \cdot a_{ij}$$

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_n \end{bmatrix} \qquad a_i = \begin{bmatrix} a_{i1} \\ a_{i2} \\ a_{in} \end{bmatrix}$$

<u>Ans 8</u> :

$$\text{Trace}\left((B^T A B)\right)$$

$$= \text{Trace}\left([b_1 \ b_2 \ \dots b_n]^T [a_1 \ a_2 \ \dots a_n][b_1 \ b_2 \ \dots b_n]\right)$$

$$= \text{Trace}\left(\begin{bmatrix} b_1^T \\ b_2^T \\ \vdots \\ b_n^T \end{bmatrix} [a_1 \ a_2 \ \dots a_n][b_1 \ b_2 \ \dots b_n]\right)$$

$$= \text{Trace}\left(\begin{bmatrix} b_1^T a_1 & b_1^T a_2 & \dots & b_1^T a_n \\ b_2^T a_1 & b_2^T a_2 & \dots & b_2^T a_n \\ \vdots & \vdots & & \vdots \\ b_n^T a_1 & b_n^T a_2 & \dots & b_n^T a_n \end{bmatrix} [b_1 \ b_2 \ \dots b_n]\right)$$

Now, the ith diagonal entry of the matrix $B^T A B$

$$= \sum_{j=1}^{n} b_i^T a_j \, b_{ij}$$

$$b_i = \begin{bmatrix} b_{i1} \\ b_{i2} \\ b_{in} \end{bmatrix}$$

$$= \sum_{j=1}^{n} \sum_{k=1}^{n} b_{ik}^{T} a_{jk} b_{ij}$$

Now, using Question #7's results

$$= b_{i}^{T} A \, b_{i}$$

~~The~~ Therefore,

$\text{Trace}(B^T A B) =$ Sum of all the diagonal entries of the matrix $B^T A B$

$$= \sum_{i=1}^{n} b_i^T A \, b_i$$

<u>Ans 9:</u>

$$c_{j,i} = \begin{cases} 1 & \text{if } j^{th} \text{ point } x_j \text{ belongs to the } i^{th} \text{ cluster} \\ 0 & \text{otherwise} \end{cases}$$

Therefore

$$c_{j,i} \|x_j - \mu_i\|_2^2 = \begin{cases} \|x_j - \mu_i\| & \text{if } x_j \in C_i \\ 0 & \text{otherwise} \end{cases}$$

Therefore $$\sum_{j=1}^{n} c_{j,i} \|x_j - \mu_i\|_2^2 = \sum_{x_j \in C_i} \|x_j - \mu_i\|_2^2$$

hence $$\sum_{i=1}^{k} \sum_{j=1}^{n} c_{j,i} \|x_j - \mu_i\|_2^2$$

$$= \sum_{i=1}^{k} \sum_{x_j \in C_i} \|x_j - \mu_i\|_2^2$$

$$= \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|_2^2$$

Let us consider the new cost function

$$f(\mu_1, \mu_2, \dots, \mu_k) = \sum_{i=1}^{k}\sum_{j=1}^{n} c_{j,i} \|x_j - \mu_i\|_2^2$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{n} c_{j,i}\left[x_j^T x_j - 2x_j^T\mu_i + \mu_i^T\mu_i\right]$$

$$\Rightarrow$$

$$\nabla_{\mu_r} f = \nabla_{\mu_r}\left(\sum_{i=1}^{k}\sum_{j=1}^{n} c_{j,i}\left(x_j^T x_j - 2x_j^T\mu_i + \mu_i^T\mu_i\right)\right)$$

$$= \sum_{j=1}^{n} \nabla_{\mu_r} c_{j,r}\left(x_j^T x_j - 2x_j^T\mu_r + \mu_r^T\mu_r\right)$$

$$= \sum_{j=1}^{n} c_{j,r}\left[2\mu_r - 2x_j\right] = \vec{0}$$

$$\Rightarrow \sum_{j=1}^{n} c_{j,r}\mu_r = \sum_{j=1}^{n} c_{j,r} x_j$$

$$\Rightarrow \mu_r = \frac{1}{\sum_{j=1}^{n} c_{j,r}} \sum_{j=1}^{n} c_{j,r} x_j$$

$$\because \sum_{j=1}^{n} c_{j,r} = |C_r|$$

$$\Rightarrow \boxed{\mu_r = \frac{1}{|C_r|}\sum_{j=1}^{n} c_{j,r} x_j}$$

$$\because c_{j,r} = \begin{cases} 1 & \text{if } x_j \in C_r \\ 0 & \text{otherwise} \end{cases}$$

$$\Rightarrow \mu_r = \frac{1}{|C_r|}\sum_{x_j \in C_r} x_j$$

Ans 10 :

$$\text{cost to minimize} = \sum_{i=1}^{k} \frac{1}{\sum_{v\in C_i} D_{v,v}} \sum_{r\in C_i} \sum_{s\notin C_i} W_{r,s}$$

Define the cluster assignment matrix $H \in \mathbb{R}^{n\times k}$ as follows

$$H_{i,j} = \begin{cases} \frac{1}{\sqrt{\sum_{v\in C_j} D_{v,v}}} & \text{if } v_i \in C_j \\ 0 & \text{otherwise} \end{cases}$$

Now, with this definition of H, please note that $H^T H \neq I$.

this we can easily show as following

$$h_i^T h_i = \sum_{j=1}^{n} h_{ij}^2 = \sum_{j\in C_i} \frac{1}{\sum_{v\in C_i} D_{v,v}}$$

$$= \frac{1}{\sum_{v\in C_i} D_{v,v}} \sum_{j\in C_i} 1$$

$$= \frac{|C_i|}{\sum_{v\in C_i} D_{v,v}} \neq 1.$$

Now, we will show that

$$h_i^T D h_i = 1$$

$$h_i^T D h_i = \sum_{j=1}^{n} h_{ij}^2 D_{j,j}$$

$$= \sum_{j\in C_i} \frac{D_{j,j}}{\sum_{v\in C_i} D_{v,v}}$$

$$= \sum_{j \in C_i} \frac{D_{j,j}}{\sum_{v \in C_i} D_{v,v}} = \frac{1}{\sum_{v \in C_i} D_{v,v}} \cdot \sum_{j \in C_i} D_{j,j}$$

$$= 1$$

Therefore, $H^T H \neq I$ and

$$H^T D H = I.$$

Now, ~~you~~ we can easily show that

$$\sum_{i=1}^{K} \frac{1}{(\sum_{v \in C_i} D_{v,v})} \sum_{r \in C_i} \sum_{s \notin C_i} W_{r,s}$$

$$= \operatorname{trace}(H^T L H)$$

Now, we have to minimize $\operatorname{trace}(H^T L H)$

such that $H^T D H = I$

$$\Rightarrow \min_{H \in \mathbb{R}^{n \times k}} \operatorname{trace}(H^T L H) \text{ s.t. } H^T D H = I.$$

$$\Rightarrow \min_{h_1, h_2, \dots h_k} \sum_{i=1}^{k} h_i^T L h_i \quad \text{st} \quad h_i^T D h_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

$$f(h_1, h_2, \dots h_k) = \sum_{i=1}^{k} h_i^T L h_i + \sum_{i=1}^{k} \lambda_i (1 - h_i^T D h_i)$$

$$\nabla_{h_i} f = 2 L h_i - \lambda_i 2 D h_i = \vec{0}$$

$$\Rightarrow \quad L h_r = \lambda_r D h_r$$

$$\Rightarrow \quad h_r^T L h_r = \lambda_r \, h_r^T D h_r$$

$$= \lambda_r$$

$$\Rightarrow \quad \operatorname{trace}(H^T L H) = \sum_{i=1}^{k} \lambda_i$$

$$\Rightarrow \quad \min \sum_{i=1}^{k} \lambda_i \quad \text{s.t.}$$

$$h_i^T D h_j = \begin{cases} 1 & \text{if } i \neq j \\ 0 & \text{if } i \neq j. \end{cases}$$

and $\cancel{L h_r = \lambda_r}$

$$L h_r = \lambda_r D h_r$$

$$D^{-1} L h_r = \lambda_r h_r$$

$\Rightarrow \quad h_r \leftarrow$ eigenvector of the matrix $D^{-1}L$ corresponding to the r^{th} ~~largest~~ smallest eigenvalue

$$\Rightarrow \quad H = [u_1 \; u_2 \; \ldots \; u_k]$$

where

$$D^{-1} L u_i = \lambda_i u_i$$

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_k.$$

Ans 11

~~we~~ given ~~this~~ a set of inputs points

$\mathcal{X} = \{x_1, \ldots x_m\}$ $x_i \in \mathbb{R}^d$

We constructed a matrix

$X = [x_1 \ \ldots \ x_m] \in \mathbb{R}^{d \times m}$

Then, we found the ~~matrix~~ EVD of the matrix $XX^T \in \mathbb{R}^{d \times d}$

if $d >> m$ then finding EVD will be computationally intractable as it requires $O(d^3)$ time.

Now let us consider ~~the~~ ~~matrix~~ EVD of the matrix $X^T X \in \mathbb{R}^{m \times m}$.

~~the~~ Our final solution was as below

$U = [u_1 \ u_2 \ \ldots \ u_k]$

where $XX^T u_i = \lambda_i u_i$ $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$

Now, can we use the EVD of $X^T X$ instead of XX^T ?

Let, following is the EVD of X^TX

$\Rightarrow$ $X^TX v_i = \bar{\lambda}_i v_i$

Then multiply this equation by X from left

$\Rightarrow$ $XX^TX v_i = \bar{\lambda}_i X v_i$

$\Rightarrow$ $(XX^T) X v_i = \bar{\lambda}_i X v_i$

$\Rightarrow$ if v_i is the eigenvector of the matrix X^TX corresponding to $\bar{\lambda}_i$

then, $X v_i$ will be the eigenvector of the matrix XX^T corresponding to $\bar{\lambda}_i$.

Therefore, given the EVD of X^TX, which is easier to find, we can find the EVD of the matrix XX^T (which is required)

$$\boxed{u_i = X v_i}.$$