Utsan Mehta 20CS10069 AI61003

Assignment-1

1) Given: $\text{avg}(x) = \frac{1}{n} 1_n^T x$ $\quad \text{std}(x) = \frac{\|x - \text{avg}(x) 1_n\|_2}{\sqrt{n}}$

(a) LHS $= \text{avg}(\alpha x + \beta 1_n)$

$= \frac{1}{n} 1_n^T (\alpha x + \beta 1_n) = \frac{1}{n} 1_n^T \alpha x + \frac{1}{n} 1_n^T \beta 1_n$

$= \frac{\alpha}{n} 1_n^T x + \frac{\beta}{n} 1_n^T 1_n$ $\quad \left[1_n^T . 1_n = [1\ 1 \cdots 1]_{1\times n} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{n\times 1} = n \right]$

$= \alpha \, \text{avg}(x) + \frac{\beta}{n} . n$

$= \alpha \, \text{avg}(x) + \beta$

$=$ RHS $\quad$ Hence proved.

(b) LHS $= \text{std}(\alpha x + \beta 1_n) = \frac{\|\alpha x + \beta 1_n - \text{avg}(\alpha x + \beta 1_n) 1_n\|_2}{\sqrt{n}}$

$= \frac{\|\alpha x + \beta 1_n - (\alpha \, \text{avg}(x) + \beta) 1_n\|_2}{\sqrt{n}}$ $\quad$ [using result of part (a)]

$= \frac{\|\alpha (x - \text{avg}(x) 1_n) + \beta 1_n - \beta 1_n\|_2}{\sqrt{n}}$

$= \frac{\|\alpha (x - \text{avg}(x) 1_n)\|_2}{\sqrt{n}}$

$= |\alpha| . \frac{\|x - \text{avg}(x) 1_n\|_2}{\sqrt{n}}$ $\quad$ [as $\alpha \in \mathbb{R}$]

$= |\alpha| . \text{std}(x)$

$=$ RHS

Hence proved.

(c) Note:

$$\text{std}(x) = \frac{\| x - \text{avg}(x) \cdot 1_n \|_2}{\sqrt{n}}$$

$$= \frac{\left\| \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_{n\times 1} - \begin{bmatrix} \text{avg}(x) \\ \text{avg}(x) \\ \vdots \\ \text{avg}(x) \end{bmatrix}_{n\times 1} \right\|_2}{\sqrt{n}} = \frac{\left\| \begin{bmatrix} x_1 - \text{avg}(x) \\ x_2 - \text{avg}(x) \\ \vdots \\ x_n - \text{avg}(x) \end{bmatrix} \right\|_2}{\sqrt{n}}$$

$$= \frac{\sqrt{\sum_{i=1}^{n} (x_i - \text{avg}(x))^2}}{\sqrt{n}} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \text{avg}(x))^2}{n}} \quad \circledast$$

Now, given that there exist k entries in x s.t. $|x_i - \text{avg}(x)| \geq a$ for some $a > 0$.

Now, for all those i s.t $|x_i - \text{avg}(x)| \geq a$,
we have $(x_i - \text{avg}(x))^2 \geq a^2$

$\Rightarrow$ Summing over all those 'k' i's :

$$\sum_{i} (x_i - \text{avg}(x))^2 \geq ka^2$$ [for all i where $|x_i - \text{avg}(x)| \geq a$]

or $$\sum_{i=1}^{n} (x_i - \text{avg}(x))^2 \geq ka^2$$ [Adding some +ve terms to the LHS of above eq^n, now $i \in [1, n]$]

Using $\circledast$ and above eq^n:

$$n \times (\text{std}(x))^2 = \sum_{i=1}^{n} (x_i - \text{avg}(x))^2 \geq ka^2$$

or $$\boxed{\left(\frac{\text{std}(x)}{a}\right)^2 \geq \frac{k}{n}}$$ Hence proved.

) Weighted norm:

$$\|x\|_w = \sqrt{\sum_{i=1}^{n} w_i x_i^2} \quad , \text{where } w_i > 0, w \in \mathbb{R}^n, x \in \mathbb{R}^n$$

In order to show that $\|.\|_w$ defines a norm, we need to prove:

(i) $\|.\|_w \geq 0$ and $\|x\|_w = 0 \Leftrightarrow x = 0$

(ii) $\|\alpha x\|_w = |\alpha| . \|x\|_w$

(iii) $\|x+y\|_w \leq \|x\|_w + \|y\|_w$.

Proof:

(i) $\|x\|_w \geq 0$ and $\|x\|_w = 0$ iff $x = 0$

$$\|x\|_w = \sqrt{\sum_{i=1}^{n} w_i x_i^2} \quad , \text{where } x_i^2 \geq 0 \Rightarrow w_i x_i^2 \geq 0$$

$$\Rightarrow \sum w_i x_i^2 \geq 0$$

$$\Rightarrow \|x\|_w \geq 0 \quad \text{--- } (*)$$

If $\|x\|_w = 0$

$$\Rightarrow \sum_{i=1}^{n} w_i x_i^2 = 0$$

[since all terms in the summation are shown non-negative, each term has to be zero.]

$$\Rightarrow w_i x_i^2 = 0 \quad \forall i \in 1, 2, -- n$$

$$\Rightarrow x_i = 0 \quad \forall i \in 1, 2, -- n \quad [\text{as } w_i > 0]$$

$$\Rightarrow x = 0$$

If $x = 0$,

Clearly $w_i x_i^2 = 0 \quad \forall i \in 1, 2, -- n$

$$\Rightarrow \|x\|_w = 0$$

$\therefore \|x\|_w = 0$ iff $x = 0$.

(ii) $\|\alpha x\|_w = |\alpha| \, \|x\|_w$

LHS $\|\alpha x\|_w = \sqrt{\sum_{i=1}^{n} w_i(\alpha x_i)^2} = \sqrt{\sum_{i=1}^{n} w_i \alpha^2 x_i^2}$

$= \sqrt{\alpha^2 \sum_{i=1}^{n} w_i x_i^2} = |\alpha| . \sqrt{\sum_{i=1}^{n} w_i x_i^2}$

$= |\alpha| . \|x\|_w$

$=$ RHS Hence proved

(iii) $\|x+y\|_w \leq \|x\|_w + \|y\|_w$

Note:

$\|x\|_w = \sqrt{\sum_{i=1}^{n} w_i x_i^2} = \sqrt{\sum_{i=1}^{n} (\sqrt{w_i} x_i)^2} = \|\tilde{x}\|_2$, where $\tilde{x} = \begin{bmatrix} \sqrt{w_1} x_1 \\ \sqrt{w_2} x_2 \\ \vdots \\ \sqrt{w_n} x_n \end{bmatrix}$

($\|\cdot\|_2 \rightarrow$ 2-norm)

Now, LHS $= \|x+y\|_w = \left\| \begin{bmatrix} x_1+y_1 \\ x_2+y_2 \\ \vdots \\ x_n+y_n \end{bmatrix} \right\|_w = \sqrt{\sum_{i=1}^{n} w_i (x_i+y_i)^2}$

$= \sqrt{\sum_{i=1}^{n} (\sqrt{w_i} x_i + \sqrt{w_i} y_i)^2} = \left\| \begin{bmatrix} \sqrt{w_1} x_1 + \sqrt{w_1} y_1 \\ \sqrt{w_2} x_2 + \sqrt{w_2} y_2 \\ \vdots \\ \sqrt{w_n} x_n + \sqrt{w_n} y_n \end{bmatrix} \right\|_2$

$= \left\| \begin{bmatrix} \sqrt{w_1} x_1 \\ \sqrt{w_2} x_2 \\ \vdots \\ \sqrt{w_n} x_n \end{bmatrix} + \begin{bmatrix} \sqrt{w_1} y_1 \\ \sqrt{w_2} y_2 \\ \vdots \\ \sqrt{w_n} y_n \end{bmatrix} \right\|_2 = \|\tilde{x} + \tilde{y}\|_2$

Now, $\|x+y\|_w = \|\tilde{x}+\tilde{y}\|_2 \leq \|\tilde{x}\|_2 + \|\tilde{y}\|_2$ [using triangle inequality for 2-norm]

$= x_i' - x_i^2$

$\therefore \quad \|x+y\|_w \le \|\tilde{x}\|_2 + \|\tilde{y}\|_2$

$\le \|x\|_w + \|y\|_w \quad [\text{as } \|a\|_w = \|\tilde{a}\|_2]$

Hence proved.

Thus using the proofs above, we can say that weighted norm is a vector norm.

3) Given: $z = (A+B)(x+y)$ where $A, B \in R^{m \times n}$, $x, y \in R^n$.

Computational complexity of approach - 1 :

$A+B$ takes mn cost

$x+y$ takes n cost

$(A+B)(x+y)$ takes $m(2n-1)$ cost

$2n-1 = n + (n-1)$

for each row

for each column

for each multiplication of element of $(A+B)$ to $(x+y)$

for each addition of those multiplication terms.

$\therefore$ Total cost $= mn + n + m(2n-1)$

$= 3mn + n - m.$

Computational complexity of approach - 2 :

Ax takes $m(2n-1)$ cost

Ay takes $m(2n-1)$ cost

Bx takes $m(2n-1)$ cost

By takes $m(2n-1)$ cost

$Ax + Ay + Bx + By$ takes $3m$ cost. [as Ax, Ay, Bx, By are m-vectors]

$= x_s^1 - x_s^2$

$\therefore$ Total cost $= 4m(2n-1) + 3m$

$= 8mn - m.$

For approach-2 to be more computationally efficient than approach-1 :

$$8mn - m < 3mn + n - m$$

$$\Rightarrow 5mn < n$$

$$\Rightarrow 5m < 1$$

$$\Rightarrow \boxed{m < 1/5}$$

, but m (= number of rows in A and B) is always an integer and greater than equal to 1.

$\therefore$ No condition on m, n exists. such that approach 2 is computationally efficient than approach-1.

5) Given x is symmetric if $x_k = x_{n-k+1}$ and

x is antisymmetric if $x_k = -x_{n-k+1}$.

Consider the decomposition $x = \frac{x+\tilde{x}}{2} + \frac{x-\tilde{x}}{2}$,

where $\tilde{x} = \begin{bmatrix} x_n \\ x_{n-1} \\ \vdots \\ x_1 \end{bmatrix}$ and $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

Now, $\frac{x+\tilde{x}}{2} = \left[\frac{x_1+x_n}{2} \quad \frac{x_2+x_{n-1}}{2} \quad \cdots \quad \frac{x_2+x_{n-1}}{2} \quad \frac{x_n+x_1}{2} \right]^T = x_s$

→ which is symmetric

and, $\frac{x-\tilde{x}}{2} = \left[\frac{x_1-x_n}{2} \quad \frac{x_2-x_{n-1}}{2} \quad \cdots \quad \frac{x_{n-1}-x_2}{2} \quad \frac{x_n-x_1}{2} \right]^T = x_a$

→ which is antisymmetric.

Hence x can be decomposed into the sum of a symmetric and antisymmetric vectors.

To prove this decomposition is unique:

Let $x = x_s^1 + x_a^1 = x_s^2 + x_a^2$ be two decompositions, where x_s^1, x_s^2 are symmetric, x_a^1, x_a^2 are antisymmetric.

Now, $x_s^1 - x_s^2 = x_a^2 - x_a^1$

$$\Rightarrow \underbrace{\begin{bmatrix} x_{s_1}^1 - x_{s_1}^2 \\ x_{s_2}^1 - x_{s_2}^2 \\ \vdots \\ x_{s_1}^1 - x_{s_1}^2 \end{bmatrix}}_{\text{symmetric } = x_s^1 - x_s^2} = \underbrace{\begin{bmatrix} x_{a_1}^2 - x_{a_1}^1 \\ x_{a_2}^2 - x_{a_2}^1 \\ \vdots \\ -x_{a_1}^2 - (-x_{a_1}^1) \end{bmatrix}}_{\text{antisymmetric } - x_a^2 - x_a^1}$$

$\therefore$ $y_s = y_a$, where $y_s = x_s^1 - x_s^2$, $y_a = x_a^2 - x_a^1$.

Now, $y_s \Rightarrow y_{s_1} = y_{s_n}$ (from def^n)

and $y_a \Rightarrow y_{a_1} = -y_{a_n}$

But, $y_{s_1} = y_{a_1}$ and $y_{a_n} = y_{s_n}$ (from $y_s = y_a$)

$\Rightarrow$ $y_{a_1} = y_{s_1} = y_{s_n} = y_{a_n} = -y_{a_1}$ $\Rightarrow y_{a_1} = 0 \Rightarrow y_{a_n} = 0 \Rightarrow y_{s_n} = 0$
$\Rightarrow y_{s_1} = 0$

$\Rightarrow$ Similarly, $y_{a_i} = 0$, $y_{s_i} = 0$

$\Rightarrow$ $y_s = 0$, $y_a = 0$

$\Rightarrow$ $x_s^1 - x_s^2 = 0$, $x_a^2 - x_a^1 = 0$

$\Rightarrow$ $x_s^1 = x_s^2$; $x_a^1 = x_a^2$

$\Rightarrow$ Both the decomposition are equal. $\Rightarrow$ Unique decomposition

Hence proved.

6) To prove that left inverse of A exists if and only if columns of A are linearly independent.

[$\Leftarrow$] Columns of A are linearly independent $\Rightarrow$ Left inverse of A exists.

We know that A^TA is invertible iff columns of A are linearly independent $\Rightarrow$ A^TA has a trivial nullspace iff columns of A are linearly independent.

$\therefore$ $(A^TA)^{-1}(A^TA) = I$, given $(A^TA)^{-1}$ exists.

$\therefore$ $((A^TA)^{-1}A^T)A = I$ (Associative property of matrix multiplication)

$\therefore$ $(A^TA)^{-1}A^T$ is the left inverse of A, which exists if columns of A are linearly independent.

[$\Rightarrow$] Left inverse of A exists $\Rightarrow$ Columns of A are linearly independent.

Consider the equation $Ax=0$, for some $x \in R^n$

Pre multiplying both sides by Left inverse of A:

$A^{-1}Ax = A^{-1}0$

$\Rightarrow x = 0$

Thus, $Ax=0$ has a unique and trivial solution $x=0$, whenever there exists a Left inverse of A.

$\therefore$ $Ax=0 \iff x=0$ [The implication $x=0 \Rightarrow Ax=0$ is obvious]

$\Rightarrow$ Linear combination of coloumns of A, given by coefficients as values of x, yields 0 iff values of x are all 0.

$\Rightarrow$ Coloumns of A are linearly independent.

Hence Left inverse of A exists iff coloumns of A are linearly independent.

Hence proved.

7) Given $A \in R^{(n+1)\times(n+1)}$ s.t $A = \begin{bmatrix} I_n & x \\ x^T & 0 \end{bmatrix}$, where $x \in R^n$ and

I_n is $n \times n$ identity matrix.

- For A to be invertible, columns of A are linearly independent

$\Rightarrow Ax=0$ iff $x=0$.

Consider $A\alpha = 0$.

$\Rightarrow \alpha_1 A_1 + \alpha_2 A_2 + \dots + \alpha_{n+1} A_{n+1} = 0$, where $\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{n+1} \end{bmatrix}$ and

$A = [A_1 \ A_2 \ \dots \ A_{n+1}]$ i.e, A_i is a column of A.

Now, $\alpha_1 A_1 + \alpha_2 A_2 + \cdots + \alpha_{n+1} A_{n+1} = \begin{bmatrix} \alpha_1 + \alpha_{n+1} x_1 \\ \alpha_2 + \alpha_{n+1} x_2 \\ \vdots \\ \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n + 0 \end{bmatrix} = 0$

$= \begin{bmatrix} \alpha_1 + \alpha_{n+1} x_1 \\ \alpha_2 + \alpha_{n+1} x_2 \\ \vdots \\ \sum_{i=1}^{n} \alpha_i x_i \end{bmatrix} = 0$ (where $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$)

$\Rightarrow \alpha_1 + \alpha_{n+1} x_1 = 0, \ \alpha_2 + \alpha_{n+1} x_2 = 0, \ \cdots, \ \alpha_n + \alpha_{n+1} x_n = 0,$
$\sum_{i=1}^{n} \alpha_i x_i = 0$

$\Rightarrow \alpha_1 = -\alpha_{n+1} x_1, \ \alpha_2 = -\alpha_{n+1} x_2, \cdots, \alpha_n = -\alpha_{n+1} x_n, \ \sum_{i=1}^{n} \alpha_i x_i = 0$

$(-\alpha_{n+1} x_1) x_1 + (-\alpha_{n+1} x_2) x_2 + \cdots + (-\alpha_{n+1} x_n) x_n = 0$

$\Rightarrow \ -\alpha_{n+1} (x_1^2 + x_2^2 + \cdots + x_n^2) = 0$

If $\sum_{i=1}^{n} x_i^2 = 0 \Rightarrow$ We can choose α s.t $\alpha_{n+1} \neq 0$ but $A\alpha = 0$.

$\Rightarrow$ If $\sum_{i=1}^{n} x_i^2 = 0$ (or $x = 0$), columns of A are linearly dependent.

If $x \neq 0$ (or $\sum_{i=1}^{n} x_i^2 \neq 0$), we have $\alpha_{n+1} = 0 \Rightarrow \alpha_i = 0 \ \forall i = 1, 2, \cdots n$

$\Rightarrow \alpha = 0$ if $\sum_{i=1}^{n} x_i^2 \neq 0 \Rightarrow$ condition for linear independence of coloumns of A. $\Rightarrow$ condition for A to be invertible.

$\therefore \boxed{x \neq 0}$ for A to be invertible.

- To find an expression of A^{-1} in terms of x.

Let $Aa = b$ be an equation which is satisfied by $a = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$ and $b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$ s.t $a_1, b_1 \in \mathbb{R}^n$ and $a_2, b_2 \in \mathbb{R}$.

$$Aa = b \Rightarrow \begin{bmatrix} I_n & x \\ x^T & 0 \end{bmatrix}\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

$$\therefore \quad I_n a_1 + x a_2 = b_1 \quad \text{and} \quad x^T a_1 = b_2$$

$$\text{or} \quad a_1 + x a_2 = b_1 \quad \text{and} \quad \boxed{x^T a_1 = b_2}$$

$$\Downarrow \qquad\qquad \Downarrow$$

$$x^T a_1 + x^T x a_2 = x^T b_1 \Rightarrow b_2 + x^T x a_2 = x^T b_1$$

$$\text{or} \quad \boxed{a_2 = \frac{x^T b_1 - b_2}{\|x\|_2^2}}$$

Again, $a_1 + x a_2 = b_1 \Rightarrow a_1 = b_1 - x a_2 = b_1 - x\left(\frac{x^T b_1 - b_2}{\|x\|_2^2}\right)$

$$\Rightarrow a_1 = \frac{b_1 \|x\|_2^2 - x x^T b_1 + x b_2}{\|x\|_2^2}$$

$$\boxed{a_1 = \frac{b_1(\|x\|_2^2 I_n - x x^T) + x b_2}{\|x\|_2^2}}$$

$$\text{Thus} \quad \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \underbrace{\frac{1}{\|x\|_2^2} \cdot \begin{bmatrix} \|x\|_2^2 I_n - x x^T & x \\ x^T & -1 \end{bmatrix}}_{A^{-1}} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

$$\text{or} \quad a = A^{-1} \cdot b$$

$$\therefore \boxed{A^{-1} = \frac{1}{\|x\|_2^2}\begin{bmatrix} \|x\|_2^2 I_n - x x^T & x \\ x^T & -1 \end{bmatrix}} \rightarrow \text{In terms of } x.$$

8) Given $A \in \mathbb{R}^{m \times n}$ is a matrix with linearly independent columns, $b \in \mathbb{R}^m$ and $\hat{x} \in \mathbb{R}^n$ is the least squares solution to $Ax = b$.

- To show for any $y \in \mathbb{R}^n$, $(Ay)^T b = (Ay)^T (A\hat{x})$.

$\because \hat{x}$ is the least squares solution to $Ax = b$,

$\hat{x} = (A^TA)^{-1}A^Tb$.

Now, RHS $= (Ay)^T A\hat{x} = y^T A^T . A . (A^TA)^{-1} A^T b$

$= y^T((A^TA)(A^TA)^{-1}) A^T b$ [associative property]

$= y^T . A^T . b$ [as $(A^TA).(A^TA)^{-1} = I_n$]

$= (Ay)^T b$

$=$ LHS.

Hence proved.

- To show that $\dfrac{(A\hat{x})^T b}{\|A\hat{x}\|_2 \|b\|_2} = \dfrac{\|A\hat{x}\|_2}{\|b\|_2}$ using the above result, let $y = \hat{x}$ in the result of first part.

$\Rightarrow (A\hat{x})^T b = (A\hat{x})^T A\hat{x} = \|A\hat{x}\|_2^2$

$\therefore$ LHS $= \dfrac{(A\hat{x})^T b}{\|A\hat{x}\|_2 \|b\|_2} = \dfrac{\|A\hat{x}\|_2^2}{\|A\hat{x}\|_2 \|b\|_2} = \dfrac{\|A\hat{x}\|_2}{\|b\|_2} =$ RHS

Hence proved.

a) Given $u = [u_1\ u_2 \cdots u_T]^T$, $y = [y_1\ y_2 \cdots y_T]^T$ are observed time series data, where

$y_t \approx \hat{y}_t = \sum_{j=1}^{n} h_j u_{t-j+1} \quad \forall\, t = 1, 2, \cdots T$, where $h \in \mathbb{R}^n$

and $u_i = 0 \ \forall\ i \leq 0$.

$\therefore \hat{y}_1 = h_1 u_1$, $\hat{y}_2 = h_2 u_1 + h_1 u_2$, $\cdots$, $\hat{y}_T = h_n u_{T-n+1} + h_{n-1} u_{T-n+2}$

$+ \cdots + h_1 u_T$

Thus $\hat{y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_T \end{bmatrix}_{T\times 1} = \begin{bmatrix} u_1 & 0 & \cdots & 0 & \cdots & 0 \\ u_2 & u_1 & 0 & \cdots & \cdots & 0 \\ \vdots & & & & & \\ u_T & u_{T-1} & \cdots & u_{T-n+2} & & u_{T-n+1} \end{bmatrix}_{T\times n} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{bmatrix}_{n\times 1} \Rightarrow y \equiv Ah,$

where $h = [h_1 \; h_2 \cdots h_n]^T$, $A = \begin{bmatrix} u_1 & 0 & \cdots & 0 \\ u_2 & u_1 & 0 \cdots & 0 \\ & & & \\ u_T & u_{T-1} & \cdots & u_{T-n+1} \end{bmatrix}_{T\times n}$

Now, RHS $= (y_1-\hat{y}_1)^2 + (y_2-\hat{y}_2)^2 + \cdots + (y_T-\hat{y}_T)^2$

$$= \sum_{i=1}^{T} (y_i-\hat{y}_i)^2 = \sum_{i=1}^{T} (\hat{y}_i - y_i)^2$$

$$= \|\hat{y}-y\|_2^2$$

$$= \|Ah-y\|_2^2$$

and LHS $= \|Ah-b\|_2^2$

$\therefore$ On comparing LHS and RHS, we find:

$A = \begin{bmatrix} u_1 & 0 & \cdots & 0 \\ u_2 & u_1 & 0 \cdots & 0 \\ & & & \\ u_T & u_{T-1} & \cdots & u_{T-n+1} \end{bmatrix}_{T\times n}$ and $b = y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}_{T\times 1}$

Note that if $n \geq T+1$, then all those values in A are zero for which u_i has $i \leq 0$.

10) K-means clustering:

Input: $x_1, x_2, \ldots x_N \in \mathbb{R}^n$ and initial clusters: $z_1, z_2, \cdots z_k$

Output: Cluster assignment: $c_1, c_2, \cdots, c_N$ of input data.

(a) Computational Complexity for cluster assignment based on cluster representatives:

Computing distance of a cluster representative from a data point = $O(n)$. (or $O(3n-1)$ to be precise).
Computing distance of all cluster representatives from a data point = $O(kn)$. [or $O((3n-1)k)$]
Computing minimum distance of a cluster representative from a data point = $O(kn)$. [or $O((3n-1)k)$]
Computing minimum cluster representative distance for all data points = $O(kNn)$. [or $O((3n-1)Nk)$]

↳ answer.

(b) Computational complexity of updating cluster representatives:

Consider each cluster to have d_i data points. $\Rightarrow \sum_{i=1}^{k} d_i = N$.

Computing average of one feature for a cluster $= O(d_i)$.
Computing average of all features for a cluster $= O(d_i n)$
Computing average of all features for each cluster

$$= O(d_1 n) + O(d_2 n) + \dots + O(d_k n)$$
$$= O((d_1 + d_2 + \dots + d_k)n)$$
$$= O(Nn)$$

answer

(c) Computations for 10 iterations:

Computation of 1 iteration = computations of step 1 + that of step-2

$$= O((3n-1)Nk) + O(Nn)$$
$$= N((3n-1)k + n) \text{ computations}$$

∴ Computations for 10 iterations = $\boxed{10N[(3n-1)k + n] \text{ computations}}$