NAME : AAYANK SINGHAI
ROLLNO : MT 2025 001
MTECH CSE 1^{st} SEMESTER

ASSIGNMENT 1 - AIT 512 Maths for ML

Q1) Consider the vectors $\vec{x_1}, \dots, \vec{x_L}$, $L \geq 1$, $\vec{x_i} \in R^{n \times 1}$ $\forall i$ and the centroid vector $\vec{\bar{x}} = \frac{1}{L}(\vec{x_1} + \dots + \vec{x_L})$
Consider another vector $\vec{z} \in R^{n \times 1}$

i) Show that it is only for $\vec{z} = \vec{\bar{x}}$ does the sum $J = \sum_{i=1}^{L} \|\vec{x_i} - \vec{z}\|^2$ have the smallest value ?
Note that $\|\cdot\|^2$ is the square of the 2-norm of any vector.

Ans

$$J = \sum_{i=1}^{L} \|\vec{x_i} - \vec{z}\|^2$$

$$= \sum_{i=1}^{L} \|\vec{x_i} - \vec{\bar{x}} + \vec{\bar{x}} - \vec{z}\|^2$$

$$= \sum_{i=1}^{L} \|(\vec{x_i} - \vec{\bar{x}}) + (\vec{\bar{x}} - \vec{z})\|^2$$

$$= \sum_{i=1}^{L} \left[\|\vec{x_i} - \vec{\bar{x}}\|^2 + 2(\vec{x_i} - \vec{\bar{x}})^T(\vec{\bar{x}} - \vec{z}) + \|\vec{\bar{x}} - \vec{z}\|^2 \right]$$

Using the squared norm property
$$\|a+b\|^2 = \|a\|^2 + 2a^Tb + \|b\|^2$$

$$= \underbrace{\sum_{i=1}^{L} \|\vec{x_i} - \vec{\bar{x}}\|^2}_{①} + 2\underbrace{\left(\sum_{i=1}^{L}(\vec{x_i} - \vec{\bar{x}})\right)^T}_{②}\underbrace{(\vec{\bar{x}} - \vec{z})}_{\text{It is constant wrt summation } \Sigma} + \underbrace{\sum_{i=1}^{L}\|\vec{\bar{x}} - \vec{z}\|^2}_{③}$$

for ③ $\sum_{i=1}^{L} \|\vec{\bar{x}} - \vec{z}\|^2$ is constant wrt Σ_i. Hence simplified to

$$\sum_{i=1}^{L} \|\vec{x} - \vec{z}\|^2 = L\|\vec{x} - \vec{z}\|^2$$

for ②

$$2\left(\sum_{i=1}^{L} (\vec{x}_i - \vec{\bar{x}})^T\right)(\vec{\bar{x}} - \vec{z})$$

$$\sum_{i=1}^{L} (\vec{x}_i - \vec{\bar{x}}) = \left(\sum_{i=1}^{L} \vec{x}_i\right) - \left(\sum_{i=1}^{L} \vec{\bar{x}}\right)$$

$$\therefore \sum_{i=1}^{L} \vec{x}_i = L\vec{\bar{x}} \quad \text{(due to centroid)}$$

Hence $\sum_{i=1}^{L} (\vec{x}_i - \vec{\bar{x}}) = L\vec{\bar{x}} - L\vec{\bar{x}} = 0$

Now; $J = \sum \|\vec{x}_i - \vec{\bar{x}}\|^2 + \sum \|\vec{\bar{x}} - \vec{z}\|^2$

$$= \underbrace{\sum_{i=1}^{L} \|\vec{x}_i - \vec{\bar{x}}\|^2}_{\text{constant wrt } z} + \underbrace{L\|\vec{\bar{x}} - \vec{z}\|^2}_{\text{dependent on } 'z'}$$

Since $\|\vec{\bar{x}} - \vec{z}\|^2 \geq 0$ (always non-negative)

$$\|\vec{\bar{x}} - \vec{z}\|^2 = 0 \Rightarrow \boxed{\begin{aligned} &\vec{\bar{x}} - \vec{z} = 0 \\ \Rightarrow\ &\vec{z} = \vec{\bar{x}} \end{aligned}} \quad \text{(Hence proved)}$$

ii) In which application would this sum be used?

Ans: It is generally used in K-Means clustering as centroid is the point considered to have least mean square distance from all the points.

Q2) Let the vectors $\vec{x}_1, \dots, \vec{x}_m \in \mathbb{R}^{n \times 1}$ be such that they can be partitioned into exactly two ~~subset~~ clusters, G_1 and G_2, whose representative vectors are $\vec{z}_1, \vec{z}_2$, resp. Let some vectors $\vec{x}_i \in G_i$. What would be the sign of difference $\|\vec{x}_i - \vec{z}_1\|^2 - \|\vec{x}_i - \vec{z}_2\|^2$? What would be the sign

of this difference if $\vec{x_i} \in G_2$?
Can you now state a classification rule, of the form $\vec{w}^T \vec{x}_K + v > 0$ or (< 0), to determine if the vector $\vec{x}_K$ belongs to G_1 or G_2?

Ans. The idea of clustering is to select a data point which belongs to the cluster whose center is closest to

1) If $\vec{x_i} \in G_1$ i.e. $\vec{x_i}$ is closer to centroid $\vec{z_1}$ than it is to $\vec{z_2}$.

$$\| \vec{x_i} - \vec{z_1} \| < \| \vec{x_i} - \vec{z_2} \|$$

$$\| \vec{x_i} - \vec{z_1} \|^2 < \| \vec{x_i} - \vec{z_2} \|^2$$

$$\| \vec{x_i} - \vec{z_1} \|^2 - \| \vec{x_i} - \vec{z_2} \|^2 < 0$$

Hence will be negative.

2) If $\vec{x_i} \in G_2$, i.e. $\vec{x_i}$ is closer to centroid $\vec{z_2}$ than it is to $\vec{z_1}$.

$$\| \vec{x_i} - \vec{z_2} \| < \| \vec{x_i} - \vec{z_1} \|$$

$$\| \vec{x_i} - \vec{z_2} \|^2 < \| \vec{x_i} - \vec{z_1} \|^2$$

$$\| \vec{x_i} - \vec{z_1} \|^2 - \| \vec{x_i} - \vec{z_1} \|^2 > 0$$

Hence will be positive.

Classification rule: Expanding the norms

$$\| \vec{x_i} - \vec{z_2} \|^2 - \| \vec{x_i} - \vec{z_1} \|^2 = \| \vec{x_i} \|^2 - 2 z_2^T x_i + \| \vec{z_2} \|^2 - \| \vec{x_i} \|^2 + 2 z_1^T x_i - \| \vec{z_1} \|^2$$

$$= -2 (\vec{z_2} - \vec{z_1})^T \vec{x} + \| \vec{z_2} \|^2 - \| \vec{z_1} \|^2$$

Hence; $w = -2(\vec{z}_2 - \vec{z}_1)$
$v = \|\vec{z}_2\|^2 - \|\vec{z}_1\|^2$

① If $\vec{w}^T \vec{x}_k + v < 0$, then $\vec{x}_k \in G_1$
② If $\vec{w}^T \vec{x}_k + v > 0$, then $\vec{x}_k \in G_2$
③ If $\vec{w}^T \vec{x}_k + v = 0$, $\vec{x}_k$ lies on decision boundary, equidistant from both centroids.

Q3) Consider the vectors $\vec{x}_1, \ldots, \vec{x}_L$ $L>1$ $\vec{x}_i \in R^{n\times 1}$ $\forall i$ and another vector $\vec{z} \in R^{n\times 1}$. The vector $\vec{x}_i$ is the distance nearest neighbour to $\vec{z}$ if $\|\vec{z} - \vec{x}_i\|^2 \leq \|\vec{z} - \vec{x}_j\|^2$, $i \neq j$
Similarly $\vec{x}_i$ is the angle nearest neighbour to $\vec{z}$ if $\angle(\vec{x}_i, \vec{z}) \leq \angle(\vec{x}_j, \vec{z})$, $i \neq j$.

$$\angle(\vec{x}_i, \vec{z}) = \gamma_i = \cos^{-1}\left(\frac{\vec{x}_i^T \vec{z}}{\|\vec{x}_i\| \, \|\vec{z}\|}\right)$$

① Give a simple numerical ex. where the distance nearest neighbour is not the same as the angle nearest neighbour.

Ans Let $\vec{x}_1$ be $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$; $\vec{x}_2$ be $\begin{bmatrix} 2 \\ 0 \end{bmatrix}$ and $\vec{z} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

Calculating distance nearest neighbour

\# Euclid distance for $\vec{x}_1$; $\|\vec{z} - \vec{x}_1\|_2$

$$= \left\| \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\|_2 = \sqrt{1^2 + 0^2}$$

$$= 1$$

\# Euclid distance for $\vec{x}_2 = \|\vec{z} - \vec{x}_2\|_2$

$$= \left\| \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 \\ 0 \end{bmatrix} \right\|_2 = \sqrt{1^2 + 1^2} = \sqrt{2}$$

M T W T F S S
Page No.:
Date:
YOUVA

→ Calculating Angle Nearest Neighbour (ANN):-

ANN for $\vec{x}_1$

$$\gamma_1 = \cos^{-1}\left(\frac{\vec{x}_1^T \vec{z}}{\|\vec{x}_1\| \|\vec{z}\|}\right) = \cos^{-1}\left(\frac{0\times1 + 1\times1}{1\times\sqrt{2}}\right)$$

$$= \cos^{-1}\left(\frac{1}{\sqrt{2}}\right)$$

$$= \cos 45^\circ$$

$$\gamma_1 = 45^\circ$$

ANN for $\vec{x}_2$

$$\gamma_2 = \cos^{-1}\left(\frac{\vec{x}_2^T \vec{z}}{\|\vec{x}_2\| \|\vec{z}\|}\right) = \cos^{-1}\left(\frac{2}{2\sqrt{2}}\right) = \cos^{-1}\left(\frac{1}{\sqrt{2}}\right)$$

$$\gamma_2 = 45^\circ$$

(2) Suppose the vectors $\vec{x}_i$ are normalized. Show that in this case, the distance nearest neighbour and the angle nearest neighbour are always the same.

Ans Lets assume that $\|z_i\| = 1$ for $i = 1, 2, 3, \ldots, n$

$$\|\vec{x} - \vec{z}_j\| \leq \|\vec{x} - \vec{z}_i\|$$

$$\|\vec{x} - \vec{z}_j\|^2 \leq \|\vec{x} - \vec{z}_i\|^2$$

$$\|\vec{x}\|^2 - 2\vec{x}^T\vec{z}_j + \|\vec{z}_j\|^2 \leq \|\vec{x}\|^2 - 2\vec{x}^T\vec{z}_i + \|\vec{z}_i\|^2$$

Cancelling $\|x\|^2$ from both sides and $\|z_j\|^2$ and $\|z_i\|^2$ are both one.

$$-2\vec{x}^T\vec{z}_j \leq -2\vec{x}^T\vec{z}_i$$

$$\vec{x}^T\vec{z}_j \geq \vec{x}^T\vec{z}_i$$

$$\frac{\vec{x}^T \vec{z_j}}{\|\vec{x}\| \|\vec{z_j}\|} \geq \frac{\vec{x}^T \vec{z_i}}{\|\vec{x}\| \|\vec{z_i}\|} \quad \because \|z_j\| = \|z_i\| = 1$$

$$\arccos\left(\frac{\vec{x}^T \vec{z_j}}{\|\vec{x}\| \|\vec{z_j}\|}\right) \leq \arccos\left(\frac{\vec{x}^T \vec{z_i}}{\|\vec{x}_i\| \|\vec{z_i}\|}\right)$$

arccosine is decreasing function

$$\angle(\vec{x}, \vec{z_j}) \leq \angle(\vec{x}, \vec{z_i})$$

if $a \geq b$ then $\arccos(a) \leq \arccos(b)$

Q4) Consider the n vectors $\vec{x}_i \in R^{n \times 1}$, $i = 1, \ldots, n$, where the vectors $\vec{x}_i$ are of the type

$$\vec{x_1} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \vec{x_2} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \vec{x_n} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

Using the Gram-Schmidt algorithm, examine if the vectors $\vec{x_i}$, form an independent set. Suppose the algorithm does not terminate early, what other property do these vectors have?

Ans Applying Gram-Schmidt algo

$$a_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, a_2 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ldots, a_K = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$k^{th} \text{ orthogonal vector } (v_K) = a_K - \sum_{j=1}^{K-1} (a_K \cdot q_j) q_j$$

$$\text{Orthogonal vector } (q_K) = \frac{v_K}{\|v_K\|}$$

(1) $v_1 = a_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ $\quad \|v_1\| = \sqrt{1^2 + 0^2 + \ldots + 0^2} = 1$

$$\underline{\underline{q_1}} = \frac{v_1}{\|v_1\|} = \frac{v_1}{1} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \underline{\underline{e_1}}$$

(2) $v_2 = a_2 - (a_2 \cdot q_1) \cdot q_1$

$$a_2 \cdot q_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = (1)(1) + (1)(0) + (0)(0) + \dots = 1$$

$$v_2 = a_2 - (1) q_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} - 1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\|v_2\| = \sqrt{0^2 + 1^2 + 0^2 + \dots} = 1$$

$$\underline{\underline{q_2}} = \frac{v_2}{\|v_2\|} = \frac{v_2}{1} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \underline{\underline{e_2}}$$

(3) $a_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 0 \end{bmatrix}$ $\quad v_3 = a_3 - (a_3 \cdot q_1) - (a_3 \cdot q_2) \cdot q_2$

$$a_3 \cdot q_1 = (1)(1) + 1 \cdot 0 + 1 \cdot 0 + \dots = 1$$

$$a_3 \cdot q_2 = 1 \cdot 0 + 1 \cdot 1 + 1 \cdot 0 + \dots = 1$$

$$v_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 0 \end{bmatrix} - (1) \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} - (1) \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$ $= e_3$ and norm $= 1$

$q_3 = e_3$

Through process, set of vectors $\{a_1, a_2, \dots, a_n\}$ cancels out the upper components, isolating one new component at each step.

\# Final orthonormal basis is standard basis for R^n.

$q_1 = e_1, \; q_2 = e_2, \; \dots\dots, q_n = e_n$.

q_k is a vector with a 1 in the k^{th} position.

Q5) A questionnaire in a magazine has 30 questions, broken into two sets of 15 questions. Someone taking the questionnaire answers each question with 'Rarely', 'Sometimes', or 'Often'. The answers are recorded as a 30-vector $\vec{a}$, with $a_i = 1, 2, 3$, if question 'i' is answered 'Rarely', 'Sometimes' or 'Often' resp.
Express the total score 's' in the form of $s = \vec{w}^T \vec{a} + v$ where $\vec{w}$ is a 30-vector and 'v' is scalar.

Ans: It is given that:
rarely (1) $\rightarrow$ 0 points
Sometimes (2) $\rightarrow$ 1 point
often (3) $\rightarrow$ 2 points

Here, points are awarded on index value of each question

i.e. for $i = 1, 2, 3, \ldots, 15$.
$a_i = 2$ sometimes $\longrightarrow$ 1 point
$a_i = 3$ often $\longrightarrow$ 2 point
$a_i = 1$ rarely $\rightarrow$ 0 point

for index range $i = 16, \ldots, 30$
$a_i = 1$ rarely $\rightarrow$ 0 point
$a_i = 2$ Sometimes $\rightarrow$ 2 point
$a_i = 3$ often $\rightarrow$ 4 point

We want to formulate score in the equation:
$$s = \vec{w}^T \vec{a} + v.$$

here ~~f(i)~~ $f_i(a_i) = \gamma_i a_i + \beta_i$
for $i = 1, 2, \ldots 15$
$\gamma + \beta = 0$ $\quad a_i = 1$

$2\gamma + \beta = 1$ — (2) $\qquad a_i = 2$

$3\gamma + \beta = 2$ — (3) $\qquad a_i = 3$

Subtracting 3 from 2

$$\begin{aligned} 3\gamma + \beta &= 2 \\ 2\gamma + \beta &= 1 \\ (-) \quad (-) \quad & (-) \\ \hline \boxed{\gamma = 1} \end{aligned}$$

hence $\boxed{\beta = 1}$

$\underline{f_i(a_i) = a_i - 1}$ on substituting

for $i = 16, 17, \ldots 30$

$\gamma + \beta = 0$ $\qquad a_i = 1$

$2\gamma + \beta = 2$ — (4) $\qquad a_i = 2$

$3\gamma + \beta = 4$ — (5) $\qquad a_i = 3$

Subtracting (5) from 4

$\boxed{\gamma = 2}$ $\quad \boxed{\beta = -2}$

$\boxed{f_i(a_i) = 2a_i - 2}$

Calc score $s = \sum_{i=1}^{15} (a_i - 1) + \sum_{i=16}^{30} (2a_i - 2)$

$s = \sum_{i=1}^{15} a_i + 2 \sum_{i=16}^{30} a_i - (15 \times 1 + 15 \times 2)$

Comparing with $s = \vec{w}^T \vec{a} + v$

we get;

$v = -45$

$w^T a = \sum_{i=1}^{15} a_i + 2 \sum_{i=16}^{30} a_i$

$$\text{score } (s) = \vec{w}^T\vec{a} - 45$$

$$w_i = \begin{cases} 1 & i = 1, 2, \ldots 15 \\ 2 & i = 16, 17, \ldots 30 \end{cases}$$

(1) if every ans is sometimes

$$\vec{w}^T\vec{a} = 15 \times 2 + 2 \times 15 \times 2$$

$$= 90$$

$$s = 90 - 45 = \underline{\underline{45}} \qquad (s = \vec{w}^T\vec{a} - v)$$

(2) if every ans is rarely

$$\vec{w}^T\vec{a} = 15 \times 1 + 15 \times 2$$

$$= 45$$

$$s = 45 - 45 = \underline{\underline{0}}$$

(3) if every ans is often

$$\vec{w}^T\vec{a} = 15 \times 3 + 15 \times 2 \times 3$$

$$= 135$$

$$s = 135 - 45 = \underline{\underline{90}}$$

——— X ———