MTH443A: End semester examination
Full Marks: 100

[1] (a) Let $\underline{X} = (X_1, \dots, X_p)^T, p > 1$, be a random vector with $E(\underline{X}) = 0$ and $Cov(\underline{X}) = \Sigma > 0$. $\underline{Y} = (Y_1, \dots, Y_p)^T$ denote the vector of principal components derived from Σ and $\underline{Z} = \begin{pmatrix} \Sigma^{-\frac{1}{2}}\underline{X} \\ \underline{Y} \end{pmatrix}$.

Prove or disprove (by giving counter example) the following statements:

Statement A: Total variation of $\underline{Z}$ = (Total variation of $\underline{X}$) $+ p$

Statement B: There does not exist any $\Sigma > 0$, such that $Y_i = X_i$ for all $i = 1, \dots, p$.

(b) Let $\underline{x}_1, \dots, \underline{x}_n$ be n p-dimensional observed feature vectors and $\mathcal{X} = (\underline{x}_1, \dots, \underline{x}_n)$ be the $p \times n$ data matrix. Let $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} \underline{x}_i$, $n\,S_n = \sum_{i=1}^{n}(\underline{x}_i - \overline{x})(\underline{x}_i - \overline{x})^T$ and $Q_n = \sum_{i=1}^{n}\sum_{j=1}^{n}(\underline{x}_i - \underline{x}_j)(\underline{x}_i - \underline{x}_j)^T$.

Prove or disprove the following statements:

Statement C: $Q_n = 2\,n\,S_n$

Statement D: $Q_n = 2\,n\,\left(\mathcal{X}\left(I_n - \frac{1}{n}\,\underline{1}_n\underline{1}_n^T\right)\mathcal{X}^T\right)$

20 (10+10) marks

[2] (a) Consider the following supermarket transactions data with 5 records

Transaction ID	Items in shopping cart
T_1	Chicken, Tomato, Cheese
T_2	Chips, Tomato, Peacan
T_3	Chicken, Chips, Tomato, Peacan
T_4	Chips, Peacan
T_5	Chicken, Chips, Tomato, Peacan

Apply apriori algorithm to derive all the association rules satisfying minimum support level of 60% and minimum confidence level 80%; mention the steps where anti-monotone property of apriori algorithm is used in deriving the rule set. List the derived association rules with their corresponding support and confidence.

(b) Consider a 2-class (π_1, π_2) classification problem with the following learning sample:

$$\mathcal{L} = \left\{ \begin{array}{l} ((0,1,0,0), \pi_1),\ ((0,1,0,1), \pi_1), ((1,1,0,1), \pi_2), \\ ((1,1,1,0), \pi_2), ((0,1,1,0), \pi_2), ((0,0,0,1), \pi_1) \end{array} \right\}.$$

Using Euclidean distance metric, find the 3-nearest neighbour classifier of the feature vector $\underline{x} = (1,0,0,1)$.

18 (12+6) marks

[3] Consider the 3 bivariate discrete populations, π_1, π_2 and π_3 with the following probability mass functions:

π_1

$x_2 \backslash x_1$	0	1
0	0.4	0.2
1	0.2	0.2

π_2

$x_2 \backslash x_1$	0	1
0	0.25	0.25
1	0.25	0.25

π_3

$x_2 \backslash x_1$	0	1
0	0.2	0.3
1	0.1	0.4

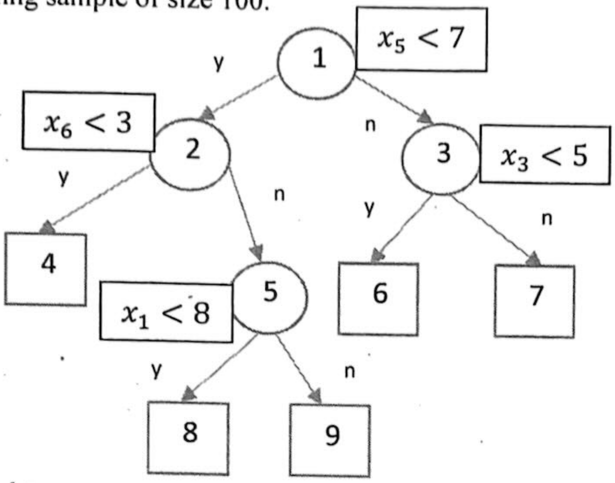
The apriori probabilities of the 3 populations are $p(\pi_1) = p(\pi_2) = p(\pi_3) = \frac{1}{3}$. Let the misclassification costs be as follows: $C(2|1) = 2, C(3|1) = 1$; $C(1|2) = 1, C(3|2) = 2$; $C(1|3) = 2, C(2|3) = 3$.

(a) Find the ECM minimizing classification rule.

(b) Find the ECM of the classification rule obtained in (a).

17 (9+8) marks

[4] Consider the following classification tree ***T*** for a 2-class (π_1, π_2) problem with 6-dimensional feature space, obtained from a learning sample of size 100.



For the constructed tree,

t	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
$N(t)$	100	60	40	10	50	16	24	20	30
$N_2(t)$	60	40	20	0	40	16	4	15	25

$N(t)$: # of training patterns reaching node t and

$N_2(t)$: # of training patterns with label π_2 reaching node t.

(a) Assign class label to the terminal node 7.

(b) Classify the feature vector $(3, 21, 0, 6, 9, 12)$ using the above tree.

(c) Find a measure of tree impurity.

(d) Find the strength of the internal node 3.

(e) Find the Gini Index of node 3.

(f) Under a weakest link pruning approach, obtain the first pruned subtree, T_1.

$$\left[\begin{array}{l}\text{Use misclassification error rate at node } t \text{ as it's impurity measure wherever required;} \\ \text{i. e. } Imp(t) = \dfrac{\sum_{i:\, \underline{x}_i \in U(t)} I(y_i \neq \pi_{j(t)})}{N(t)}\text{; where } j(t) = \underset{i}{\operatorname{argmax}}\, p(\pi_i|t).\end{array}\right]$$

25 (2+2+4+4+3+10) Marks

[5] Let $(X_1, \ldots, X_n)$ be a random sample from a population having a mixture density

$$p(x) = \sum_{j=1}^{2} \pi_j\, p(x|\theta_j);\ \text{where } p(x|\theta_j) = \begin{cases}\theta_j\, x^{\theta_j - 1}, & 0 < x < 1 \\ 0, & \text{otherwise}\end{cases},\ \theta_j > 0,\ i = 1,2.$$

$\pi_1, \pi_2, \theta_1, \theta_2$ are unknown parameters. Formulate the maximum likelihood method for the parameter estimation under E-M algorithm framework and derive the E-M algorithm update equations for π_js and θ_js for density estimation.

12 marks

[6] Consider the learning sample $\mathcal{L} = \{((1,2), \pi_1), ((2,3), \pi_2), ((3,2), \pi_2), ((-1,2), \pi_1)\}$ for 2-class (π_1, π_2) classification problem.

(a) Are the training patterns linearly separable?

(b) Using the instantaneous mode perceptron learning rule with linear classifier (without constant and without margin), obtain the first 2 steps of iteration of the weight vector, starting from the initial weight vector $\underline{v}_0 = (1, 2)'$ and presenting the learning patterns sequentially.

8 (2+6) marks

Solution and marking scheme

1 (a)

$$\text{Cov}(\underline{z}) = \begin{pmatrix} I_p & \Sigma^{-1/2}\,\text{Cov}(\underline{x},\underline{y}) \\ \text{Cov}(\underline{y},\underline{x})\,\Sigma^{-1/2} & \text{Cov}(\underline{y}) \end{pmatrix} = \Sigma_z$$

1

$$\begin{aligned} \text{total variation of } \underline{z} &= \text{tr}\,\Sigma_z \\ &= \text{tr}(I_p) + \text{tr}(\text{Cov}(\underline{y})) \\ &= p + \text{tr}\,\Sigma \quad \left(\because \text{total variation of } \underline{y} = \text{total variation of } \underline{x}\right) \\ &= p + \text{total variation of } \underline{x} \end{aligned}$$

4

Statement A is proved.

Counter example for Statement B

Take $p = 2$ and $\Sigma = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} > 0$

$\lambda_1 = 2,\ \lambda_2 = 1;\ \underline{e}_1 = (1,0)',\ \underline{e}_2 = (0,1)'$

PC 1 : $Y_1 = X_1$ $\quad Y_i = X_i \ \forall i$

PC 2 : $Y_2 = X_2$

5

Statement B is disproved

* There can be many more similar counter examples

(b)

$$Q_n = \sum_i \sum_j (\underline{x}_i - \underline{x}_j)(\underline{x}_i - \underline{x}_j)'$$

$$= \sum_i \sum_j \left(\underline{x}_i \underline{x}_i' - \underline{x}_i \underline{x}_j' - \underline{x}_j \underline{x}_i' + \underline{x}_j \underline{x}_j' \right)$$

$$= n \sum_i \underline{x}_i \underline{x}_i' + n \sum_j \underline{x}_j \underline{x}_j' - (n \bar{\underline{x}})(n \bar{\underline{x}}) - (n \bar{\underline{x}})(n \bar{\underline{x}})$$

$$= 2n \left(\sum \underline{x}_i \underline{x}_i' - n \bar{\underline{x}} \bar{\underline{x}}' \right)$$

$$= 2n \left(\mathcal{X} \mathcal{X}' - n \left(\frac{1}{n} \mathcal{X} \underline{1}_n \right) \left(\frac{1}{n} \mathcal{X} \underline{1}_n \right)' \right)$$

$$= 2n \; \mathcal{X} \left(I_n - \frac{1}{n} \underline{1}_n \underline{1}_n' \right) \mathcal{X}' \qquad 5$$

statement D is proved.

Also,

$$n S_n = \sum_{i=1}^{n} (\underline{x}_i - \bar{\underline{x}})(\underline{x}_i - \bar{\underline{x}})'$$

$$= \sum_{i=1}^{n} \left(\underline{x}_i \underline{x}_i' - \underline{x}_i \bar{\underline{x}}' - \bar{\underline{x}} \underline{x}_i' + \bar{\underline{x}} \bar{\underline{x}}' \right)$$

$$= \sum_i \underline{x}_i \underline{x}_i' - n \bar{\underline{x}} \bar{\underline{x}}' - \cancel{n \bar{\underline{x}} \bar{\underline{x}}'} + \cancel{n \bar{\underline{x}} \bar{\underline{x}}'}$$

$$= \sum_i \left(\underline{x}_i \underline{x}_i' - \bar{\underline{x}} \bar{\underline{x}}' \right)$$

$$= \frac{Q_n}{2n} \qquad 5$$

$\Rightarrow Q_n = 2n^2 S_n$ — statement C is disproved.

2(a) Let A ≡ Chicken

B ≡ Chips

C ≡ Tomato

D ≡ Cheese

E ≡ peacorn

min sup = 60%.

min conf = 80%.

Generation of frequent Itemsets

C_1

Item	Count
A	3
B	4
C	4
D	1 — X (infrequent w.r.t minsup)
E	4

→

L_1

Item	Count
A	3
B	4
C	4
E	4

→

C_2

itemset	count
AB	– 2 — X
AC	– 3
AE	– 2 — X
BC	– 3
BE	– 4
CE	– 3

using apriori algo (anti monotone property)

L_2

Itemset	Count
AC	– 3
BC	– 3
BE	– 4
CE	– 3

→

C_3

Itemset	count
BCE	– 3

using apriori algorithm

Freq Itemsets using apriori algo and w.r.t min sup level

BCE, AC, BC, BE, CE

6 (Give part marking)

(One may also write A, B, C, E are frequent items)

ARM Rule generation

From BCE:		conf
	BC ⇒ E	1 ✓
	BE ⇒ C	.75 < .8 X
	CE ⇒ B	1 ✓
	B ⇒ CE	.75 < .8 X
	C ⇒ BE	.75 < .8 X
	E ⇒ BC	.75 < .8 X

Prune subrules using anti monotone property to get the following ARM rules only

$A \Rightarrow C$ - conf 1

$B \Rightarrow E$ - conf 1

$E \Rightarrow B$ - conf 1

Final ARM rules

$BC \Rightarrow E$ (sup = .6, conf = 1)

$CE \Rightarrow B$ (sup = .6, conf = 1)

$A \Rightarrow C$ (sup = .6, conf = 1)

$B \Rightarrow E$ (sup = .8, conf = 1)

$E \Rightarrow B$ (sup = .8, conf = 1)

6 (Give part marking for subset of these rules)

i.e. {chips, Tomato} $\Rightarrow$ peacan (.6, 1)

{Tomato, peacan} $\Rightarrow$ chips (.6, 1)

Chicken $\Rightarrow$ Tomato (.6, 1)

Chips $\Rightarrow$ peacan (.8, 1)

peacan $\Rightarrow$ chips (.8. 1)

2

(b) Distance of $\underset{\sim}{x}$ from pattern 1 = 3

. . . . 2 = 2

. . . . 3 = 1

. . . . 4 = 3

. . . . 5 = 4

. . . . 6 = 1

$\Rightarrow$ 3 nearest neighbors of $\underset{\sim}{x} = (1,0,0,1)$ are patterns

3, 6 & 2

Majority voting $\Rightarrow$ $\underset{\sim}{x}$ is classified to π_1 6

(3)

(a) ECM rule

$\underset{\sim}{x}$ to π_K if $\sum_{\substack{i=1 \\ i \neq K}}^{3} p_i \, f_i(x) \, C(K|i)$ is min

$p_1 = p_2 = p_3 = \frac{1}{3}$

$\Leftrightarrow \sum_{\substack{i=1 \\ i \neq K}}^{3} f_i(x) \, C(K|i)$ is min

$\underline{(0,0)}$

$K = 1 \rightarrow f_2 \, C(1|2) + f_3 \, C(1|3) = \underline{.65}$

$K = 2 \rightarrow f_1 \, C(2|1) + f_3 \, C(2|3) = 1.4$

$K = 3 \rightarrow f_1 \, C(3|1) + f_2 \, C(3|2) = .9$

$\Rightarrow (0,0)$ in π_1 partition

$2\frac{1}{2}$

$(0,1)$

$K = 1 \quad f_2 \, C(1|2) + f_3 \, C(1|3) = \underline{.45}$

$K = 2 \quad f_1 \, C(2|1) + f_3 \, C(2|3) = .7$

$K = 3 \quad f_1 \, C(3|1) + f_2 \, C(3|2) = .7$

$2\frac{1}{2}$

$(0,1)$ to π_1

$(1,0)$

$K = 1 \quad f_2 \, C(1|2) + f_3 \, C(1|3) = .85$

$K = 2 \quad f_1 \, C(2|1) + f_3 \, C(2|3) = 1.3$

$K = 3 \quad f_1 \, C(3|1) + f_2 \, C(3|2) = \underline{.7}$

2

$(1,0)$ to π_3

$(1,1)$

$K = 1 \quad f_2 \, C(1|2) + f_3 \, C(1|3) = 1.05$

$K = 2 \quad f_1 \, C(2|1) + f_3 \, C(2|3) = 1.6$

$K = 3 \quad f_1 \, C(3|1) + f_2 \, C(3|2) = \underline{.7}$

$(1,1)$ to π_3

2

Any 3 correct give $7\frac{1}{2}$

Any 2 correct give 5

Any 1 correct give $2\frac{1}{2}$

Assignment rule

Assign $\overbrace{(0,0) \;\&\; (0,1)}^{R_1}$ to π_1

Assign $\underbrace{(1,0) \;\&\; (1,1)}_{R_3}$ to π_3

$R_2 = \phi$

(b) ECM of the rule obtained in (a)

$$ECM = p_1 \left(P(2|1)\, c(2|1) + P(3|1)\, c(3|1) \right)$$
$$+ p_2 \left(P(1|2)\, c(1|2) + P(3|2)\, c(3|2) \right)$$
$$+ p_3 \left(P(1|3)\, c(1|3) + P(2|3)\, c(2|3) \right) \quad 2$$

- Note that $P(2|i) = \sum_{\underset{\sim}{x} \in R_2} f_i(\underset{\sim}{x}) = 0$ as $R_2 = \phi$
  $i \neq 2$

$$P(3|1) = \sum_{\underset{\sim}{x} \in R_3} f_1(\underset{\sim}{x}) = .4$$

$$P(1|2) = \sum_{\underset{\sim}{x} \in R_1} f_2(x) = .5$$

$$P(3|2) = \sum_{\underset{\sim}{x} \in R_3} f_2(x) = .5$$

$$P(1|3) = \sum_{\underset{\sim}{x} \in R_1} f_3(x) = .3$$

3 Give part marking

$$\Rightarrow ECM = \frac{1}{3}(.4 \times 1) + \frac{1}{3}(.5 \times 1 + .5 \times 2)$$
$$+ \frac{1}{3}(.3 \times 2) \quad 3$$
$$= \frac{1}{3}(.4 + 1.5 + .6) = \frac{2.5}{3} = 0.8\dot{3}$$

.4

(a) Class label of node 7 : π_1 2

(b) (3, 21, 0, 6, 9, 12) reaches node 6

$\Rightarrow$ assignment is π_2 2

(c) Impurity of tree $= \sum_{t \in \tilde{T}} p(t)\, Imp(t)$

$$= \underbrace{\frac{24}{100} \times \frac{4}{24}}_{\text{node } 7} + \underbrace{\frac{20}{100} \times \frac{5}{20}}_{\text{node } 8} + \underbrace{\frac{30}{100} \times \frac{5}{30}}_{\text{node } 9} + \underbrace{0}_{\text{node } 4} + \underbrace{0}_{\text{node } 6}.$$

$$= \frac{14}{100} = 0.14$$ 4

(d) Strength of node 3

$$g(3) = \frac{R(3) - R(T_3)}{|\tilde{T}_3| - 1}$$

$$R(3) = p(3)\, r(3) = \frac{20}{100}$$

$$R(T_3) = \sum_{6,7} p(t)\, r(t) = \frac{4}{100}$$

$$g(3) = \frac{\frac{20}{100} - \frac{4}{100}}{2-1} = \frac{16}{100}$$ 4

(e) Gini Index for node 3

$$\text{Gini Index}(3) = \sum_{i=1}^{2} \sum_{\substack{j=1 \\ j \neq i}}^{2} p(\pi_i|3)\, p(\pi_j|3)$$

$$= 2\, p(\pi_1|3)\, p(\pi_2|3)$$

$$= 2 \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{2} \quad 3$$

(f) $g(3) = \frac{16}{100}$ (calculated in (d))

$$g(2) = \frac{R(2) - R(T_2)}{|\tilde{T}_2| - 1}$$

$$R(2) = p(2)\, r(2) = \frac{20}{100}$$

$$R(T_2) = \sum_{4,8,9} p(t)\, r(t) = p(8)\, r(8) + p(9)\, r(9)$$

$$= \frac{10}{100}$$

$$g(2) = \frac{\frac{20}{100} - \frac{10}{100}}{3-1} = \frac{5}{100} \quad 2\tfrac{1}{2}$$

$$g(5) = \frac{R(5) - R(T_5)}{|\tilde{T}_5| - 1} \; ; \; R(5) = p(5)\, r(5) = \frac{10}{100}$$

$$R(T_5) = \sum_{8,9} p(t)\, r(t) = \frac{10}{100}$$

$$g(5) = 0 \quad 2\tfrac{1}{2}$$

$$g(1) = \frac{R(1) - R(T)}{|\tilde{T}| - 1}; \quad R(1) = p(1)r(1) = \frac{40}{100}$$

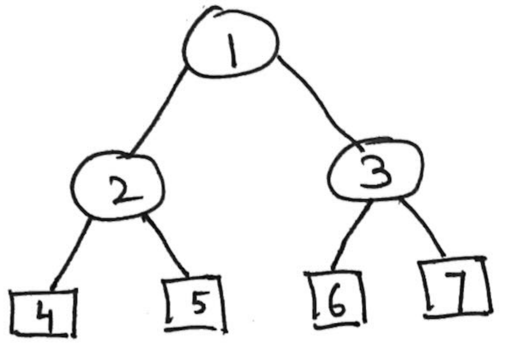
$$R(T) = \sum_{4,6,7,8,9} p(t)r(t) = \frac{14}{100}$$

$$g(1) = \frac{\frac{40}{100} - \frac{14}{100}}{4} = \frac{1}{4} \times \frac{26}{100} \quad 2\frac{1}{2}$$

$g(5) = 0$ is lowest

$\Rightarrow$ 5 is weakest link $\Rightarrow$ prune at 5

T_1



$2\frac{1}{2}$ – Give $2\frac{1}{2}$ if the answer is "prune at 5" without giving the subtree T_1

(Give part marking for minor calculation mistakes)

(5)

$$p(x) = \sum_{j=1}^{2} \pi_j \, p(x|\theta_j) \qquad p(x|\theta_j) = \begin{cases} \theta_j x^{\theta_j - 1}, & 0<x<1 \\ 0, & o/w \end{cases}$$

$$\Phi = (\pi_1, \pi_2, \theta_1, \theta_2)'$$

$\underset{\sim}{x}$: incomplete data w/o class lablels

$\underset{\sim}{y}' = (\underline{x}', \underline{z}')$

$\underset{\sim}{z}$: latent variable data

$$g(\underset{\sim}{y}_1, \ldots, y_n | \Phi) = \prod_{i=1}^{n} \left(\prod_{j=1}^{2} \left(p(x_i|\theta_j)\pi_j\right)^{z_{ji}} \right)$$

log likelihood

$$\log g = \sum_{i=1}^{n} \sum_{j=1}^{2} z_{ji} \log(p(x_i|\theta_j)) + \sum_{i=1}^{n} \sum_{j=1}^{2} z_{ji} \log \pi_j$$

2 (either g or log g)

E-Step : $$w_{ji} = E(z_{ji} | x_i, \Phi^{(m)})$$

$$= \frac{\pi_j^{(m)} p(x_i|\theta_j^{(m)})}{\sum_{j=1}^{2} \pi_j^{(m)} p(x_i|\theta_j^{(m)})} \quad \text{— } 2$$

Form the f^n

$$Q(\Phi, \Phi^{(m)}) = \sum_{i=1}^{n} \sum_{j=1}^{2} w_{ji} \log p(x_i|\theta_j) + \sum_{i=1}^{n}\sum_{j=1}^{2} w_{ji} \log \pi_j$$

— 2

M-Step Maximize Q w.r.t. π_j & θ_j

Max Q w.r.t. π_j subject to $\sum_1^2 \pi_j = 1$

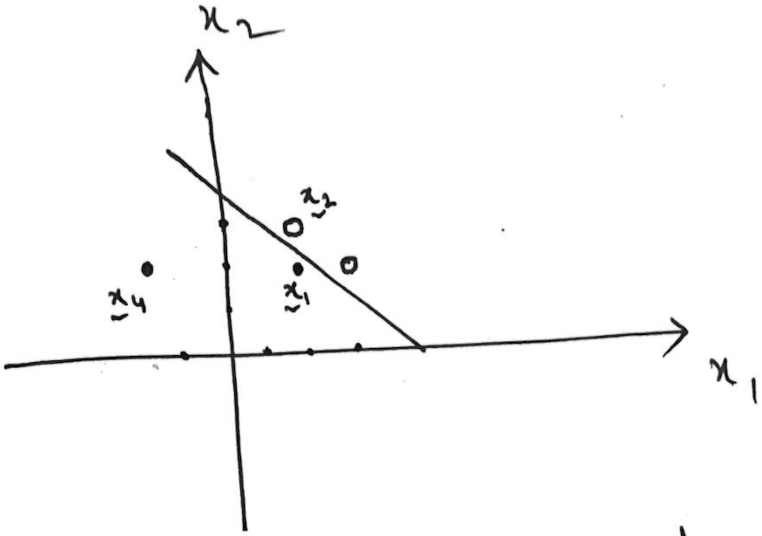
eq^n $$\hat{\pi}_j = \sum_1^n w_{ji}/n \quad \text{— } 3$$

$$\frac{\partial Q}{\partial \theta_j} = 0 \Rightarrow \sum_{i=1}^{n} w_{ji}\left(\frac{1}{\theta_j} + \log x_i\right) = 0$$

i.e. $$\hat{\theta}_j = -\frac{\sum_{i=1}^{n} w_{ji}}{\sum_{i=1}^{n} w_{ji} \log x_i} \quad \text{—③}$$

(6)

(a)



feature space

2

Training patterns are linearly separable

(b) If $\underline{v}'\underset{\sim}{y} > 0$ then it is correct classification

If $\underline{v}'\underline{y} < 0$ then misclassification

$$\underline{y}_i' = \begin{cases} \underset{\sim}{x}_i' & \text{If } \underset{\sim}{x}_i \in \pi_1 \\ -x_i' & \text{If } \underset{\sim}{x}_i \in \pi_2 \end{cases}$$

i.e. $\underset{\sim}{y}_1 = (1, 2)$

$\underline{y}_2 = (-2, -3)$

$\underline{y}_3 = (-3, -2)$

$\underset{\sim}{y}_4 = (-1, 2)$

updation If $\underline{v}'\underset{\sim}{y}_i < 0$ using

$\underline{v}_{K+1} = \underset{\sim}{v}_K + 0.5\, y_i$

$\underline{v}_0 = (1, 2)$ with cyclical instantaneous mode learning

$\underline{v}_0'\underset{\sim}{y}_1 = 5 > 0$ no updation for $\underset{\sim}{y}_1$ reqd.

$\underline{v}_0'\underset{\sim}{y}_2 < 0$ — updation reqd.

$\underset{\sim}{v}_0 \rightarrow \underset{\sim}{v}_1$

$$\underset{\sim}{v}_1 = \underset{\sim}{v}_0 + \frac{1}{2}\underset{\sim}{y}_2$$

$$= \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} -2 \\ -3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} -1 \\ -3/2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1/2 \end{pmatrix}$$

$\underset{\sim}{v}_1$ is new weight vector (updated) 3

$$\underset{-}{v}_1' \underset{-}{y}_3 = \left(0, \frac{1}{2}\right)\begin{pmatrix} -3 \\ -2 \end{pmatrix} = -1 < 0 \quad \text{- updation reqd.}$$

$\underset{\sim}{v}_1 \rightarrow \underset{-}{v}_2$

$$\underset{-}{v}_2 = \underset{\sim}{v}_1 + \frac{1}{2}\underset{\sim}{y}_3 = \begin{pmatrix} 0 \\ \frac{1}{2} \end{pmatrix} + \frac{1}{2}\begin{pmatrix} -3 \\ -2 \end{pmatrix}$$

$$= \begin{pmatrix} -3/2 \\ -\frac{1}{2} \end{pmatrix}. \quad 3$$

$\underset{-}{v}_2 = -\begin{pmatrix} 3/2 \\ 1/2 \end{pmatrix}$ is the weight vector after 2 steps of iteration