N. KAUSIK
CS21M037

Page No.
Date

Tutorial 10

1. 1) Precision $= \frac{TP}{TP+FP} = \frac{10}{10+5} = \frac{2}{3}$

2) Recall $= \frac{TP}{TP+FN} = \frac{10}{10+3} = \frac{10}{13}$

3) Accuracy $= \frac{TP+TN}{Total} = \frac{10+16}{10+16+5+3} = \frac{13}{17}$

4) Sensitivity $= \frac{TP}{TP+FN} = \frac{10}{13}$

5) Specificity $= \frac{TN}{TN+FP} = \frac{16}{16+5} = \frac{16}{21}$

2. i) C as it dominates other curves and it has largest area under curve.

3. Start

$C_1 \rightarrow (4,2)\ (3,2)\ (2,3)\ (-1,1)$

mean $C_1 = (2, 2)$

$C_2 \rightarrow (2,2)\ (-2,0)\ (-1,-1)$

mean $C_2 = \left(\frac{-1}{3}, \frac{1}{3}\right)$

For each point,

Point	dist from mean C_1	dist from mean C_2	
(4,2)	(2)	4.64	C_1
(3,2)	(1)	3.73	C_1
(2,3)	(1)	3.54	C_1
(-1,1)	3.16	(0.94)	C_2
(2,2)	(0)	2.87	C_1
(-2,0)	4.47	(1.69)	C_2
(-1,-1)	4.24	(1.49)	C_2.

Updated clusters are,

$C_1 \rightarrow (4,2)\ (3,2)\ (2,3)\ (2,2)$

$C_2 \rightarrow (-1,1)\ (-2,0)\ (-1,-1)$

Page No.
Date

4. i) $(-2,-2)$ $(-1,-2)$ $(2,1)$ $(1,2)$

Distances,

$(-2,-2) \leftrightarrow (-1,-2) \Rightarrow 1$

$(-2,-2) \leftrightarrow (2,1) \Rightarrow 5$

$(-2,-2) \leftrightarrow (1,2) \Rightarrow 5$

$(-1,-2) \leftrightarrow (2,1) \Rightarrow 4.24$

$(-1,-2) \leftrightarrow (1,2) \Rightarrow 4.47$

$(2,1) \leftrightarrow (1,2) \Rightarrow 1.41$

Avg dist $\Rightarrow$ $(-2,-2) = \boxed{\frac{11}{3}}$

$\Rightarrow (-1,-2) = \frac{9.71}{3}$

$\Rightarrow (2,1) = \frac{10.65}{3}$

$\Rightarrow (1,2) = \frac{10.88}{3}$

Max Avg L2 dist is for $(-2,-2)$

ii) $C_1 \rightarrow (-2,-2)$ mean $C_1 \rightarrow (-2,-2)$

$C_2 \rightarrow (-1,-2)\ (2,1)\ (1,2)$ mean $C_2 \rightarrow (1, 0.5)$

For pts in C_2,

Point	dist from mean C_1	dist from mean C_2	
$(-1,-2)$	(1)	3.2	C_1
$(2,1)$	5	(1.12)	C_2
$(1,2)$	5	(1.5)	C_2

Final $C_1 \rightarrow (-2,-2)\ (-1,-2)$

$C_2 \rightarrow (2,1)\ (1,2)$

iii) Heirarchical

Page No.
Date

5. Yes

6. b)

7. c)