# Seng 474 – Assignment 3

RYAN WOODWARD
V00857268

1a)

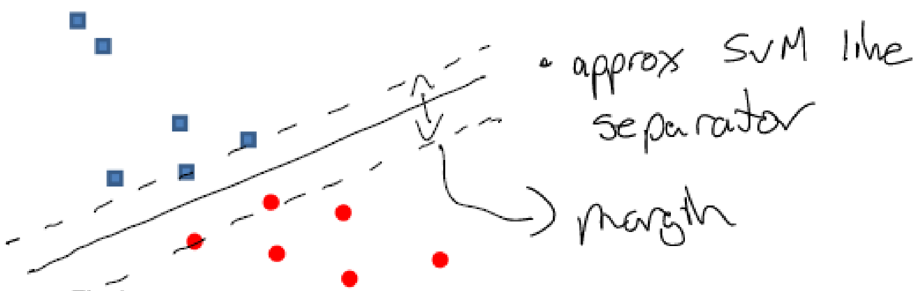


Fig. 1

b). If $\frac{1}{2} \cdot w^2 = 2$ in SVM optimization

$$w^2 = 4$$

$$w = 2$$

$$\|w\| = \sqrt{w \cdot w} = 2$$

$$\text{Margin} = \frac{1}{\|w\|} = \frac{1}{2}$$

c) With red points shifted further from the blue pts. The margin will be greater. As a result of this if $\frac{1}{\|w\|} > \frac{1}{2}$ this means

- $2 > \|w\|$

  $2 > \sqrt{w^2}$

  $4 > w^2$

  $2 > w$

- this means that the weight vector will have to be lower to maximize the margin.

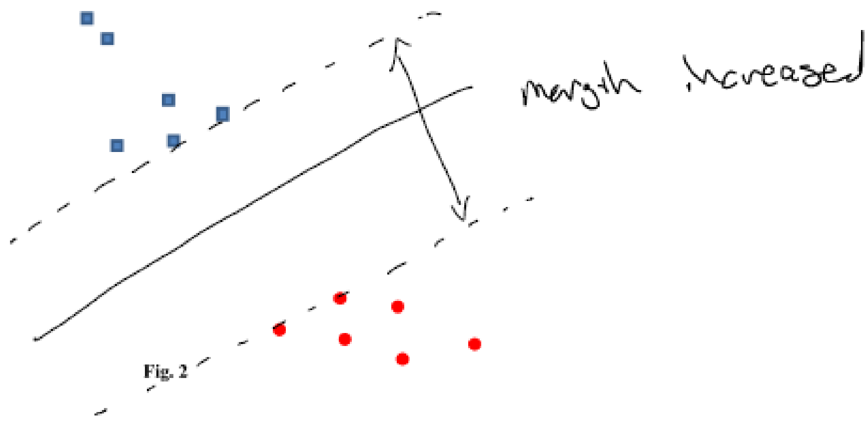



Fig. 2

d) Fig.1 = approx 1cm Margin

Fig.2 = approx 4 cm Margin

- $\frac{1}{\|w'\|} = 4$

  $\|w'\| = \frac{1}{4}$

  $\sqrt{(w')^2} = \frac{1}{4}$

  $(w')^2 = \frac{1}{16}$

  $\boxed{w' = \frac{1}{4}}$

e)

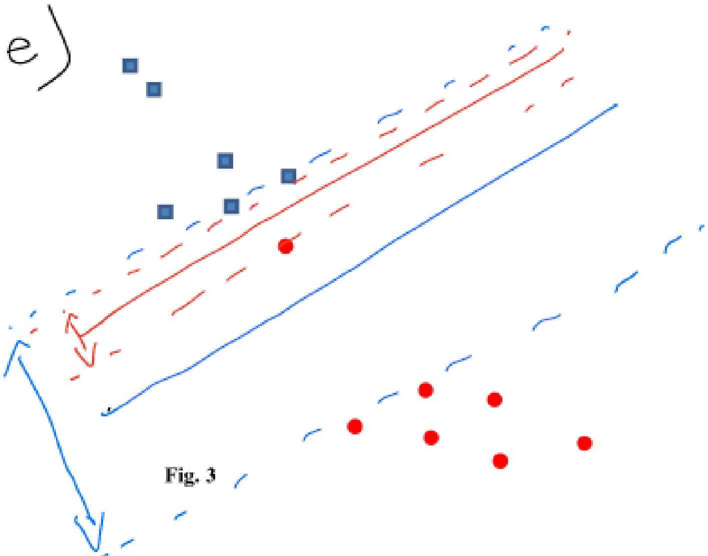


Fig. 3

- The red line that separates the points will have a higher cost than the blue line. This is due to the fact that $\xi$ for the red circle $\approx 1.5$
- $\xi = 1.5$ due to the fact that this new red circle exists with the margin and has a cost associated with it
- The cost of the blue line
$$= \frac{1}{2} \cdot (w)^2 + \xi$$
↳ the other $\xi$ will be zero due to the fact that they exist outside of the margin and don't have a cost associated with them.
- The cost of red line

$= \frac{1}{2} \cdot (w)^2$

↳ all of the $\xi$ are zero since they exist outside of the margin

Blue Line

- $\frac{1}{2} \cdot (w')^2 + \xi$

$= \frac{1}{2} \cdot \frac{1}{16} + 1.5 \approx \frac{49}{32}$

Red Line

- $\frac{1}{2} \cdot (w)^2$

$= \frac{1}{2} \cdot (4) = 2$

- Cost Blue < Cost Red

f)

- We can conclude that the blue line is a better SVM line separator since it has a lower cost

2. Take 100 records in the test set where 50 are positive and 50 are negative.

↳ 100% ⊕ records are classified correctly
0% ⊖ records " " correctly

↳ This error is 50/100 = 50%.

b) 80% ⊕ are classified correctly
20% ⊖ " "

$(50(80\%) + 50(20\%))/100 = 50\%$.

c) Of 99 records. 66 are positive, 33 are ⊖

100% ⊕ classified correctly
0% ⊖ " "

$1 - \{[66(100\%) + 33(0\%)]/99\} = 33\%$

d) 66% of + records are classified correctly
33% of − " "

$1 - [[66(2/3)] + [33(1/3)]] = 44.4\%$

3b) Confusion Matrix

Classifier A

| | | Predicted Class | |
|---|---|---|---|
| | | + | − |
| Actual class | +P | #TP 3 | #FN 2 |
| | −N | #FP 1 | #TN 4 |

$TPR = \frac{3}{5} = 0.6$

$FPR = \frac{FP}{N} = \frac{1}{5} = 0.2$

Classifier B

| | | Predicted Class | |
|---|---|---|---|
| | | + | − |
| Actual class | + | TP 1 | FN 4 |
| | − | FP 1 | TN 4 |

$TPR = \frac{1}{5} = 0.2$

$FPR = \frac{1}{5} = 0.2$

5) Classifier A

$\text{precision} = \frac{TP}{TP+FP} = \frac{3}{4} = 0.75$

$\text{recall} = \frac{TP}{P} = \frac{3}{5} = 0.6$

$\text{f-measure} = \frac{0.6+0.75}{2} = 0.675$

c) class B

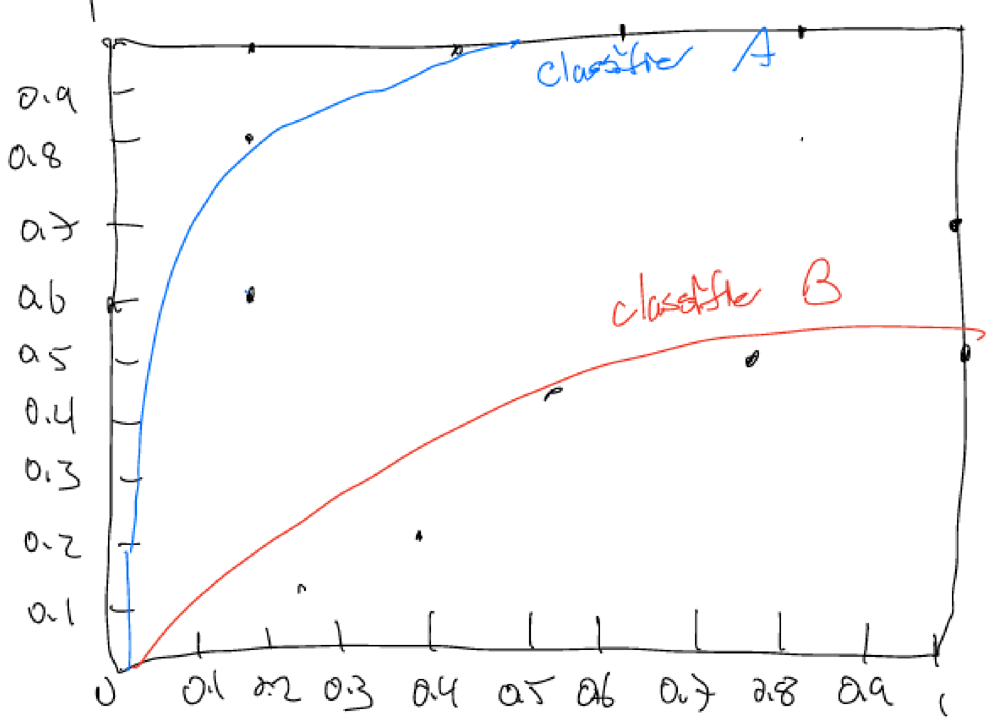
$\text{prec} = 0.5$

$\text{recall} = \frac{1}{5} = 0.2$

$\text{f-measure} = 0.35$

classifier A is better as it has a higher f-measure.

3a)




| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| pred | − | − | − | − | − | − | − | − | + | + |
| actual | − | − | − | − | + | + | − | + | + | + |

Class A

| | 0.08 | 0.15 | 0.35 | 0.44 | 0.45 | 0.47 | 0.55 | 0.67 | 0.69 | 0.73 |
|---|---|---|---|---|---|---|---|---|---|---|
| TP | 5 | 5 | 5 | 5 | 5 | 4 | 3 | 3 | 2 | |
| FP | 5 | 4 | 3 | 2 | 1 | 1 | 1 | 0 | 0 | |
| TN | 0 | 1 | 2 | 2 | 3 | 3 | 3 | 4 | 0 | |
| FN | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 2 | 3 | |
| TPR | 1 | 1 | 1 | 1 | 1 | 0.8 | 0.6 | 0.6 | 0.4 | |
| FPR | 1 | 0.8 | 0.6 | 0.4 | 0.2 | 0.2 | 0.2 | 0 | 0 | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| pred | − | − | − | | | | | + | + | − |
| actual | + | + | − | − | + | − | − | + | + | − |

| class B | 0.01 | 0.03 | 0.04 | 0.05 | 0.09 | 0.31 | 0.38 | 0.45 | 0.61 | 0.68 |
|---|---|---|---|---|---|---|---|---|---|---|
| TP | 4 | 3 | 2 | 2 | 2 | 1 | 1 | 0 | | |
| FP | 4 | 4 | 4 | 3 | 2 | 2 | 1 | 1 | | |
| TN | 1 | 1 | 1 | 1 | 2 | 2 | 3 | 3 | | |
| FN | 1 | 2 | 3 | 3 | 3 | 4 | 3 | 1 | | |
| TPR | 1 | 0.75 | 0.5 | 0.5 | 0.5 | 0.25 | 0.25 | 0 | | |
| FPR | 1 | 1 | 1 | 0.75 | 0.5 | 0.5 | 0.25 | 0.25 | | |

4.

nil

| F | F | F | F | F |
|---|---|---|---|---|
| a | b | c | d | e |

| F | I | F | F | F | F | F | F | I | F |
|---|---|---|---|---|---|---|---|---|---|
| a,b | a,c | a,d | a,e | b,c | b,d | b,e | c,d | c,e | d,e |

| \| | \| | \| | \| | \| | F | \| | \| | F | \| |
|---|---|---|---|---|---|---|---|---|---|
| a,b,c | a,b,d | a,b,e | a,c,d | a,c,e | a,d,e | b,c,d | b,c,e | b,d,e | c,d,e |

| N | N | I | I | I |
|---|---|---|---|---|
| a,b,c,d | a,b,c,e | a,b,d,e | a,c,d,e | b,c,d,e. |

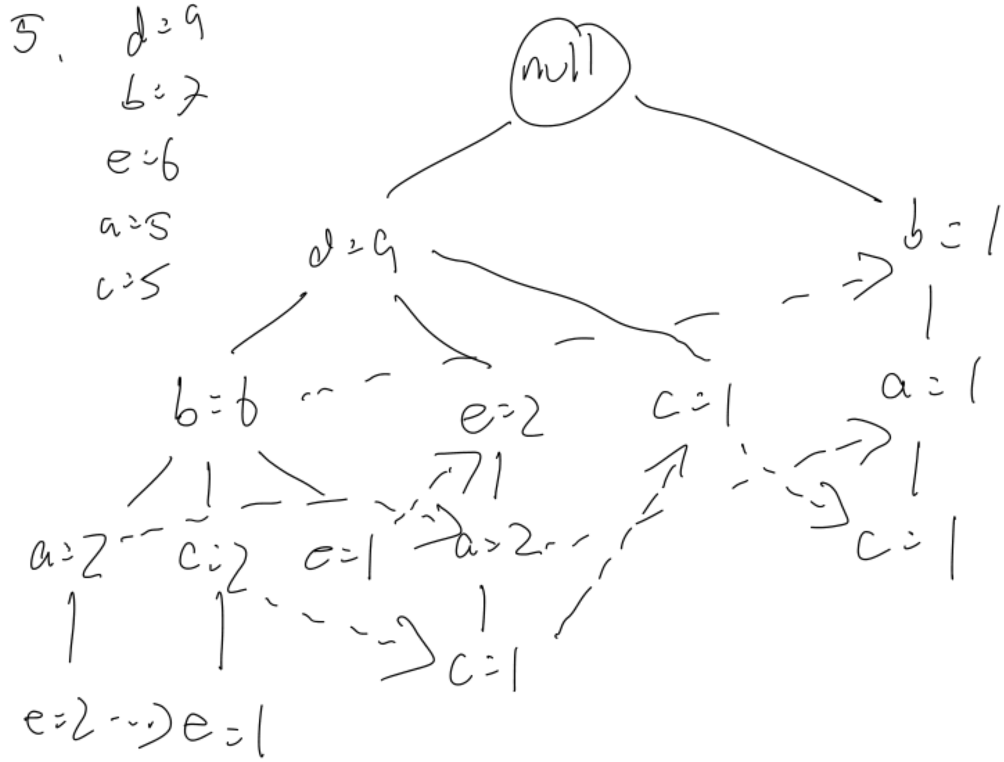
N
a,b,c,d,e

b) 10/31 = 32.3%.

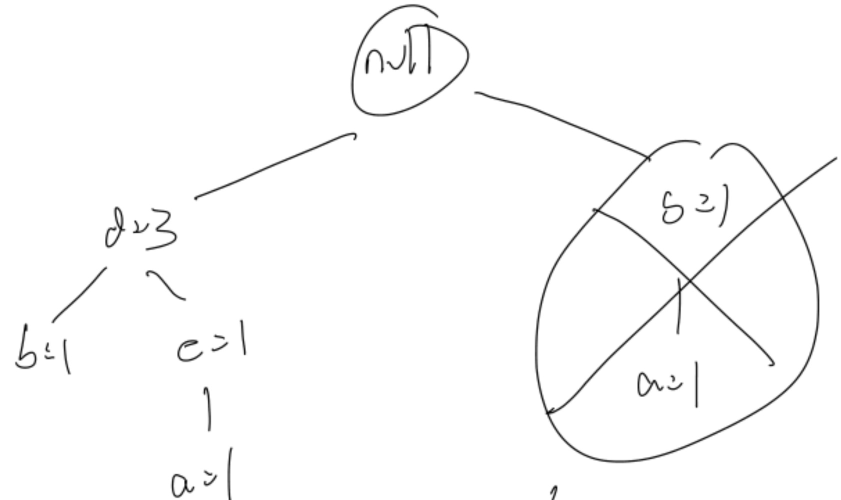
c) 21/31 = 67.7%.

d) 18/31 = 58.1%.

5.




FP-Tree Suffix "c"




↳ not supported enough
output "c"

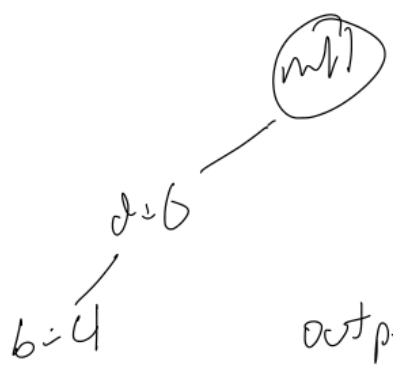
"ac", "bc", "dc", "ec" leave null trees and output ac, bc, dc, ec respectively.

FP Tree Suffix "a"


output a

"da" outputs da and leaves a null tree

FP Tree Suffix "e"




outputs e

F.P Tree Suffix "bc"


output bc, dbe

"de" outputs de

FP-Tree Suffix "b"

null

d:6

output b, db

FP-Tree Suffix "d"

Leaves null and outputs d

FI = a, b, c, d, e, ac, bc, dc, ec, da, be, de, be

6. MIN (p1), (p2)

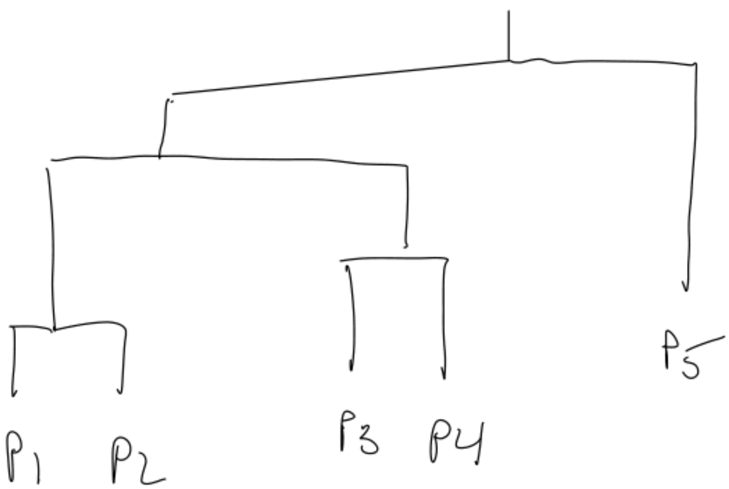
| | p1, p2 | p3 | p4 |
|---|---|---|---|
| p3 | 0.525 | | |
| p4 | 0.551 | 0.44 | |
| p5 | 0.665 | 0.85 | 0.76 |

MIN (P3), (P4)

| | P1, P2 | P3, P4 | P5 |
|---|---|---|---|
| P5 | 0.665 | 0.805 | x |
| P3P4 | 0.538 | x | |

MIN (P1, P2), (P3, P4)

MIN (P1, P2, P3, P4), (P5)




. MAX (P2), (P5)

| | P1 | P2, P5 | P3 | P4 |
|---|---|---|---|---|
| P1 | | 0.225 | 0.41 | 0.55 |
| P2, P5 | | | 0.745 | 0.615 |
| P3 | | | | 0.44 |

MAX $(p_2, p_5), (p_3)$

| | $p_1$ | $p_2, p_3, p_5$ | $p_4$ |
|---|---|---|---|
| $p_1$ | | 0.3175 | 0.55 |
| $p_2, p_3, p_5$ | | | 0.5275 |

MAX $(p_1, p_4)$
MAX $(p_2, p_3, p_5), (p_1, p_4)$