CSCE-5380 - DATA MINING SECTION-002

HOMEWORK-5

Santhoshi Karedla
11600006

1) Perceptron Model

Truth table for AND for pair of boolean variables

A	B	O/p
0	0	0
0	1	0
1	0	0
1	1	1

Let $w_1 = 1.2$ $w_2 = 0.6$ Threshold $= 1$

L.R $= 0.5$

$\Rightarrow$ A=0 , B=0 , target=0

$\Sigma w_i x_i = 0 * 1.2 + 0 * 0.6 = 0$

$\rightarrow$ Not greater threshold so O/p is '0'

$\Rightarrow$ A=0 B=1 target=0

$\Sigma w_i x_i = 0 * 1.2 + 1 * 0.6 = 0.6 \sim 0$

$\therefore$ O/p is '0'

$\Rightarrow$ A=1 B=0 target=0

$\Sigma w_i x_i = 1 * 1.2 + 0 * 0.6 = 1.2$

$\therefore$ O/p = 1 $\rightarrow$ this is greater threshold

So, now weights are modified based on perceptron rule.

So, $w_i = w_i + \eta(t - o)x_i$ ⇒ t = target, o = o/p

$w_1 = 1.2 + 0.5(0-1) = 0.7$

$w_2 = 0.8 + 0.5(0-1) = 0.6$

0.7 - 0.6 are modified weights

Restart the process from start

→ $w_1 = 0.7$, $w_2 = 0.6$ Thr = 1 LR(η) = 0.5

$w_i x_i = 0 * 0.7 + 0 * 0.6 = 0$

not greater than threshold ∴ o/p = 0

A = 0, B = 1 T = 0

$w_i x_i = 0 * 0.7 + 1 * 0.6 = 0.6$

not greater than threshold → o/p = 0.

A = 1, B = 0 ⇒ Target = 0

$w_i x_i = 1 * 0.7 + 0 * 0.6 = 0.7$

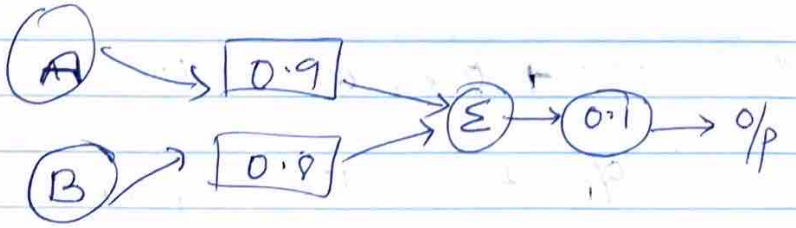
not greater than threshold so o/p = 0

A = 1, B = 1 target = 1

$1 * 0.7 + 1 * 0.6 = 1.3$

greater than 1 so o/p = 1

$w_1 = 0.7$, $w_2 = 0.6$ ⇒ weights



Approach 2

If we set bais value of -1.5

$$z = w_1 x_1 + w_2 x_2 + b$$

$$z = x_1 + x_2 - 1.5$$

$x_1 = x_2 = 1$

$1 + 1 - 1.5 = 0.5$

$\text{step}(0.5) = 1$

If $x_1 = 1 \quad x_2 = 0$

$1 + 0 - 1.5$

$= (0.5)$

$Y = \text{step}(-0.5) = 0.$

$\Sigma (x_1 + x_2 - 1.5) \rightarrow$ proves true for this condition.

AND gate

Approach 3

$w_1 = 0.5 \quad w_2 = 0.3 \quad b = -0.7 \quad x_1 = x_2 = 1$

$$z = w_1 * x_1 + w_2 * x_2 + b$$

$$= 0.5 * 1 + 0.5 * 1 - 0.7 = 0.3$$

$Y = \text{step}(z) = 1$

$\text{step}(0.3) = 1$

$\text{step}(x) = 0$ if $x < 0$

$= 1$ if $x > 0$

$x_1 + x_2 - 0.7 \rightarrow$ proves true for AND functions

3a) $x_1 \cdot x_2 \geq x_3 \cdot x_4$

$\rightarrow x_1 \cdot x_2 - x_3 \cdot x_4 \geq 0$.

$$F(x) = \sum_{i=1}^{n} w_i \phi(x).$$

$$x_1 x_2 - x_3 x_4$$

$$= 1 \times x_1 x_2 + x_3 x_4 \times 1 \quad -1 (x_1 x_2 + x_3 x_4) * 1)$$

$$= (1-1)\ (x_1 x_2 + x_3 x_4).$$

$w_1 = 1$, $w_2 = -1$

Feature Function $\phi_1(x) = x_1 x_2$ $\phi_2(x) = x_3 \cdot x_4$

Mathematical Expression.

$$\boxed{F(x) = x_1 x_2 - x_3 x_4}$$

b) Mathematical Expression for linear classifier of forms.

$F(x) = w \cdot \phi(x)$

w = weight $\phi(x)$ = Feature function.

$F(x) = -1\ (x_1 (>) x_2))$

$w = -1$ $\phi(x) = x_1 x_2$

$\rightarrow$ Hence we get $\boxed{F(x) = -x_1 \cdot x_2}$

But this case is not valid when $w = 1$

as we get $\boxed{F(x) = x_1 \cdot x_2}$

Which can't classify.

5) Classifier G

1) Threshold = 0.15

Predicted = + + + + + + + + + +

Actual = − − + − + − − + + +

TP = 5 FN = 0 FP = 5 TN = 0

$TPR = \frac{TP}{TP+FN} = 5/5+0 = 1$

$FPR = FP/FP+TN = 5/5+0 = 1$

2) Threshold = 0.2

Predicted = − + + + + + + + + +

Actual = − − + − + − − + + +

TP = 5 FN = 0 FP = 4 TN = 1

$TPR = TP/TP+FN = 5/5+0 = 1$

$FPR = FP/FP+TN = 4/5 = 0.8$

3) Threshold = 0.25

predicted = − − + + + + + + + +

Actual = − − + − + − − + + +

TP = 5 TN = 2 FP = 3 FN = 0

$TPR = TP/TP+FN = 5/5+0 = 1$

$FPR = FP/FP+TN = 3/5 = 0.6.$

4) Threshold = 0.37

Predicted = − − − + + + + + + +

Actual = − − + − + − − + + +

TP = 4 FP = 3 FN = 1 TN = 2

TPR = TP/TP + FN = 4/5 = 0.8

FPR = FP/FP+FN = 3/3+2 = 0.6.

5) Threshold = 0.41

Predicted = − − − − + + + + + +

Actual = − − + − + − − + + +

TP = 4 FP = 2 FN = 1 TN = 3

TPR = TP/TP+FN = 4/5 = 0.8

FPR = FP/FP+TN = 2/5 = 0.4

6) Threshold = 0.55

Predicted = − − − − − + + + + +

Actual = − − + − + − − + + +

TP = 3 FP = 2 FN = 2 TN = 3.

TPR = TP/TP+FN = 3/3+2 = 0.6

FPR = FP/FP+TN = 2/5 = 0.4

7) Threshold = 0.65

Predicted = − − − − − − + + + + TP = 3 FP = 1

Actual = − − + − + − − + + + FN = 2 TN = 4

TPR = TP/TP+FN = 3/5 = 0.6

FPR = FP/FP+TN = 1/5 = 0.2

8) Threshold = 0.8

Predicted = − − − − − − + + +

Actual = − − + − + − − + + +

TP = 3 FP = 0 FN = 2 TN = 5

TPR = TP / TP+FN = 3/3+2 = 0.6

FPR = FP / FP+TN = 0/0+5 = 0.

9) Threshold = 0.92

Predicted = − − − − − − − − + +

Actual = − − + − + − − + + +

TP = 2 FP = 0 FN = 3 TN = 5

TPR = TP / TP+FN = 2/2+3 = 0.4

FPR = FP / FP+TN = 0/0+5 = 0

Classifier 2

1) Threshold = 0.33

⇒ TP = 4 FP = 4 TN = 1 FN = 1

TPR = 0.8

FPR = 0.8

2) Threshold = 0.22

TP = 4 FP = 5 TN = 0 FN = 1

TPR = 0.8

FPR = 0.1

3) Threshold = 0.1

TP = 5 FP = 5 TN = 0 FN = 0

TPR = 0.1

FPR = 1

6) Threshold = 0.41

TP = 4 FP = 3 TN = 2 FN = 1

TPR = 0.8

FPR = 0.6.

7) Threshold = 0.68

TP = 3 FP = 1 TN = 2 FN = 1

TPR = 0.6

FPR = 0.2

8) Threshold = 0.59

TP = 4 FP = 2 TN = 3 FN = 2

TPR = 0.8 FPR = 0.4

9) Threshold = 0.72

TP = 2 FP = 1 TN = 4 FN = 3.

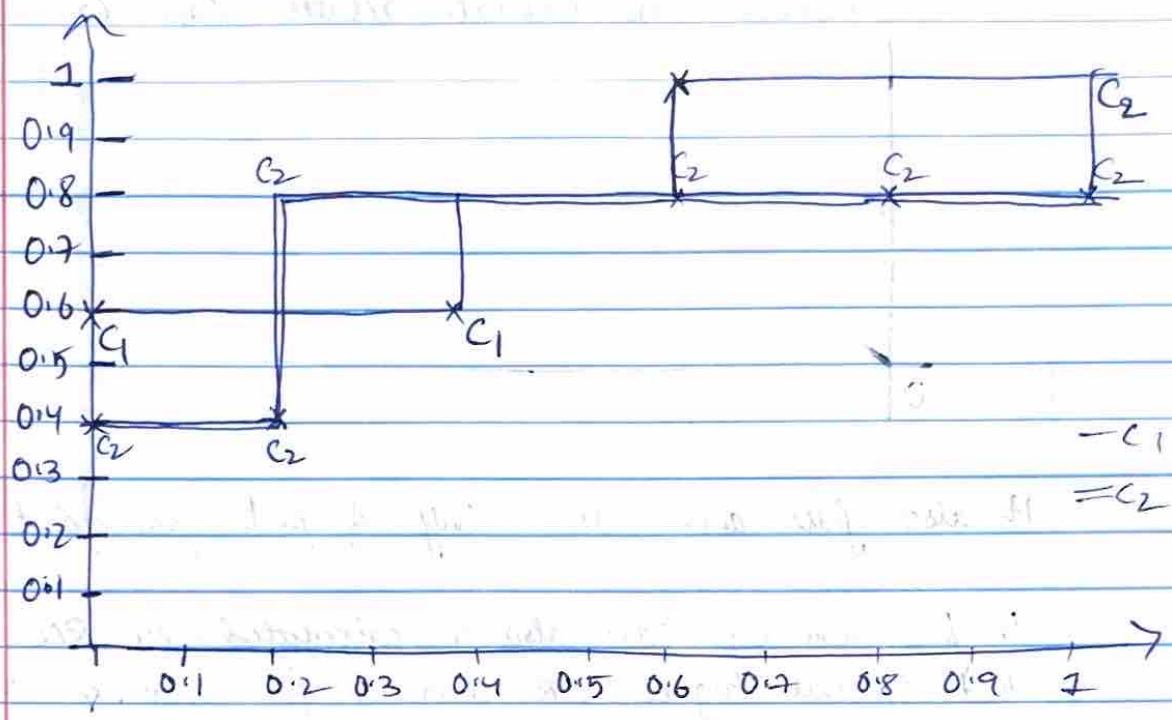
TPR = 0.4 FPR = 0.2

10) Threshold = 0.75

TP = 2 FP = 0 TN = 5 FN = 3

TPR = 0.4 FPR = 0.

a ROC of C_1 & C_2



b Area Under the Curve C_1 :

$\frac{1}{2} * \Sigma [i=1 \text{ to } n-1] (TPR_i + 1 - TPR_i) * (FPR_i + 1 + FPR_i)$

↑ TPRdiff ↑ FPRdiff

$\Rightarrow 0.6 \times 0.4 + 0.8 \times 0.2 + 1 \times 0.4$

$= 0.24 + 0.16 + 0.40 \Rightarrow 0.80$

$\Rightarrow$ **$C_1 \rightarrow 0.80$**

Area Under the Curve C_2 :-

$\Rightarrow 0.4 \times 0.2 + 0.8 \times 0.8$

$\Rightarrow 0.08 + 0.64$

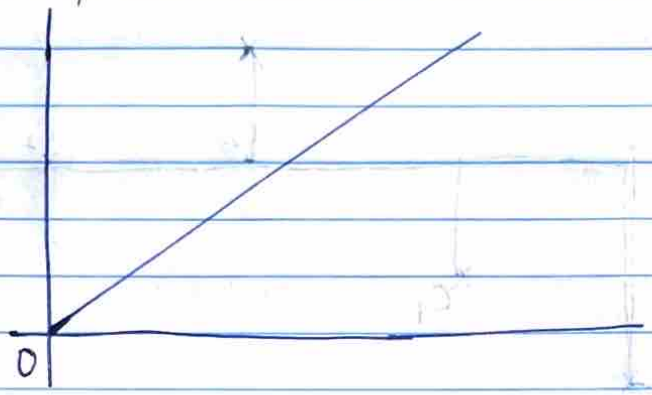
$\Rightarrow 0.72$

$C_2 = 0.72$

C $C_1 = 0.8$ $C_2 = 0.72$

Classifier 1 has more Area Under Curve

C_1 provides more accurate results than C_2



It also falls above lower half of plot from point (0,0) to (1,1)

In few cases C_2 can also be calculated as ROC curves that C_2 has higher TPR than C_1 for most of FPR values

Overall C_1 (or) C_2 either of them can be considered