Name : Areeka Aijaz
Roll no : 17K-3913

Assignment # 02

* Naive Bayes classifier

① Calculate prior probability

$P(Cheat = Yes) = 7/13$, $P(Cheat = No) = 6/13$.

② Calculate probability for Refund.

$P(Refund = Yes | Yes) = 4/7$, $P(Refund = No | Yes) = 3/7$
$P(Refund = Yes | No) = 3/6$, $P(Refund = No | No) = 3/6$.

③ Calculate probability for Status.

$P(Single | Yes) = 1/7$, $P(Single | No) = 3/6$
$P(Married | Yes) = 3/7$, $P(Married | No) = 2/6$
$P(Divorced | Yes) = 3/7$, $P(Divorced | No) = 1/6$

④ Calculate mean and st.dev.

$\mu_{class=Yes} = 100.71$, $\sigma^2_{class=Yes} = 1078.571$
$\mu_{class=No} = 74.16$ $\sigma^2_{class=No} = 64.166$

(5) Predict Label:

→ P(Yes , Single , 110) = ?

P(Income = 110 | Yes) = 0.0116

P(Income = 110 | No) = 2.24×10^{-6}.

P(Yes, Single, 110 | Yes) = $7/13 \times 4/7 \times 1/7 \times 0.0116$

= 0.000509

P(Yes, Single, 110 | No) = $6/13 \times 3/6 \times 3/6 \times 2.24 \times 10^{-6}$

= 2.584×10^{-7}

Hence P(Yes, Single, 110) can Cheat (Yes).

→ P(Yes, Single, 5) = ?

P(Income = 5 | Yes) = 1.7391×10^{-4}

P(Income = 5 | No) = 3.2406×10^{-18}

P(Yes, Single, 5 | No) = $4/13 \times 3/6 \times 3/6 \times 3.2406 \times 10^{-18}$

= 3.739×10^{-19}

P(Yes, Single, 5 | Yes) = $7/13 \times 4/7 \times 1/7 \times 1.7391 \times 10^{-4}$

= 7.644×10^{-5}.

PER312.

P(Yes, Single, 5) can Cheat(Yes)

→ P(Yes, Married, 95) = ?

P(Income = 95 | Yes) = 0.0119 , P(Income = 95 | No) = 0.0016

P(Yes, Married, 95 | Yes) = $7/13 \times 4/7 \times 3/7 \times 0.0119$ = 0.00156

P(Yes, Married, 95 | No) = $6/13 \times 3/6 \times 2/6 \times 0.0016$ = 0.07692

P(Yes, Married, 95) cannot Cheat (No)

→ P(No, Divorced, 63) = ?

P(Income = 63 | Yes) = 0.00628 , P(Income = 63 | No) = 0.01887

P(No, Divorced, 63 | Yes) = $7/13 \times 3/7 \times 3/7 \times 0.00628$ = 0.000621

P(No, Divorced, 63 | No) = $6/13 \times 3/6 \times 1/6 \times 0.01887$ = 0.000725

P(No, Divorced, 63) cannot Cheat (No).

* Decision Tree Classifier

– Using C4.5

① Calculate entropy of Cheat (Label) - (Yes-7 , No-6)

entropy ($7/13$, $6/13$) = $-7/13 \log(7/13) - 6/13 \log(6/13)$

= 0.2997.

② Calculate information gain
Total attributes 3 ie Refund, Status & Taxincome.

(i) Refund
- Yes - 7 (Yes - 4, No - 3)
- No - 6 (Yes - 3, No - 3)

$info([4,3]) = entropy(4/7, 3/7)$
$= -4/7 \log(4/7) - 3/7 \log(3/7)$
$info([4,3]) = 0.2965$

$info([3,3]) = entropy(3/6, 3/6)$
$= -3/6 \log(3/6) - 3/6 \log(3/6)$
$info([3,3]) = 0.3010$

$overall\ info([4,3],[3,3]) = 7/13 (0.2965) + 6/13 (0.3010)$
$= 0.2985$

$Gain(Refund) = 0.2997 - 0.2985$
$= 0.0012$

(ii) Status
- Single - 4 (Yes - 1, No - 3)
- Married - 5 (Yes - 3, No - 2)
- Divorced - 4 (Yes - 3, No - 1)

$info([1,3]) = entropy(1/4, 3/4)$
$= -1/4 \log(1/4) - 3/4 \log(3/4)$
$info([1,3]) = 0.2442$

PER 3/12.
PER 3/12.

info([3,2]) = entropy($^3/_5$, $^2/_5$)

$= -^3/_5 \log(^3/_5) - ^2/_5 \log(^2/_5)$

info([3,2]) = 0.2922

overall info([1,3],[3,2],[3,1])

$= ^4/_{13}(0.2442) + ^5/_{13}(0.2922) + ^4/_{13}(0.2442)$

$= 0.2626.$

Gain (Status) = 0.2997 − 0.2626.

= 0.0371.

(iii) Tax Income.

50	60	65	70	75	80	80	80	90	115	125	130	130
Yes	No	Yes	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes

	55		62		67		72		77		80		85		102		120		127	
	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>
Yes	1	6	1	6	2	5	2	5	2	5	2	5	2	5	3	4	4	3	5	2
No	0	6	1	5	1	5	2	4	3	3	6	0	6	0	6	0	6	0	6	0
Gain	0.2778		0.2994		0.2952		0.2991		0.2892		0.1502		0.1502		0.1913		0.2248		0.2531	

↳ $^1/_{13}(-^1/_1 \log(^1/_1) - 0\log 0) + ^{12}/_{13}(-^6/_{12}\log(^6/_{12}) - ^6/_{12}\log(^6/_{12}))$

(6)

– Using CART.

(i)

	Cheat
Yes	7
No	6

$$\text{Gini(Cheat)} = 1 - (7/13)^2 - (6/13)^2$$
$$= 0.4970$$

(ii)

	Refund	
	Yes	No
Yes	4	3
No	3	3

$$\text{Gini(Refund)} = \frac{7}{13}\left[1 - (4/7)^2 - (3/7)^2\right] + \frac{6}{13}\left[1 - (3/6)^2 - (3/6)^2\right]$$

$$\text{Gini(Refund)} = 0.4945$$

(iii)

	Status		
	Single	Married	Divorced
Yes	1	3	3
No	3	2	1

$$\text{Gini(Status)} = 4/13\left[1 - (1/4)^2 - (3/4)^2\right] + 5/13\left[1 - (3/5)^2 - (2/5)^2\right] + 4/13\left[1 - (1/4)^2 - (3/4)^2\right]$$

$$\text{Gini(Status)} = 0.4153$$

(iv) Taxable Income

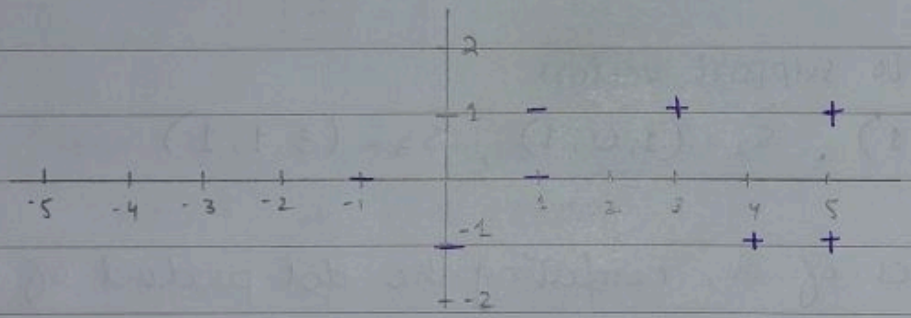
Cheat	Yes	No	Yes	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
	50	60	65	70	75	80	80	80	90	115	125	130	130

	55		62		67		72		77		80		85		102		120		127		130	
	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>
Yes	1	6	1	6	2	5	2	5	2	5	2	5	2	5	3	4	4	3	5	2	7	0
No	0	6	1	5	1	5	2	4	3	3	6	0	6	0	6	0	6	0	6	0	6	0
Gini	0.9230		0.4965		0.4871		0.5578		0.4730		0.2307		0.2307		0.3076		0.3692		0.4195		0.4970	

$$\frac{2}{13}\left(1-\left(\frac{1}{2}\right)^2-\left(\frac{1}{2}\right)^2\right)+\frac{11}{13}\left(1-\left(\frac{6}{11}\right)^2-\left(\frac{5}{11}\right)^2\right)$$

* Support Vector Machine (SVM)

1.



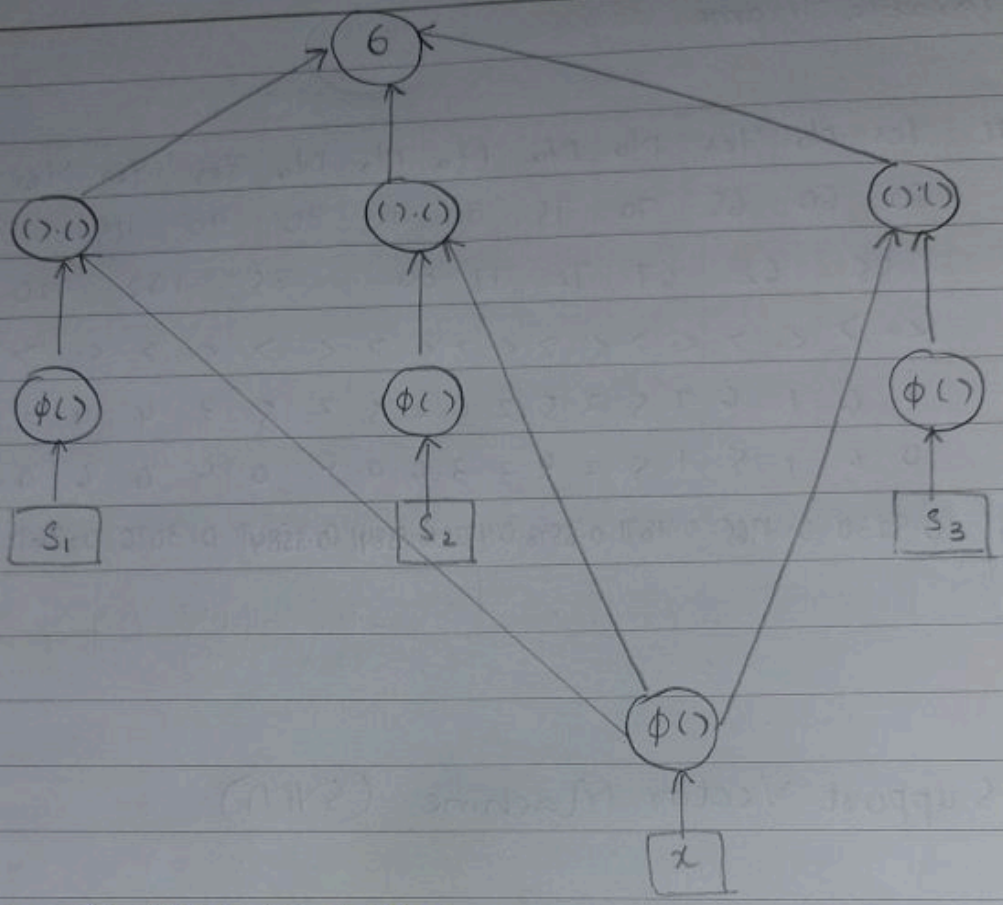
2. Support vectors be

$S_1 = (1,1) \rightarrow -ve$

$S_2 = (1,0) \rightarrow -ve$

$S_3 = (3,1) \rightarrow +ve$

3.



4 Adding bias to support vectors

$\tilde{s}_1 = (1,1,1)$, $\tilde{s}_2 = (1,0,1)$, $\tilde{s}_3 = (3,1,1)$

to find values of α, computing the dot product of the result

$$3\alpha_1 + 2\alpha_2 + 5\alpha_3 = -1$$
$$2\alpha_1 + 2\alpha_2 + 4\alpha_3 = -1$$
$$5\alpha_1 + 4\alpha_2 + 11\alpha_3 = 1$$

$\alpha_1 = -1.5$, $\alpha_2 = -2$, $\alpha_3 = 1.5$

Now, $\tilde{w} = \sum \alpha_i \tilde{s}_i$

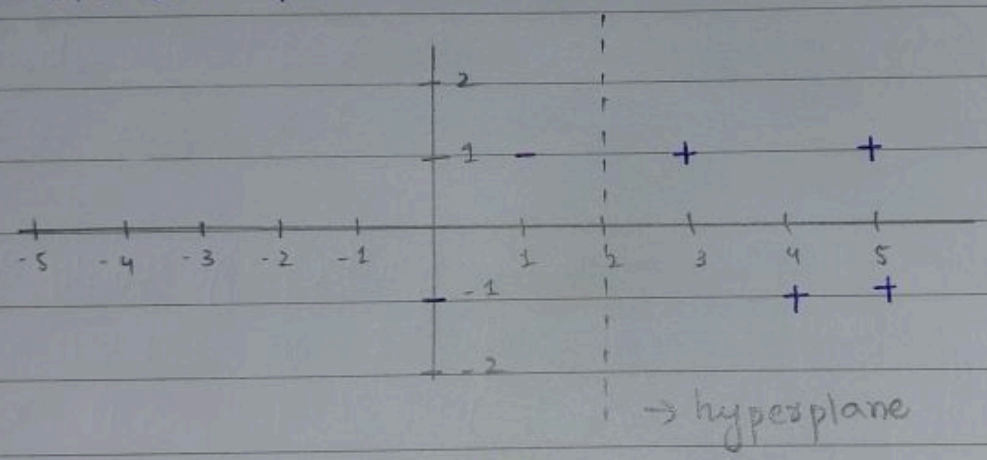
$$= -1.5\begin{pmatrix}1\\1\\1\end{pmatrix} - 2\begin{pmatrix}1\\0\\1\end{pmatrix} + 1.5\begin{pmatrix}3\\1\\1\end{pmatrix}$$

$$= \begin{pmatrix}1\\0\\-2\end{pmatrix}$$

As we have $y = wx + b$

so $w = (1, 0)$ and $b = -2$

We can plot line as;



5) From the above plot

(3, 2) → +ve class

(1, 5) → -ve class

(1, -1) → -ve class