

CSE 482 HW3 PART 2

1. (a)

Risk score = $0.2BP - 2.4Height - 5$
3.32
9.6
4
10.84
7.12

(b) Root mean square error = $\sqrt{\frac{1}{5}(3-3.32)^2 + (8.5-9.6)^2 + (2-0-4)^2 + (9-10.84)^2 + (4-0-7.12)^2}$
 $= \sqrt{\left(\frac{1}{5}\right)(18.4324)} = 1.92$

(c) The most important attribute for predicting risk score in the equation (Risk score = $0.2BP - 2.4Height - 5$) is Height because it has the highest absolute weight.

(d)

	Blood Pressure	Height
mean	129	5.76
Standard Deviation	18.276	0.413

(e) Risk score = $0.2(20.43Z_{BP} + 129) - 2.4(0.46Z_H + 5.76) - 5$
 $= 4.09Z_{BP} - 1.10Z_H + 6.98$

(f) Blood pressure has the highest absolute weight (4.09) so it is the most important.

(g) No, they are not consistent. I would use part (f) because it uses standardized values so outliers impact is reduced.

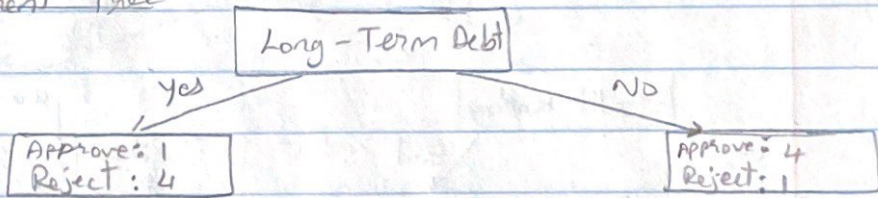
$$\text{Entropy} = - \sum p(i) \log p(i)$$

2.

		Class		Entropy (child)	Entropy Total
		Approve	Reject		
Long-Term Debt	Yes	1	4	0.7219	0.7219
	No	4	1	0.7219	
Unemployed	Yes	0	2	0	0.7635
	No	5	3	0.9544	
Credit Rating	Good	2	1	0.9183	0.9651
	Bad	3	4	0.9852	

← Best condition split

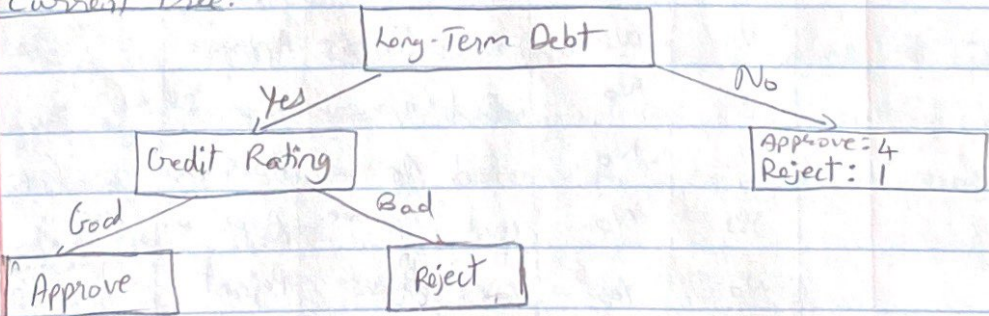
Current Tree



		Class		Entropy (child)	Entropy Total
		Approve	Reject		
Long-Term Debt = Yes	Yes	0	1	0	0.6490
	No	1	3	0.9113	
Credit Rating	Good	1	0	0	0
	Bad	0	4	0	

← Best condition split

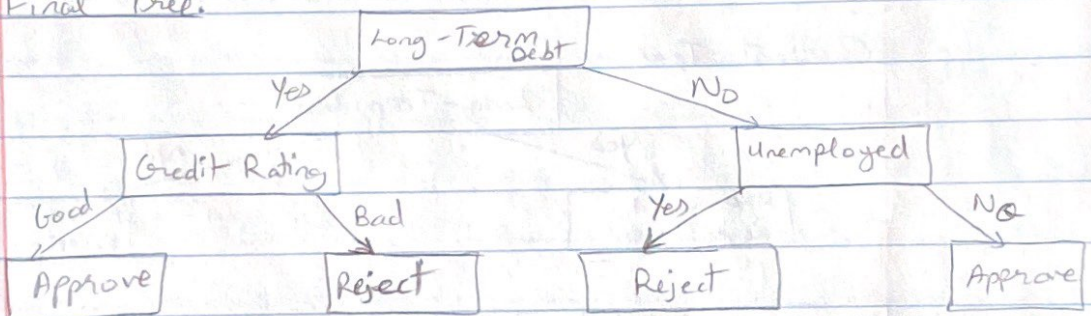
Current Tree:



Long-Term Debt = No		Class		Entropy (child)	Entropy Total
		Approve	Reject		
unemployed	Yes	0	1	0	0
	No	4	0	0	
Credit Rating	Good	1	1	1	0.4000
	Bad	3	0	0	

← Best condition split

Final Tree:



Long-Term Debt	unemployed	Credit Rating	Class (Actual)	Class (Predicted)
No	No	Good	Approve	Approve
No	No	Bad	Approve	Approve
No	No	Bad	Approve	Approve
No	No	Bad	Approve	Approve
Yes	No	Good	Approve	Approve
No	Yes	Good	Reject	Reject
Yes	No	Bad	Reject	Reject
Yes	No	Bad	Reject	Reject
Yes	No	Bad	Reject	Reject
Yes	Yes	Bad	Reject	Reject

$$\text{Accuracy} = 1$$

$$\text{Training Error} = 1 - 1 = 0$$

3. 20 unique ID's

	Positive	Negative
Left-Handed	9	1
Right-Handed	1	9

a) $Gini = 0$

b) $Gini = 1 - \left(\frac{1}{10}\right)^2 - \left(\frac{9}{10}\right)^2 = 0.18$

c) ID will be used as a splitting function

d) No, it is not reasonable to ID because it is just a ^{unique} number that is assigned to every entry so it has no predictive power. But handedness has predictive power.