Classmate

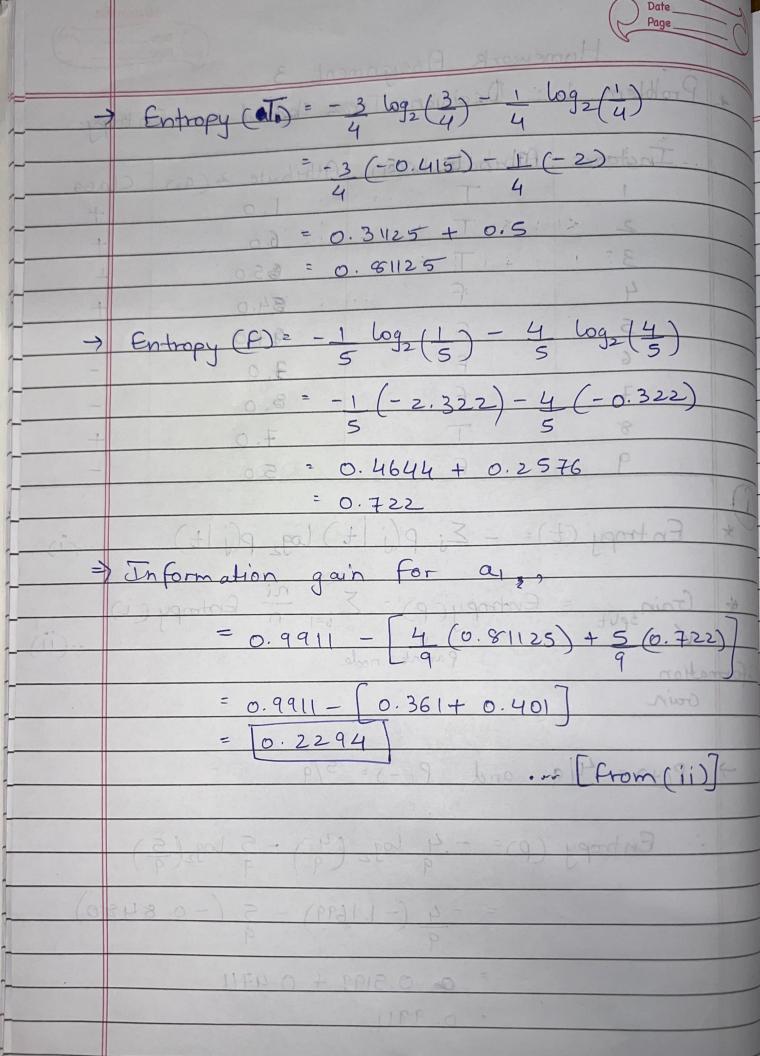
Date

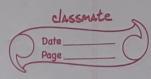
Page

| | Hor | nework Assignm | ent 3 | |
|--------|--------------------------------------|-----------------|------------------|--|
| * | Problem 1 | : Decision Tre | es based on F | in troopy |
| | | | | 7 |
| | Instance | Attribute (a) | Attribute 2 (as) | Class |
| | 1 | | 1.0 | + |
| | 2 | 3.0.T4 7.8VE | 6.0 | + |
| | 3 | SII25 T | \$ 5.0 | - |
| 100 | 4 | F | 84.0 | + |
| | (25 00) | 1 - FI 100 | \$ 7.0 | 3 + |
| 3 | 6 | F | 3.0 | - |
| | (-4322) | 1-(5(8,2-)) | 8.0 | - |
| | 8 | | 7.0 | + |
| 0 | 9 34 | 500 F 4494 | 5.0 | - |
| (1) | | - CSF. | 0 | |
| * | Entropy (t | = - Z; P(j 2 | t) log_p(j t) | (1) |
| | | n For al , | ke noth moda | 了传, |
| * | Gain = | Entropy(p) - 2 | i=1 n Entropy | (i) |
| - | 2 4 (D) 5 4 (| Parent nod | 11PPO = | (11) |
| Inform | ration Gain | | | |
| | Garc | 0 361+ 0 401 | -118903 | |
| | 0 41 | 10 | PSS 0 = | |
| 9 | r(+) = 110 | and PC-)= | 3/9 | |
| | | | | |
| | : Entropy (p)= -4 log (4) -5 log (5) | | | |
| | | | | |
| | | 9 (-1.1 | 699) - 5 (-0.8 | (480) |
| | | = 0 0 510 | 9 1 0 1711 | The state of the s |
| | | = 0.519 | 1 + 0,441 | |
| | | = 0,9911 | | |
| * | For -11. 14 | te 1 (a1), cour | it matrix | |
| U | agn bu | ie I (al), ww | | |
| | | | a, - | + |

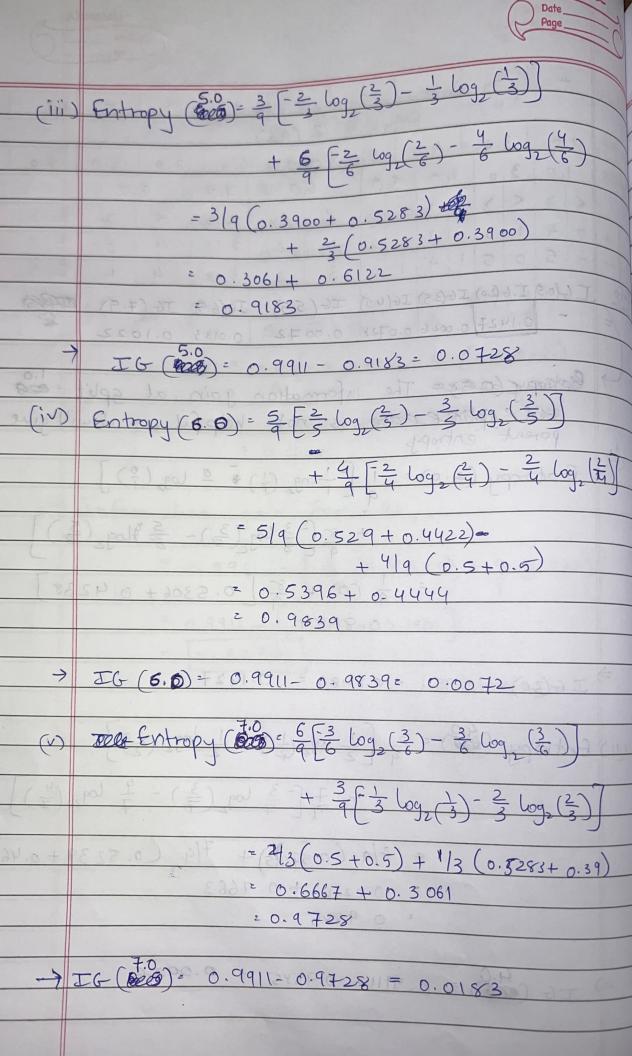
F

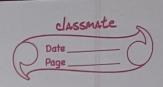
4





| Clease | + 12 + 2 + 3 + 3 + 3 + 3 + 3 + 3 + 3 + 3 + |
|--------|--|
| | Attribute 2 (a2) |
| Sorted | 1.0 3.0 4.0 5.0 5.0 6.0 7.0 7.0 8.0 |
| splits | 1.0 3.0 4.0 5.0 6.0 7.0 8.0 |
| | < > < > < > < > < > < > < > < > < > < > |
| + | 0 4 10 3 10 3 2 2 2 2 3 1 4 0 |
| | 050514143324150 |
| Info. | [6.5] I.6(2.0) I6(3.5) I6(4.5) I6(5.5) I6(6.5) I6(7.5) |
| | 0.14270.00260.0728 0.0072 0.0183 0.1022 |
| | 7 + 6 (200 - 10 9911 - 0 91 x = 0 0 72 x |
| 9 | Corporation gain at golit and |
| | porent entropy. |
| (3) | |
| 163 | Entropy (3.0) = 1 [-1 log (1) + 0 log (0)] |
| | Y |
| 724 | 9 (3) - 5 \$ log (5) |
| | |
| | - 1 (0) + 8 [0.5306 + 0.4238] |
| | = 0.8484 |
| 3 | The state of the s |
| | IG(3.0) = 0.9911 - 0.8484 = 0.1427 |
| 100 | 0 10 (4.0) - 5 (-11) |
| (11) | Entropy (4.0) = 2 (-1 log (1) - 1 log (2) |
| | |
| | (=) rol = (=) (=) (=) (=) (=) (=) (=) (=) (=) (|
| 1 | = 2/a (0.5+0/5) + 7/9 (0.5239 + 0.4613) |
| | |
| | 2 0.9885 |
| | |
| = | IG (808) = 0.9911 - 0.9885 2 0.0026 |
| | 16 (808) - 0.4111-0.1303 |





(vi) Entropy (8.0) = 8 = 4 log (4) - 4 log (4)

+ \frac{1}{9} [- ? log_ (?) - + log_2 (-1)]

 $= \frac{8}{9} \left[0.5 + 0.5 \right] + 0$ = 0.8889

) IG (8.0) = 0.9911-0.8889= 0.1022

st => 50, maximum information gain is at gplit= 3.0 which is 0.1427 for attribute 2 (a2)

Among a and az, a, puil be chosen as the first applitting for decision tree, as it has maximum inforgration gain.

2) I don't think that "Instance" althibute
should be used for a decision in the
thee, as for each new Instance, new
number will be assigned and hence it
has no predictive power.