

Problem 1 Entropy $H(S) = \frac{5}{14} \log_2 \frac{14}{5} + \frac{9}{14} \log_2 \frac{14}{9} = 0.940$

$$IG(S, Outlook) = 0.940 - \frac{5}{14} \times (\frac{3}{5} \log_2 \frac{5}{3} + \frac{2}{5} \log_2 \frac{5}{2}) - 0 - \frac{5}{14} \times (\frac{2}{5} \log_2 \frac{5}{2} + \frac{3}{5} \log_2 \frac{5}{3})$$

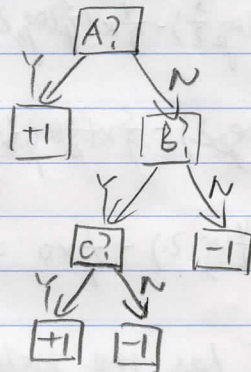
$$= 0.246$$

$$IG(S, Temperature) = 0.940 - \frac{4}{14} \times (\frac{2}{4} \log_2 \frac{4}{2} + \frac{2}{4} \log_2 \frac{4}{2}) - \frac{6}{14} \times (\frac{4}{6} \log_2 \frac{6}{4} + \frac{2}{6} \log_2 \frac{6}{3})$$

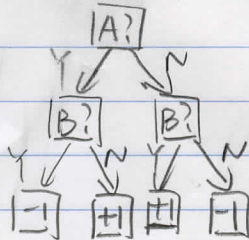
$$- \frac{4}{14} \times (\frac{3}{4} \log_2 \frac{4}{3} + \frac{1}{4} \log_2 4)$$

$$= 0.029$$

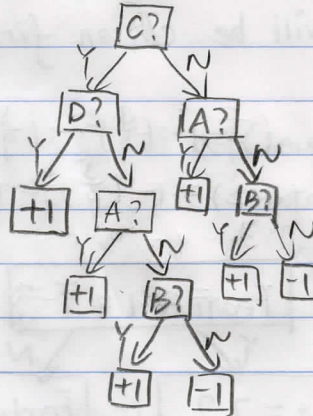
Problem 2 1. $AV(B \wedge C)$



2. $A \text{ XOR } B$



3. $(A \vee B) \vee (C \wedge D)$



Problem 3 1. Entropy $H(S) = \frac{4}{9} \log_2 \frac{9}{4} + \frac{5}{9} \log_2 \frac{9}{5} = 0.991$

$$2. IG(S, Feature1) = 0.991 - \frac{4}{9} \times (\frac{3}{4} \log_2 \frac{4}{3} + \frac{1}{4} \log_2 4) - \frac{5}{9} \times (\frac{1}{5} \log_2 \frac{5}{1} + \frac{4}{5} \log_2 \frac{5}{4})$$

$$= 0.229$$

$$IG(S, Feature2) = 0.991 - \frac{5}{9} \times (\frac{2}{5} \log_2 \frac{5}{2} + \frac{3}{5} \log_2 \frac{5}{3}) - \frac{4}{9} \times (\frac{1}{4} \log_2 \frac{4}{1} + \frac{3}{4} \log_2 \frac{4}{3})$$

$$= 0.007$$

3. ① threshold = 2.5

$$IG = 0.991 - \frac{1}{9} \times 0 - \frac{8}{9} \times \left(\frac{3}{8} \log_2 \frac{8}{3} + \frac{5}{8} \log_2 \frac{8}{5} \right) = 0.143$$

② threshold = 3.5

$$IG = 0.991 - \frac{2}{9} \times \left(\frac{1}{2} \log_2 2 + \frac{1}{2} \log_2 2 \right) - \frac{7}{9} \times \left(\frac{3}{7} \log_2 \frac{7}{3} + \frac{4}{7} \log_2 \frac{7}{4} \right) = 0.002$$

③ threshold = 4.5

$$IG = 0.991 - \frac{3}{9} \times \left(\frac{2}{3} \log_2 \frac{3}{2} + \frac{1}{3} \log_2 3 \right) - \frac{6}{9} \times \left(\frac{2}{6} \log_2 3 + \frac{4}{6} \log_2 \frac{6}{4} \right) = 0.073$$

④ threshold = 5.5

$$IG = 0.991 - \frac{5}{9} \times \left(\frac{2}{5} \log_2 \frac{5}{2} + \frac{3}{5} \log_2 \frac{5}{3} \right) - \frac{4}{9} \times \left(\frac{2}{4} \log_2 2 + \frac{2}{4} \log_2 2 \right) = 0.007$$

⑤ threshold = 6.5

$$IG = 0.991 - \frac{6}{9} \times \left(\frac{3}{6} \log_2 2 + \frac{3}{6} \log_2 2 \right) - \frac{3}{9} \times \left(\frac{1}{3} \log_2 3 + \frac{2}{3} \log_2 \frac{3}{2} \right) = 0.018$$

⑥ threshold = 7.5

$$IG = 0.991 - \frac{8}{9} \times \left(\frac{4}{8} \log_2 2 + \frac{4}{8} \log_2 2 \right) - \frac{1}{9} \times 0 = 0.102$$

Threshold value 7.5 has the highest information gain.

4. Feature 1 will be chosen first.

$$5. \text{Gini}(S, \text{Feature 1}) = 1 - \left(\frac{4}{9} \right)^2 - \left(\frac{5}{9} \right)^2 = 0.494$$

$$\text{Gini}(S, \text{Feature 2}) = 1 - \left(\frac{5}{9} \right)^2 - \left(\frac{4}{9} \right)^2 = 0.494$$

6. Gini(S, Feature 1)

