OPT. ASSMT 2

TASK 1

S: Sensor is in Sahara.

T_i: Temp on day i is > 80

$P(S) = 0.95 \qquad P(\neg S) = 0.05$

$P(T_i / S) = 0.9 \qquad P(T_i / \neg S) = 0.2$

$P(\neg T_i / S) = 0.1 \qquad P(\neg T_i / \neg S) = 0.8$

(a)

$$P(\neg S / T_1) = \frac{P(T_1 / \neg S)\, P(\neg S)}{P(T_1)}$$

$$= \frac{P(T_1 / \neg S)\, P(\neg S)}{P(T_1 / S) P(S) + P(T_1 / \neg S) P(\neg S)}$$

$$= \frac{0.2 \times 0.05}{0.9 \times 0.95 + 0.2 \times 0.05} = 0.01156$$

(b)

$$P\left(T_2 / T_1\right) = \frac{P(T_2 \wedge T_1)}{P(T_1)}$$

$$= \frac{P(T_2 \wedge T_1 \wedge S) + P(T_2 \wedge T_1 \wedge \neg S)}{P(T_1 \wedge T_2 \wedge S) + P(T_1 \wedge T_2 \wedge \neg S) + P(T_1 \wedge \neg T_2 \wedge S) + P(T_1 \wedge \neg T_2 \wedge \neg S)}$$

$$= \frac{P\left(T_2 / T_1 \wedge S\right) P\left(T_1 / S\right) P(S) + P\left(T_2 / T_1 \wedge \neg S\right) P\left(T_1 / \neg S\right) P(\neg S)}{P\left(T_1 / T_2 \wedge S\right) P\left(T_2 / S\right) P(S) + P\left(T_1 / T_2 \wedge \neg S\right) P\left(T_2 / \neg S\right) P(\neg S) + P\left(T_1 / \neg T_2 \wedge S\right) P\left(\neg T_2 / S\right) P(S) + P\left(T_1 / \neg T_2 \wedge \neg S\right) P\left(\neg T_2 / \neg S\right) P(\neg S)}$$

$$= \frac{P\left(T_2 / S\right) P\left(T_1 / S\right) P(S) + P\left(T_2 / \neg S\right) P\left(T_1 / \neg S\right) P(\neg S)}{P\left(T_1 / S\right) P\left(T_2 / S\right) P(S) + P\left(T_1 / \neg S\right) P\left(T_2 / \neg S\right) P(\neg S) + P\left(T_1 / S\right) P\left(\neg T_2 / S\right) P(S) + P\left(T_1 / \neg S\right) P\left(\neg T_2 / \neg S\right) P(\neg S)}$$

$$= \frac{0.9 \times 0.9 \times 0.95 + 0.2 \times 0.2 \times 0.05}{0.9 \times 0.9 \times 0.95 + 0.2 \times 0.2 \times 0.05 + 0.9 \times 0.1 \times 0.95 + 0.2 \times 0.8 \times 0.05}$$

$$= 0.7718$$

(c)

$$P(T_1 \wedge T_2 \wedge T_3)$$

$$= P(T_1 \wedge T_2 \wedge T_3 \wedge S) + P(T_1 \wedge T_2 \wedge T_3 \wedge \neg S)$$

$$= P(T_1 / T_2 \wedge T_3 \wedge S) P(T_2 / T_3 \wedge S) P(T_3 / S) P(S) + P(T_1 / T_2 \wedge T_3 \wedge \neg S) P(T_2 / T_3 \wedge \neg S) P(T_3 / \neg S) P(\neg S)$$

$$= P(T_1 / S) P(T_2 / S) P(T_3 / S) P(S) + P(T_1 / \neg S) P(T_2 / \neg S) P(T_3 / \neg S) P(\neg S)$$

$$= 0.9 \times 0.9 \times 0.9 \times 0.95 + 0.2 \times 0.2 \times 0.2 \times 0.05$$

$$= 0.69295$$

TASK 2

A takes 6 values

Each B_i takes 5 values.

Part a

JPD needs

$6^1 \times 5^{10}$ values

58,593,750 values

(In practice only need 58,593,749 values)

Part b

$$P(A, B_1, B_2, \dots B_{10}) = P(B_1/A)\, P(B_2/A) \dots P(B_{10}/A)\, P(A)$$

by condl. independance.

$P(A)$ needs 6 values (5 in practice)

$P(B_i/A)$ needs 6×5 values (6×4 in practice)

So in total we need.

$30 \times 10 + 6 = 306$ values

($24 \times 10 + 5 = 245$ values in practice)

TASK 3

$$P(A \mid B=f, E=t)$$

$$=$$

$$\langle P(A=t \mid B=f, E=t) \quad P(A=f \mid B=f, E=t) \rangle$$

$$\alpha \langle P(A=t, B=f, E=t) \quad P(A=f, B=f, E=t) \rangle$$

$$\alpha \langle P(A=t, B=f, C=t, D=t, E=t) + P(A=t, B=f, C=f, D=t, E=t) + P(A=t, B=f, C=t, D=f, E=t) + P(A=t, B=f, C=f, D=f, E=t)$$

$$P(A=f, B=f, C=t, D=t, E=t) + P(A=f, B=f, C=f, D=t, E=t) + P(A=f, B=f, C=t, D=f, E=t) + P(A=f, B=f, C=f, D=f, E=t) \rangle$$

$$\alpha \langle P(A=t)\, P(B=f)\, P(C=t \mid A=t, B=f)\, P(D=t \mid A=t)\, P(E=t \mid D=t, C=t)$$

$$+ P(A=t)\, P(B=f)\, P(C=f \mid A=t, B=f)\, P(D=t \mid A=t)\, P(E=t \mid D=t, C=f)$$

$$+ P(A=t)\, P(B=f)\, P(C=t \mid A=t, B=f)\, P(D=f \mid A=t)\, P(E=t \mid C=t, D=f)$$

$$+ P(A=t)\, P(B=f)\, P(C=f \mid A=t, B=f)\, P(D=f \mid A=t)\, P(E=t \mid C=f, D=f)$$

$$P(A=f)\, P(B=f)\, P(C=t \mid A=f, B=f)\, P(D=t \mid A=f)\, P(E=t \mid D=t, C=t)$$

$$+ P(A=f)\, P(B=f)\, P(C=f \mid A=f, B=f)\, P(D=t \mid A=f)\, P(E=t \mid D=t, C=f)$$

$$+ P(A=f)\, P(B=f)\, P(C=t \mid A=f, B=f)\, P(D=f \mid A=f)\, P(E=t \mid D=f, C=t)$$

$$+ P(A=f)\, P(B=f)\, P(C=f \mid A=f, B=f)\, P(D=f \mid A=f)\, P(E=t \mid D=f, C=f) \rangle$$

$$\alpha \langle\, 0.71 \times 0.31 \times 0.65 \times 0.84 \times 0.11 \quad 0.29 \times 0.31 \times 0.06 \times 0.76 \times 0.11$$

$$+ 0.71 \times 0.31 \times 0.35 \times 0.84 \times 0.19 \quad 0.29 \times 0.31 \times 0.94 \times 0.76 \times 0.19$$

$$+ 0.71 \times 0.31 \times 0.65 \times 0.16 \times 0.85 \quad 0.29 \times 0.31 \times 0.06 \times 0.24 \times 0.85$$

$$+ 0.71 \times 0.31 \times 0.35 \times 0.16 \times 0.62 \quad 0.29 \times 0.31 \times 0.94 \times 0.24 \times 0.62 \,\rangle$$

$$\alpha \langle\, 0.05261 \quad 0.02833 \,\rangle$$

$$\alpha = \frac{1}{0.07894}$$

$$\langle\, 0.6665 \quad 0.3335 \,\rangle$$

TASK 4

(a)

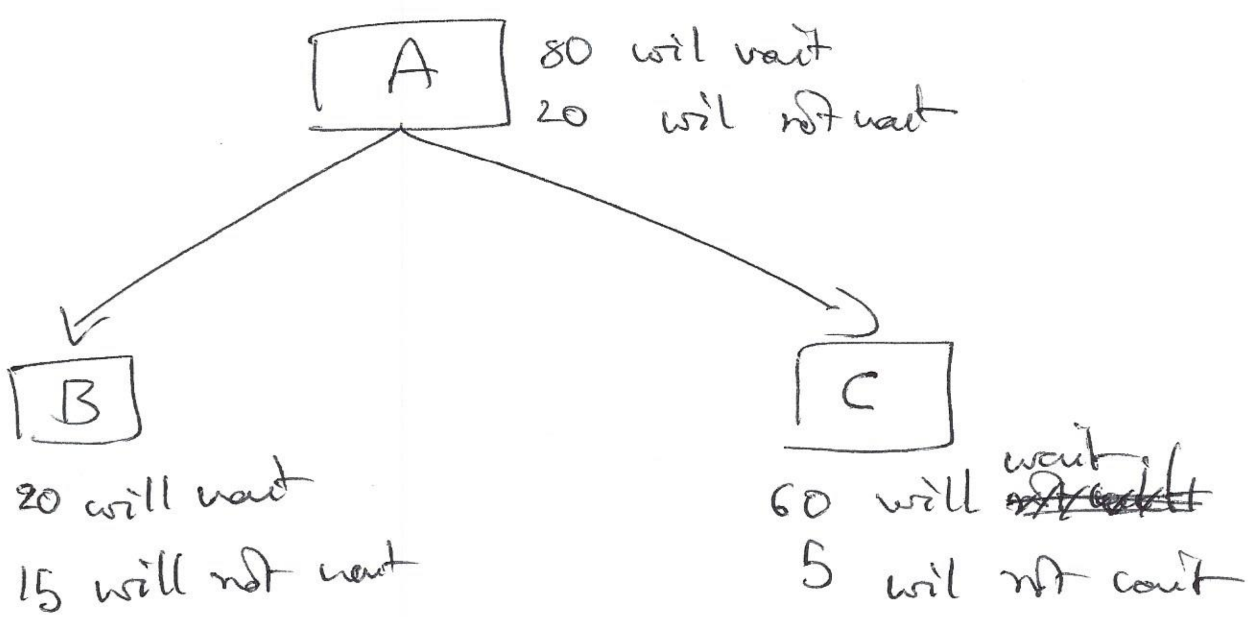
$\boxed{\text{Node A}}$

80 will wait
20 wil not wait

$$H_A = -\frac{80}{100} \log_2 \frac{80}{100} - \frac{20}{100} \log_2 \frac{20}{100}$$

$$= 0.7219$$

(b)



$H_A = 0.7219.$

$$H_B = -\frac{20}{35}\log_2\frac{20}{35} - \frac{15}{35}\log_2\frac{15}{35}$$

$$= 0.9852$$

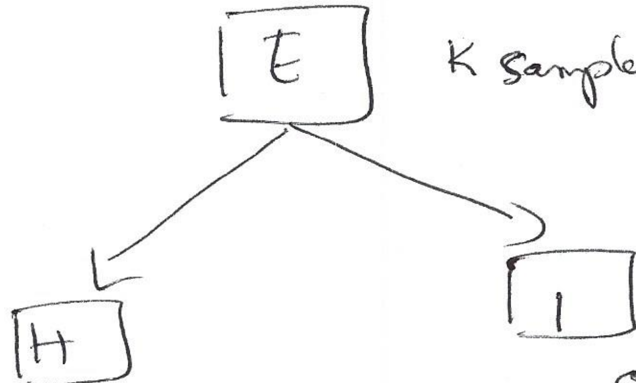
$$H_C = -\frac{60}{65}\log_2\frac{60}{65} - \frac{5}{65}\log_2\frac{5}{65}$$

$$= 0.3912$$

$$I = 0.7219 - \frac{35}{100}(0.9852) - \frac{65}{100}(0.3912)$$

$$= 0.1228$$

(c)



K samples with Entropy H

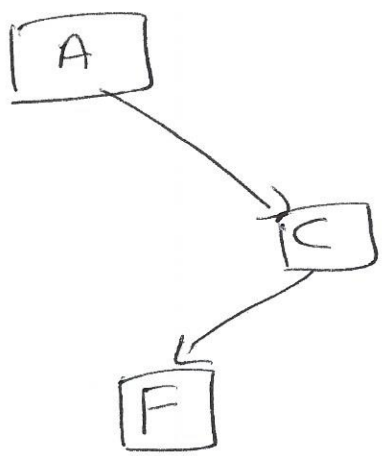
K sampls

0 sampls

$$I = H - \frac{K}{K}(H) - \frac{0}{K}(0)$$

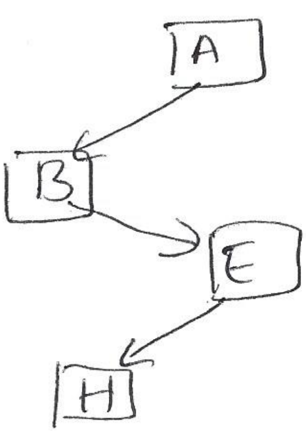
$$= 0$$

(d)



'will wait'

(e)



'will not wait'

TASK 5 Before Split: $\begin{matrix} 5x \\ 5y \end{matrix}$

A

$$H = -\frac{5}{10}\log_2\frac{5}{10} - \frac{5}{10}\log_2\frac{5}{10} = 1$$

A=1

$\begin{matrix} 3x \\ 0y \end{matrix}$

$$H_{A=1} = -\frac{3}{3}\log_2\frac{3}{3} - \frac{0}{3}\log_2\frac{0}{3} = 0$$

A=2

$\begin{matrix} 1x \\ 3y \end{matrix}$

$$H_{A=2} = -\frac{1}{4}\log_2\frac{1}{4} - \frac{3}{4}\log_2\frac{3}{4} = 0.8113$$

A=3

$\begin{matrix} 1x \\ 2y \end{matrix}$

$$H_{A=3} = -\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3} = 0.9183$$

$$I_A = 1 - \frac{3}{10}(0) - \frac{4}{10}(0.8113) - \frac{3}{10}(0.9183)$$

$$= 0.39999$$

B

B=1

1x
3y

$$H_{B=1} = -\frac{1}{4}\log_2\frac{1}{4} - \frac{3}{4}\log_2\frac{3}{4}$$

$$= 0.8113$$

B=2

3x
1y

$$H_{B=2} = -\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4}$$

$$= 0.8113$$

B=3

1x
1y

$$H_{B=3} = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}$$

$$= 1.$$

$$I_B = 1 - \frac{4}{10}(0.8113) - \frac{4}{10}(0.8113) - \frac{2}{10}(1)$$

$$= 0.15096$$

C

C=1

1 x
4 y

$$H_{c=1} = -\frac{1}{5}\log_2\frac{1}{5} - \frac{4}{5}\log_2\frac{4}{5}$$
$$= 0.7219.$$

C=2

3 x
1 y

$$H_{c=2} = -\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4}$$

$$= 0.8113$$

C=3

1 x
0 y

$$H_{c=3} = -\frac{1}{1}\log_2\frac{1}{1} - \frac{0}{1}\log_2\frac{0}{1}$$

$$= 0.$$

$$I_c = 1 - \frac{5}{10}(0.7219) - \frac{4}{10}(0.8113) - \frac{1}{10}(0)$$

$$= 0.31453$$

A is the attribute with highest Information gain.

A is the best attribute.

<u>TASK 6</u>

(a) Highest Entropy $= \log_2 N$

$= \log_2 4$

$= 2.$

Lowest Entropy $= 0$

(b) Highest Information gain. $= H - \sum_i \frac{k_i}{k} (0)$

$= H$

[H is entropy before split]

Lowest Information gain $= H - \sum_i \frac{k_i}{k} (H)$

$= 0$ //