Task 1:

a. Let $K1=80$, $K2=20$, $K=K1+K2=100$

$\therefore H(A)=H(K1/K, K2/K)=-K1/K \log K1/K - K2/K \log K2/K$

$= -0.8\cdot\log 0.8 - 0.2\cdot\log 0.2 = 0.7219$

b. Let $K_3=20$, $K4=15$, $K5=60$, $K6=5$.

$\therefore Gain_A = H(A)-\sum_i \frac{K_i}{K}\cdot H(A_i) = H(A)-[K3+K4/K\cdot H(K_3/K3+K4, K4/K3+K4) + K5+K6/K\cdot H(K5/K5+K6, K6/K5+K6)]$

$= 0.7219-[35/100\cdot H(20/35, 15/35) + 65/100\cdot H(60/65, 5/65)]$

$= 0.7219-[0.35\cdot 0.9852+0.65\cdot 0.3912]$

$= 0.1228$

c. Node E is a duplicate node. And people in node E they all those will not eat.

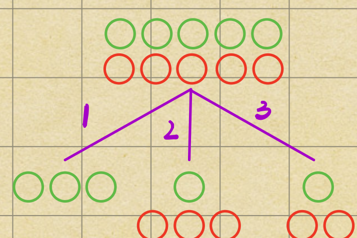
$Gain_E = H(E)-\sum_i \frac{K_i}{K} H(E_i) = 0-[0+0]=0$

d. $A \to B \to D$,

The test case will end up in Node D and output: will wait.

Task 2:

A:



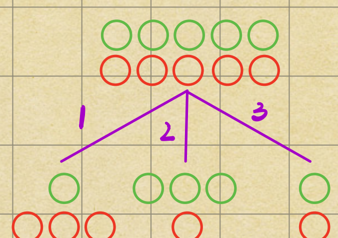
$Gain_A = H(A)-\sum_i \frac{K_i}{K} H(A_i)$

$= 1-[0+\frac{4}{10}(-\frac{1}{4}\log\frac{1}{4}-\frac{3}{4}\log\frac{3}{4})+\frac{3}{10}(-\frac{1}{3}\log\frac{1}{3}-\frac{2}{3}\log\frac{2}{3})]$

$= 1-[\frac{2}{5}(0.5+0.3112)+\frac{3}{10}(0.5283+0.3900)]$

$= 1-0.5997 \approx 0.4$

B:

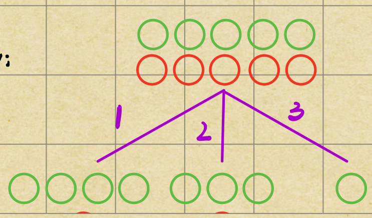


$Gain_B = H(B)-\sum_i \frac{K_i}{K} H(B_i)$

$= 1-[\frac{2}{5}(0.5+0.3112)+\frac{2}{5}(0.5+0.3112)+\frac{1}{5}\cdot 1]$

$= 1-0.649 = 0.15104$

C:



$Gain_C = H(C)-\sum_i \frac{K_i}{K} H(C_i)$

$= 1-[\frac{1}{2}(-\frac{4}{5}\log\frac{4}{5}-\frac{1}{5}\log\frac{1}{5})+\frac{2}{5}\cdot 0.8112+0]$

$= 1-[\frac{1}{2}\cdot 0.7219+\frac{2}{5}\cdot 0.8112] = 0.3148$

$\therefore$ Attribute A acheives the highest information gain at the root

Task 3:

a. Lowest entropy value : 0 , when all examples are from same class.

(A:1000, B,0, C:0, D:0, $H = -1 \cdot \log_2 1 - 0 - 0 - 0 = 0$

Highest entropy value when 4 classes have equal number (A:250, B:250, C:250, D:250)

where $H = -\frac{1}{4}\log\frac{1}{4} \cdot 4 = -\log\frac{1}{4} = 2$

b. Base on the answer of part a:

$Gain_{low} = 0-0-0-0-0 = 0$

$Gain_{high} = 2 - [\frac{1}{4} \cdot 4 \cdot 0] = 2$

Task 4:

Yes, reverse operation then we can obtain a 72% accuracy classifier.

Task 5:

1. calculate means and variances:

$\mu_{A_1} = {15+25}/{2} = 20 \quad \delta^2_{A_1} = \frac{1}{2-1}[(15-20)^2 + (25-20)^2 = 50$

$\mu_{A_2} = {28+32}/{2} = 30 \quad \delta^2_{A2} = \frac{1}{2-1}[(30-28)^2 + (32-30)^2] = 8$

$\mu_{B_1} = {20+32+25}/{3} = 25.67 \quad \delta^2_{B_1} = \frac{1}{3-1} \cdot [(20-25.67)^2 + (32-25.67)^2 + (25-25.67)^2] = 36.33$

$\mu_{B2} = {10+15+15}/{3} = 13.33 \quad \delta^2_{B_2} = \frac{1}{3-1}[(10-13.33)^2 + (15-13.33)^2 \cdot 2] = 8.33$

2. Calalate P(Attribute | Class) by $1/\sqrt{2\pi\delta^2} \cdot \exp(-{(x_i-\mu)^2}/{2\delta^2})$

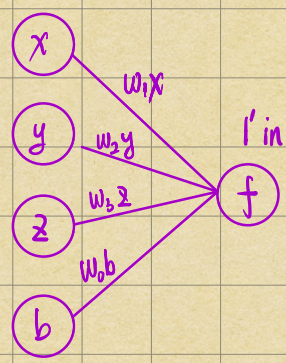
$P(A1|A) = 1/\sqrt{2\pi \cdot 50} \cdot \exp(-{x-20}/{100})$

$P(A2|A) = 1/\sqrt{2\pi \cdot 8} \cdot \exp(-{x-30}/{16})$

$P(A1|B) = 1/\sqrt{2\pi \cdot 36.33} \cdot \exp(-{x-25.67}/{72.66})$

$P(A2|B) = 1/\sqrt{2\pi \cdot 8.33} \cdot \exp(-{x-13.33}/{16.66})$

Task 6:



1' $in = w_1x + w_2y + w_3z + w_0b$ 2' $a = g(in)$

$\because$ $4X-7Y+2Z=6$, and we want the neural network alway return 1.

$\therefore$ We need to make sure $4X-7Y+2Z-6>0$ 0 0 3

According to the question, the bias input is +1 and we use "ReLu" as activation function

As a result, we can set $w_1=4$, $w_2=-7$, $w_3=2$, $w_0=1$. to simulate $4X-7y+2Z+1$, which is definitely larger than 0.