Task 1:

a) W: 65
Not-W: 35

$$H\left(\frac{65}{100}, \frac{35}{100}\right) = -\frac{65}{100}\log_2\frac{65}{100} - \frac{35}{100}\log_2\frac{35}{100}$$

$$= 0.9340$$

b) $k = 100$

$k_1 = 65 - W$

$k_2 = 35 - \text{not } W$

$k_3 = 25 - \text{week}, W$

$k_4 = 40 - \text{not week}, W$

$k_5 = 20 - \text{week}, \text{not } W$

$k_6 = 15 - \text{not week}, \text{not } W$

$$N_A = 0.9340 \qquad W_B = \frac{k_1 + k_5}{100} = 0.45$$

$$W_C = \frac{k_4 + k_6}{100} = 0.55$$

$$N_B = H\left(\frac{25}{45}, \frac{20}{45}\right) = 0.9907$$

$$N_C = H\left(\frac{40}{55}, \frac{15}{55}\right) = 0.8453$$

$$0.9340 - (0.45)(0.9907) - (0.55)(0.8453)$$

$$0.02307\ldots$$

$$\boxed{0.023}$$

c) ~~N_A = 1.20499~~
~~N_D = 1.02684~~

It is 0 because the test has already been done.

d) Node D

Node A -> Node B -> Node D

Outputs: will wait

Task 2:

A: 20

B: 18

C: 16

total: 54

info gain at root

$A: H\left(\frac{20}{54}, \frac{34}{54}\right) = -\frac{20}{54}\log_2\frac{20}{54} - \frac{34}{54}\log_2\frac{34}{54} = 0.9509$

$B: H\left(\frac{18}{54}, \frac{36}{54}\right) = -\frac{18}{54}\log_2\frac{18}{54} - \frac{36}{54}\log_2\frac{36}{54} = 0.9182$

$C: H\left(\frac{16}{54}, \frac{38}{54}\right) = -\frac{16}{54}\log_2\frac{16}{54} - \frac{38}{54}\log_2\frac{38}{54} = 0.8767$

A has the most information gain

Task 3:

a) Lowest is 0 (All Training examples are one class)

Highest is 8

because:

$$-\left(\frac{250}{1000}\log_2\frac{250}{1000}\right) = 2$$

done 4 times is 8

(All Training examples are evenly distributed between Attributes)

b)

Lowest: 0 (only 1 class type)

Highest: 1 (Even distribution of classes)

Task 4:

Increase the number of ~~fe~~ Attributes considered.

The more ~~information~~ Attributes used to make information gain the more accurate ~~the~~ the classifier.

Yes, by considering enough attributes to get .60+ information gain we can guarantee 60% accuracy.