Problem 1)

- $P(C_i)$: $P(\text{buys\_computer} = \text{"yes"}) = 12/20 = 0.6$
  $P(\text{buys\_computer} = \text{"no"}) = 8/20 = 0.4$

(a) $P(\text{age} <= 30 \mid \text{buys\_computer} = \text{"yes"}) = 3/12 = 0.25$
$P(\text{age} <= 30 \mid \text{buys\_computer} = \text{"no"}) = 4/8 = 0.5$
$P(\text{age}: 31...40 \mid \text{buys\_computer} = \text{"yes"}) = 5/12 = 0.4166$
$P(\text{age}: 31...40 \mid \text{buys\_computer} = \text{"no"}) = 1/8 = 0.125$
$P(\text{age} > 40 \mid \text{buys\_computer} = \text{"yes"}) = 4/12 = 0.3333$
$P(\text{age} > 40 \mid \text{buys\_computer} = \text{"no"}) = 3/8 = 0.375$

(b) $P(\text{student} \mid \text{buy\_computer} = \text{"yes"}) = 7/12 = 0.5384$
$P(\text{student} \mid \text{buy\_computer} = \text{"no"}) = 2/8 = 0.25$
$P(\text{not\_student} \mid \text{buys\_computer} = \text{"yes"}) = 5/12 = 0.4166$
$P(\text{not\_student} \mid \text{buys\_computer} = \text{"no"}) = 6/8 = 0.75$

(c) $P(\text{medium\_income} \mid \text{buys\_computer} = \text{"yes"}) = 5/12 = 0.4166$
$P(\text{medium\_income} \mid \text{buys\_computer} = \text{"no"}) = 3/8 = 0.375$

- $P(\overbrace{\text{buys\_computer} = \text{yes}}^{B} \mid \overbrace{>40, \text{medium\_income}, \text{student}}^{X})$

$$= \frac{P(X \mid B)\, P(B)}{P(X)}$$

$$= \frac{1}{P(X)} \times P(>40 \mid B) \times P(\text{medium\_income} \mid B) \times P(\text{student} \mid B) \times P(B)$$

$$= \frac{1}{P(X)} \times \frac{4}{12} \times \frac{5}{12} \times \frac{7}{12} \times \frac{12}{20} = \frac{1}{P(X)} \times \frac{7}{144}$$

- $P(\sim B \mid X) = \dfrac{P(X \mid \sim B)\, P(\sim B)}{P(X)}$

$$= \frac{1}{P(X)} \times \frac{3}{8} \times \frac{3}{8} \times \frac{2}{8} \times \frac{8}{20} = \frac{1}{P(X)} \times \frac{9}{640}$$

$$P(B|X) + P(\sim B|X) = 1$$

$$\Rightarrow \frac{1}{P(X)} \times \frac{7}{144} + \frac{1}{P(X)} \times \frac{9}{640} = 1$$

$$\Rightarrow P(X) = \frac{7}{144} + \frac{9}{640} = \frac{361}{5760} = 0.06267$$

$$\Rightarrow P(B|X) = \frac{7}{144} \times \frac{5760}{361} = \frac{280}{361} = 0.7756$$

⇒ Probability of customer with profile (>40, medium income, student) buying a computer from MicroShop is 0.7756

Problem 2)

- Class P: buys_computer = "yes"
- Class N: buys_computer = "no"

Expected info needed to classify tuple D:

$$Info(D) = I(10,5) = -\frac{10}{15}\log_2\left(\frac{10}{15}\right) - \frac{5}{15}\log_2\left(\frac{5}{15}\right) = 0.91829$$

Assume split by age:

Age	p_i	n_i
<=30	3	4
31..40	4	0
>40	3	1

Info needed after using "age" to split D

$$= Info_{Age}(D) = \frac{7}{15} I(3,4) + \frac{4}{15} I(4,0) + \frac{4}{15} I(3,1)$$

$$= \frac{7}{15}\left(-\frac{3}{7}\log_2\frac{3}{7} - \frac{4}{7}\log_2\frac{4}{7}\right) + \frac{4}{15}\left(-\frac{4}{4}\log_2\frac{4}{4} - 0\right) + \frac{4}{15}\left(-\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4}\right) = 0.6761$$

Hence, $Gain(age) = Info(D) - Info_{age}(D) = 0.24219$

↓ Assume split by Income:

Income	p_i	n_i
High	2	2
Medium	5	2
Low	3	1

Info needed after using Income to split D:

$$Info_{income}(D) = \frac{4}{15} I(2,2) + \frac{7}{15} I(5,2) + \frac{4}{15} I(3,1).$$

$$= \frac{4}{15}\left(-\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4}\right) + \frac{7}{15}\left(-\frac{5}{7}\log_2\frac{5}{7} - \frac{2}{7}\log_2\frac{2}{7}\right) + \frac{4}{15}\left(-\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4}\right)$$

$$= 0.88579.$$

Hence, $Gain(Income) = Info(D) - Info_{income}(D) = 0.0325$

Since $Gain(age) > Gain(income)$, age should be used as the split factor for first level. Gain info is 0.24219

Problem 3)

a) k = 1, distance = Euclidean, weight = uniform
Prediction of (h = 69, w = 155) = 0, i.e. male
Prediction of (h = 72, w = 160) = 0, i.e. male

b) k = 3, distance = Manhattan, weight = uniform
Prediction of (h = 69, w = 155) = 1, i.e. female
Prediction of (h = 72, w = 160) = 0, i.e male