Name: Venkata Subba Narasa Bharath, Meadam

SBU ID: 112672986

CSE - 512 (HW 4)

1)

(a) $P(x = red) = \frac{1}{2}$ $P(x = yellow) = \frac{1}{5}$

$P(x = blue) = \frac{1}{4}$ $P(x = black) = \frac{1}{20}$

$$\boxed{H(x) = -\sum_{x=x} Pr(x = x) \log_2 (Pr(x = x))}$$

$$= -\left(\frac{1}{2}\log_2\left(\frac{1}{2}\right) + \frac{1}{4}\log_2\left(\frac{1}{4}\right) + \frac{1}{5}\log_2\left(\frac{1}{5}\right) + \frac{1}{20}\log_2\left(\frac{1}{20}\right)\right)$$

$= 1.68$ bits

b) X = color of a sock randomly picked

Y = which drawer (up, or down)

P(top drawer) = 2 P(bootom drawer) — 1

P(top drawer) + P(bottom drawer) = 1 — 2

Solving the above 2 equations, we get

$$\boxed{\begin{array}{l} P(\text{top drawer}) = 2/3 \\ P(\text{bottom drawer}) = 1/3 \end{array}}$$

$$\boxed{H(X|Y) = - \sum_{X=x, Y=y} Pr(X=x, Y=y) \log_2 (Pr(X=x|Y=y))}$$

<u>calculating relevant Probabilities</u>

$Pr(X=x, Y=y) = Pr(X|Y) * Pr(Y)$

$Pr(X=red | Y=top) = 1$

$Pr(X=red, Y=top) = 1 \times \frac{2}{3} = \frac{2}{3}$

$Pr(X=red | Y=bottom) = 0$

$Pr(X=blue | Y=top) = 0$

$Pr(X=blue | Y=below) = \frac{1}{2}$

$Pr(X=blue, Y=below) = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$

$Pr(x = \text{yellow} \mid y = \text{top}) = 0$

$Pr(x = \text{yellow} \mid y = \text{below}) = 2/5$

$Pr(x = \text{yellow}, y = \text{below}) = \frac{2}{5} \times \frac{1}{3} = \frac{2}{15}$

$Pr(x = \text{black} \mid y = \text{top}) = 0$

$Pr(x = \text{black} \mid y = \text{below}) = \frac{1}{10}$

$Pr(x = \text{black}, y = \text{below}) = \frac{1}{10} \times \frac{1}{3} = \frac{1}{30}$

we, will assume, $0 \log_2(0)$ as 0

$$H(X|Y) = -\left[\frac{2}{3}\log_2(1) + \frac{1}{6}\log_2\left(\frac{1}{2}\right) + \frac{2}{15}\log_2\left(\frac{2}{5}\right) + \frac{1}{30}\log_2\left(\frac{1}{10}\right)\right]$$

$$= 0.45 \text{ bits}$$

c) Information gain:-

$$I(X:Y) = H(X) - H(X|Y)$$

$$= 1.68 - 0.45$$

$$= 1.23$$

2)

a) p(word = the) = 6/141

p(word = rabbit) = 3/141

p(word = a) = 5/141

i) Pr (current word = rabbit | Previous word = the)

$\Rightarrow \frac{\cancel{2}\,1}{\cancel{6}\,3} = \frac{1}{3}$

ii) Pr (current word = a | Previous word = rabbit)

$= \frac{1}{3}$

iii) Pr (current word = the | Previous word = rabbit)

$= \frac{2}{3}$

iv) Pr (current word = the or a | previous word = rabbit)

= 1

v) $P(A \cap B) = P(A) * P(B)$ -> condition for Naive Baye's

p(current word = the ∩ previous_word = rabbit)

$= \frac{2}{141}$

$p(\text{word} = \text{the}) = \frac{6}{141}$

$p(\text{word} = \text{rabbit}) = \frac{3}{141}$

$\Rightarrow \frac{2}{141} \neq \frac{6}{141} \times \frac{3}{141}$

$\boxed{\therefore \text{ Naive Bayes assumption is not valid here}}$