CS598: CXZ **Fall 2016**

Problem Set 2

chen zhang, NetId: czhang49 *Handed In: September 10, 2016*

1 Entropy

1.1 a

$$H(W) = -\sum_{x \in \Omega} P(x) \log_2 P(x) \tag{1}$$

The minimum happens when $P(W = w_k) = 1$, and $P(W = w_i) = 0$ for all $i \neq k$. Then equation (1) becomes

$$H(W) = -P(W = w_k) \log_2 P(W = w_k) = 0 \tag{2}$$

Theoretical maximum happens when W has a even distribution among all possible N words. Then we have

$$H(W) = -\sum_{x \in \Omega} P(x) \log_2 P(x) = -N \frac{1}{N} \log_2 \frac{1}{N} = \log_2 N \tag{3}$$

1.2 b

Article1 – minimum : w_1, w_1, w_1, w_1.
Article2 – maximum : w_1, w_2, w_3, w_4, w_5, w_6.

1.3 c

Since for A_1 and A_2, we all have $H(W) = 0$, then A_1 and A_2 are constructed with only one word, respectively. So when concatenating A_1 and A_2, maximum entropy happens if A_1 and A_2 are constructed with different words, and with equal length(minimum entropy happens when they are constructed with the same word).
Article A_1: w_1, w_1, w_1, w_1.
Article A_2: w_2, w_2, w_2, w_2.

2 Conditional Entropy and Mutual Information

2.1 a

$$H(X|X) = -\sum_{X \in \Omega} P(X|X) \log_2 P(X|X), \tag{4}$$

since $P(X|X) = 1$, then above equation becomes

$$H(X|X) = -N \times 1 \times \log_2 1 = 0 \tag{5}$$

2.2 b

We have the following equation:

$$I(X;Y) = -\sum_{x,y} P(x,y) \log_2 \frac{P(x,y)}{p(x)p(y)}, \tag{6}$$

if x and y are mutually independent, then we have $p(x,y) = p(x)p(y)$, plug in the above equation, we have

$$I(X;Y) = -\sum_{x,y} P(x)P(y) \log_2 \frac{P(x)P(y)}{p(x)p(y)} = -\sum_{x,y} P(x)P(y) \log_2 1 = 0. \tag{7}$$

We can also understand it in this way, since $I(X;Y) = H(X) - H(X|Y)$, if x and y are mutually independent, then $H(X|Y) = H(X)$, and then $I(X;Y) = 0$.

3 Mutual Information of Words (Programming Exercise)

3.1 a

$$P(X_A = 0, X_B = 1) = \frac{N_B - N_{AB}}{N} \tag{8}$$

$$P(X_A = 0, X_B = 0) = \frac{N - N_A - N_B + N_{AB}}{N} \tag{9}$$

3.2 b

paper january 181
language programming 153
january time 150
systems january 149
program january 149
data january 142
presented january 141
programming january 139
program programs 133
method january 125

3.3 c

3.3.1 i

program storage
programming method

programming time
data january
paper presents
language time
paper describes
intelligence artificial
output input
jr thacher

Many of the words are different, only (data january) is also in part (b).

3.3.2 ii

method
time
paging
polynomial
distribution

I think it makes sense, because programming is associated with these terms . (e.g. method in programming).