Ankit Agrawal : 2581532
Akshay Joshi : 2581346

SNLP Assn 5

1] Mutual Information

1.2 Mutual Info & Entropy :

→ Given: $f(o,q)$

To deduce: Amount of uncertainity about O that we can remove if the values of question Q & answer A are known.

i.e $\boxed{I(O;Q,A) = H(A|Q)}$ —— ⊛

• Hint : $H(x)=0$ if x is deterministic

[Reference: University of Cambridge : Info Theory & Coding lecture by J G Daugman]

• using the formula of mutual information

i.e $I(x,y) = H(x) - H(x|Y)$

applying this to LHS of eqn ⊛

$$I(O;Q,A) = H(O) - H(O|Q,A)$$
$$= \cancel{H(O)} + H(Q,A) - H(O|Q,A)$$
$$= H(Q,A) - H(O|Q,A)$$

• It is given in the question that, the answer is a deterministic function of object O & question Q, so we have to find out the entropy $H(O,Q)$.

WKT, entropy of 2 independent variable X & Y is computed as:

$$\boxed{H(X,Y) = H(X) + H(Y|X) \Leftrightarrow H(Y) + H(X|Y)}$$

Thus, $H(O,Q) = H(O) + H(O|Q)$

$$= H(O,Q) + (-H(O))$$

∴ using the previous values in $I(O;Q,A)$

i.e $I(O;Q,A) = H(O) + H(Q) - H(Q,A|O) + H(A|Q)$

$= \cancel{H(Q)} + H(\cancel{Q|O}) - H(A|\cancel{Q,O}) + \cancel{H(A|Q)}$

$= H(A|Q)$

From the above solution it is clear that the uncertainity about O that could be removed if values of Q & A are known is equal to the uncertainity of answer, given the question Q.

(1.1) Independence of variables

→ Assumption:

- $I(x_1;y) = 0$
- $I(x_2;y) = 0$

Is it correct that $I(x_1,x_2;y) = 0$ holds good?

Answer: NO!

if $I(x_1;y) = 0$ & $I(x_2;y) = 0$, it is not true that $I(x_1,x_2;y) = 0$ because if we expand these mutual info expressions we get the following.

WKT:

$$\boxed{I(x;y) = H(x) - H(x|y)} \quad \text{—— (1)}$$

- Applying (1) on $I(x_1;y)$, the resultant mutual info is 0 (as both x_1 & y are independent)

Illy, the mutual info b/n x_2 & y is also 0.

- But the values x_1 & x_2 are not mutually independent like $(x_1;y)$ & $(x_2;y)$. so on applying (1) on $I(x_1,x_2;y)$ we get:

(P.T.O)

$$I(X_1, X_2; Y) = H(X_1, X_2) - H(X_1, X_2 | Y)$$

$$= H(X_1, X_2) + H(Y) - H(X_1, X_2, Y)$$

$$= \sum_{x} p(x_1 x_2) \cdot \log \frac{1}{P(x_1 x_2)} + \sum_{y} p(y) \cdot \log \frac{1}{P(y)} +$$

$$\sum_{x,y} p(x_1 x_2, y) \log p(x_1 x_2, y)$$

$$= \sum_{x,y} p(x_1 x_2, y) \cdot \log \frac{p(x_1 x_2, y)}{p(x_1 x_2)\, p(y)}$$

$\therefore I(x_1; y) \neq I(x_1, x_2; y)$ &

$I(x_2; y) \neq I(x_1, x_2; y)$

so $I(x_1, x_2; y) \neq 0$

4) Bonus:

→ Given:

$Q = 6.5$

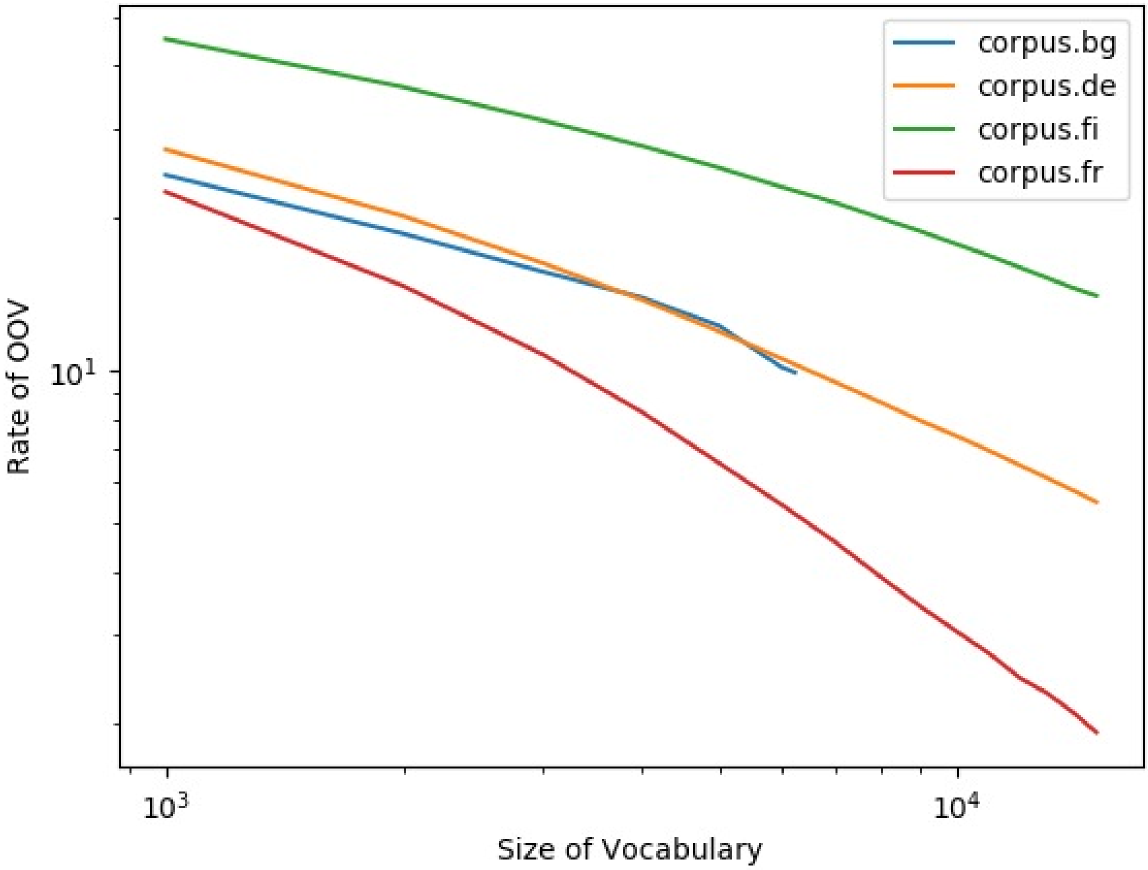
$A = y/n$

$I(Q, A; O) = ?$

$I(Q, A; O) = H(Q, A) - H(Q, A | O)$

- The search space can be reduced to half by just considering the true values of answer. All the questions (considering are good) which may lead to false answer can be pruned.
- The lower bound of the objects could be 0.5 as 6.5 questions are needed to guess each object & based on the pruned answer space this may perhaps be the possibility.

3. d


corpus.bg
corpus.de
corpus.fi
corpus.fr
Rate of OOV
10^1
10^3
10^4
Size of Vocabulary

2.1. Huffman Coding:

→ BBBDBACCDBDBACC.

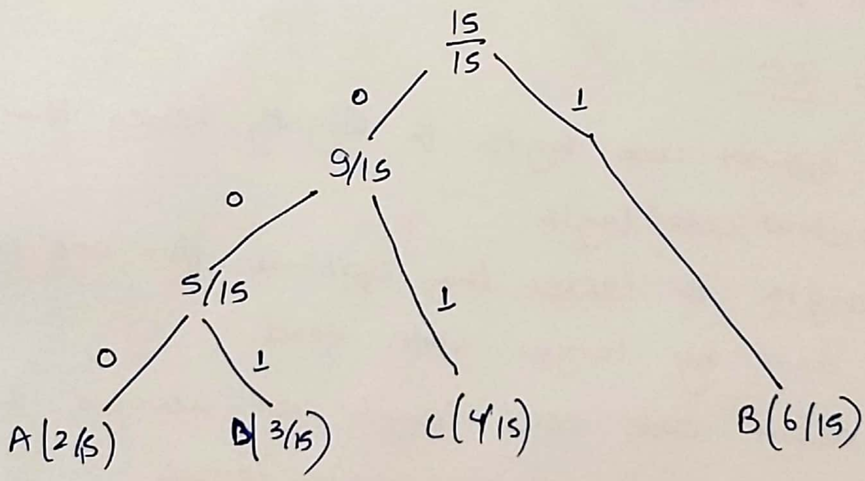
$P(A) = 2/15$

$P(B) = 6/15$

$P(C) = 4/15$

$P(D) = 3/15$

we join smallest first:



we get; codes,

$C(A) = 000$

$C(B) = 1$

$C(D) = 001$

$C(C) = 01$

<u>Encoded string</u>:

111 001 1 000 01 01 001 1 001 1 000 01 01 } total 29

optimal code length $l_i = -\log_2 (P(w_i))$.

	P_i	l_i	Actual length
A	2/15	2.90	3
B	6/15	1.32	1
C	4/15	1.90	2
D	3/15	2.32	3

$\Rightarrow$ optimal code length

According to optimal code length,

length of encoded string:

$\Rightarrow 2(A l_i) + 6(B) + 4(C) + 3(D)$

$\Rightarrow 2(2.906) + 6(1. \quad \ldots)$

$\Rightarrow 28.325$

$\simeq \underline{29}$

we see optimal code length is slightly lesser than the actual code length.

If codelength is lesser than optimal for one codeword, it is done by longer code word.

Hence actual code word length will always be greater than the optimal code length.

2.2

(a) Given

p	t	k	a	i	u
1/8	1/4	1/8	1/4	1/8	1/8

per letter entropy $(H(L)) \Rightarrow -\sum p_i \log_2(p_i)$

$\Rightarrow -\left[4 \cdot \frac{1}{8} \log_2(1/8) + 2 \cdot 1/4 \log_2(1/4)\right]$

$\Rightarrow -\left[3/2 - 1\right]$

$H(L) \Rightarrow$ ~~2½~~ $2\frac{1}{2}$ bits/letter

we take $\log_2$ to get in bits

(5)

	p	t	k	
a	1/16	3/8	1/16	1/2
i	1/16	3/16	0	1/4
u	0	3/16	1/16	1/4
	1/8	3/4	1/8	

now,

$$P(p|a) = \frac{P(p,a)}{P(a)} = \frac{1/16}{1/2} = 1/8$$

$P(t|a) = 3/4$

$P(k|a) = 1/8$

$P(p|i) = 1/4$

$P(t|i) = 3/4$

$P(k|i) = 0$

$P(p|u) = 0$

$P(t|u) = 3/4$

$P(k|u) = 1/4$

now,

entropy $H(L) = -\sum p_i \log_2(p_i)$

$$\Rightarrow \left[\frac{1}{16} \log_2 16 + \frac{3}{8} \log_2 (8/3) + \frac{1}{16} \log_2 (1/6) \right.$$

$H(L) \Rightarrow$ 2.28 bits/letter

now,

$$H(C|V) = -\sum_{x \in V} \sum_{y \in C} P(x,y) \log (y/x)$$

$$\Rightarrow -\left[p(p,a) \log[p(p|a)] + p(p,i) \log[p(p|i)] + p(p,u) \log[p(p|u)] + \ldots \right].$$

$$H(C|V) = -\left[\frac{1}{16} \log_2(1/8) + \frac{3}{8} \log(3/4) + 1/16 \log(1/8) + \frac{1}{16} \log 1/4 + 3/16 \log(3/4) + 0 + 0 + 3/16 \log 3/4 + 1/16 \log(1/4) \right]$$

$\underline{H(C|V) \Rightarrow 0.936 \text{ bits}}$.

also, $$H(V) = -\sum_{x \in V} P(x) \log_2 P(x)$$

$$\Rightarrow 1 \log_2 2 + \frac{1}{4} \log_2 4 + \frac{1}{4} \log_2 4$$

$$\Rightarrow 3/2 \text{ bits/letter}$$

now, joint entropy: $H(C,V) = H(C|V) + H(V)$

$$\Rightarrow 0.936 + 1.5$$

$$\Rightarrow \boxed{2.436 \text{ bits/letter}}$$

(c) Joint probability.

	p	t	k
a	1/32	1/16	1/32
i	1/64	1/32	1/64
u	1/64	1/32	1/64

$P(p, a) = P(p)\,P(a)$

Since they are assumed to be independant.

So we need to consider the conditional probability otherwise we donot get per syllable probabilies since the letters are not independent.