LING 572 – HW1

Q1(a)

$P(X_1) = \sum_Y P(X,Y) = 0.10 + 0.05 = 0.15$

$P(X_2) = \sum_Y P(X,Y) = 0.20 + 0.15 = 0.35$

$P(X_3) = \sum_Y P(X,Y) = 0.30 + 0.20 = 0.50$

Q1(b)

$P(Y_a) = \sum_X P(Y,X) = 0.10 + 0.20 + 0.30 = 0.60$

$P(Y_b) = \sum_X P(Y,X) = 0.05 + 0.15 + 0.20 = 0.40$

Q1(c)

$P(X_1|Y_a) = P(X_1,Y_a) / P(Y_a) = 0.10 / 0.60 = 0.167$

$P(X_1|Y_b) = P(X_1,Y_b) / P(Y_b) = 0.05 / 0.40 = 0.125$

$P(X_2|Y_a) = P(X_2,Y_a) / P(Y_a) = 0.20 / 0.60 = 0.333$

$P(X_2|Y_b) = P(X_2,Y_b) / P(Y_b) = 0.15 / 0.40 = 0.375$

$P(X_3|Y_a) = P(X_3,Y_a) / P(Y_a) = 0.30 / 0.60 = 0.5$

$P(X_3|Y_b) = P(X_3,Y_b) / P(Y_b) = 0.20 / 0.40 = 0.5$

Q1(d)

$P(Y_a|X_1) = P(X_1,Y_a) / P(X_1) = 0.10 / 0.15 = 0.667$

$P(Y_a|X_2) = P(X_2,Y_a) / P(X_2) = 0.20 / 0.35 = 0.571$

$P(Y_a|X_3) = P(X_3,Y_a) / P(X_3) = 0.30 / 0.50 = 0.6$

$P(Y_b|X_1) = P(X_1,Y_b) / P(X_1) = 0.05 / 0.15 = 0.333$

$P(Y_b|X_2) = P(X_2,Y_b) / P(X_2)$ = 0.15 / 0.35 = 0.429

$P(Y_b|X_3) = P(X_3,Y_b) / P(X_3)$ = 0.20 / 0.50 = 0.4

Q1(e)

X and Y are not independent.

Proof 1:

$P(X_2,Y_a) = 0.20$
$P(X_2)P(Y_a) = 0.35*0.60 = 0.21$

$P(X_2,Y_a) \neq P(X_2)P(Y_a)$

Proof 2:

$P(Y_a|X_2) = 0.571$
$P(Y_a) = 0.60$

$P(Y_a|X_2) \neq P(Y_a)$

Proof 3:

$P(X_2|Y_a) = 0.333$
$P(X_2) = 0.35$

$P(X_2|Y_a) \neq P(X_2)$

Q1(f)

$$
\begin{aligned}
H(X) &= -\sum_x P(x)\log P(x) \\
&= -(0.15\log(0.15) + 0.35\log(0.35) + 0.5\log(0.5)) \\
&= -(0.15(-2.737) + 0.35(-1.515) + 0.5(-1)) \\
&= 0.411 + 0.53 + 0.5 \\
&= 1.441 \textit{ bits}
\end{aligned}
$$

Q1(g)

$$
\begin{aligned}
H(Y) &= -\sum_y P(y)\log P(y) \\
&= -(0.6\log(0.6) + 0.4\log(0.4)) \\
&= -(0.6(-0.737) + 0.4(-1.322)) \\
&= 0.442 + 0.529
\end{aligned}
$$

$= 0.971\ bits$

Q1(h)

$H(X,Y) = -\sum_x \sum_y P(x,y) log P(x,y)$

When X=1,

$\sum_y P(x,y) log P(x,y)$ = 0.10log(0.10) + 0.05log(0.05) = -0.548

When X=2,

$\sum_y P(x,y) log P(x,y)$ = 0.20log(0.20) + 0.15log(0.15) = -0.875

When X=3,

$\sum_y P(x,y) log P(x,y)$ = 0.30log(0.30) + 0.20log(0.20) = -0.985

$H(X,Y)$ = - (-0.548 – 0.875 – 0.985)
= 2.408 *bits*

Q1(i)

$H(X|Y) = H(X,Y) - H(Y)$
= 2.408 – 0.971
= 1.437 *bits*

Q1(j)

$H(Y|X) = H(X,Y) - H(X)$
= 2.408 – 1.441
= 0.967 *bits*

Q1(k)

$MI(X,Y) = H(X) + H(Y) - H(X,Y)$
= 1.441 + 0.971 – 2.408
= 0.004 *bits*

Q1(l)

Part 1:

Hc(X,Y) = $-\sum_x \sum_y P(x,y) log Q(x,y)$

When X=1,

$\sum_y P(x,y) log Q(x,y)$ = 0.10log(0.10) + 0.05log(0.01) = -0.664

When X=2,

$\sum_y P(x,y) log Q(x,y)$ = 0.20log(0.20) + 0.15log(0.09) = -0.985

When X=3,

$\sum_y P(x,y) log Q(x,y)$ = 0.30log(0.40) + 0.20log(0.20) = -0.861

Hc(X,Y) = - (-0.664 – 0.985 – 0.861) = 2.51 *bits*

KL(P(X,Y) || Q(X,Y)) = Hc(X,Y) - H(X,Y)
= 2.51 – 2.408
= 0.102 *bits*

Part 2:

Find *KL(Q(X,Y) || P(X,Y));*

Hc(X,Y) = $-\sum_x \sum_y Q(x,y) log P(x,y)$

When X=1,

$\sum_y Q(x,y) log P(x,y)$ = 0.10log(0.10) + 0.01og(0.05) = -0.375

When X=2,

$\sum_y Q(x,y) log P(x,y)$ = 0.20log(0.20) + 0.09log(0.15) = -0.711

When X=3,

$\sum_y Q(x,y) log P(x,y)$ = 0.40log(0.30) + 0.20log(0.20) = -1.159

Hc(X,Y) = - (-0.375 – 0.711 – 1.159) = 2.245 *bits*

$H(X,Y) = -\sum_x \sum_y Q(x,y) log Q(x,y)$

When X=1,

$\sum_y Q(x,y) log Q(x,y)$ = 0.10log(0.10) + 0.01og(0.01) = -0.399

When X=2,

$\sum_y Q(x,y) log Q(x,y)$ = 0.20log(0.20) + 0.09log(0.09) = -0.777

When X=3,

$\sum_y Q(x,y) log Q(x,y)$ = 0.40log(0.40) + 0.20log(0.20) = -0.993

H(X,Y) = - (-0.399 – 0.777 – 0.993) = 2.169 *bits*

KL(Q(X,Y) || P(X,Y)) = Hc(X,Y) - H(X,Y)
= 2.245 – 2.169
= 0.076 *bits*

KL(P(X,Y) || Q(X,Y)) = 0.102 *bits*
KL(Q(X,Y) || P(X,Y)) = 0.076 *bits*

Hence, KL(P(X,Y) || Q(X,Y)) is not the same as KL(Q(X,Y) || P(X,Y)).

Q2(a)

H(X) = $-\sum_x P(x) \log P(x)$ = - (*plogp* + *(1-p)log(1-p)*)

Q2(b)

H(X) = $-\sum_x P(x) \log P(x)$ = - (*plogp* + *(1-p)log(1-p)*)

A uniform distribution would achieve the maxima value of *H(X).* For a coin, a uniform distribution would be *P(X = h)* = ½ and *P(X = t)* = ½.

H(X) = - (*0.5log0.5 + (1-0.5)log(1-0.5)) = - (-0.5 – 0.5) = 1*

Hence *p* = 0.5 would give the maximal value of *H(X)* = 1

Q2(c)

Proof

$H(X) = -(p\log p + (1-p)\log(1-p))$

The first derivative of $H(X) = -[\log p + p*(1/p) - \log(1-p) + (1-p)*1/(1-p)]$
$= -\log p + \log(1-p) - 2$

The second derivative of H(X) $= -\log p + \log(1-p) - 2$
$= -(1/p) + 1/(1-p)$

We have an inflection point when $-(1/p) + 1/(1-p) = 0$

Solving for p to find the maximal point;

$-(1/p) + 1/(1-p) = 0$
$1/(1-p) = 1/p$
$p = 1-p$
$2p = 1$
$p = ½$ *or* 0.5

Q3(a)

Number of color sequences = 10! / (5!*3!*2!) = 2520

Q3(b)

Total possible number of documents = $N!$

Let t be the count of word w, so that count of $w_1 = t_1$, $w_2 = t_2$, …, $w_i = t_i$

The number of documents that satisfy the condition stated in Q3(b)
$= N! / (t_1! * t_2! * t_3! \ldots * t_i!)$

Q3(c)

Probability of picking a word $w_i = P(w_i) = p_i$

Number of ways to get exactly $t_i = N! / t_i!$

Probability of getting exactly $t_i = (N! / t_i!) / N! = 1/ t_i!$

Probability where the occurrence of w_i is exactly t_i for each w_i

$= \sum_i p_i * (1 / t_i!)$

$= \sum_i p_i / t_i!$

Q4(a)

Trigram = $\prod_i P(w_i|t_i)P(t_i|t_{i-2},t_{i-1})$

Q4(b)

Each state in a trigram model corresponds to a tag pairs.

For example: *From State (IN, NN) => To State (NN, RB)*

If T is the number of tagset then the total number of states will be

= [T! / (T-2)!] + T

a_{ij} is the transition probability that corresponds to $P(t_i|t_{i-2},t_{i-1})$

b_{jk} is the emission probability that corresponds to $P(w_i|t_i)$

Q5(a)

Number of features will be = $3*|V| + |V|^2 + |T| + |T|^2$

Q5(b)

x is the feature vector that contains the current word.

y is the targeted POS tag based on x.

Q5(c)

Sentence = Mike/NN likes/VBP cats/NNS

Using binary value feature:

Mike NN prevW=<s> 1 curW=Mike 1 nextW=likes 1 surroundW=<s>+likes 1 prevT=BOS 1

likes VBP prevW=Mike 1 curW=likes 1 nextW=cats 1 surroundW=Mike+cats 1 prevT=NN 1 prevTwoTags=BOS+NN 1

cats NNS prevW=likes 1 curW=cats 1 nextW=</s> 1 surroundW=likes+</s> 1 prevT=VBP 1 prevTwoTags=NN+VBP 1

End of HW1 – submitted by Wee Teck Tan

Student ID: 0937003

Course Name: LING 572