**LING 572 – HW1**

**Q1(a)**

*P(X1)* = = 0.10 + 0.05 = 0.15

*P(X2)* = = 0.20 + 0.15 = 0.35

*P(X3)* = = 0.30 + 0.20 = 0.50

**Q1(b)**

*P(Ya)* = = 0.10 + 0.20 + 0.30 = 0.60

*P(Yb)* = = 0.05 + 0.15 + 0.20 = 0.40

**Q1(c)**

*P(X1|Ya)* = *P(X1,Ya) / P(Ya)* = 0.10 / 0.60 = 0.167

*P(X1|Yb)* = *P(X1,Yb) / P(Yb)* = 0.05 / 0.40 = 0.125

*P(X2|Ya)* = *P(X2,Ya) / P(Ya)* = 0.20 / 0.60 = 0.333

*P(X2|Yb)* = *P(X2,Yb) / P(Yb)* = 0.15 / 0.40 = 0.375

*P(X3|Ya)* = *P(X3,Ya) / P(Ya)* = 0.30 / 0.60 = 0.5

*P(X3|Yb)* = *P(X3,Yb) / P(Yb)* = 0.20 / 0.40 = 0.5

**Q1(d)**

*P(Ya|X1)* = *P(X1,Ya) / P(X1)* = 0.10 / 0.15 = 0.667

*P(Ya|X2)* = *P(X2,Ya) / P(X2)* = 0.20 / 0.35 = 0.571

*P(Ya|X3)* = *P(X3,Ya) / P(X3)* = 0.30 / 0.50 = 0.6

*P(Yb|X1)* = *P(X1,Yb) / P(X1)* = 0.05 / 0.15 = 0.333

*P(Yb|X2)* = *P(X2,Yb) / P(X2)* = 0.15 / 0.35 = 0.429

*P(Yb|X3)* = *P(X3,Yb) / P(X3)* = 0.20 / 0.50 = 0.4

**Q1(e)**

*X and Y are not independent.*

*Proof 1:*

*P(X2,Ya) = 0.20*

*P(X2)P(Ya) = 0.35\*0.60 =0.21*

*P(X2,Ya) P(X2)P(Ya)*

*Proof 2:*

*P(Ya|X2) = 0.571*

*P(Ya) = 0.60*

*P(Ya|X2) P(Ya)*

*Proof 3:*

*P(X2|Ya) = 0.333*

*P(X2) = 0.35*

*P(X2|Ya) P(X2)*

**Q1(f)**

*H(X)*  =

*=* - (0.15log(0.15) + 0.35log(0.35) + 0.5log(0.5))

*=* - (0.15(-2.737) + 0.35(-1.515) + 0.5(-1))

*=* 0.411 + 0.53 + 0.5

*=* 1.441 *bits*

**Q1(g)**

*H(Y)*  =

*=* - (0.6log(0.6) + 0.4log(0.4))

*=* - (0.6(-0.737) + 0.4(-1.322))

*=* 0.442 + 0.529

*=* 0.971 *bits*

**Q1(h)**

*H(X,Y)*  =

When X=1,

= 0.10log(0.10) + 0.05log(0.05) = -0.548

When X=2,

= 0.20log(0.20) + 0.15log(0.15) = -0.875

When X=3,

= 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) =*

When X=1,

= 0.10log(0.10) + 0.05log(0.01) = -0.664

When X=2,

= 0.20log(0.20) + 0.15log(0.09) = -0.985

When X=3,

= 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) =*

When X=1,

= 0.10log(0.10) + 0.01og(0.05) = -0.375

When X=2,

= 0.20log(0.20) + 0.09log(0.15) = -0.711

When X=3,

= 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) =*

When X=1,

= 0.10log(0.10) + 0.01og(0.01) = -0.399

When X=2,

= 0.20log(0.20) + 0.09log(0.09) = -0.777

When X=3,

= 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)* = = - (*plogp* + *(1-p)log(1-p)*)

**Q2(b)**

*H(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)* = - (*plogp* + *(1-p)log(1-p)*)

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

= -*logp + log(1-p) - 2*

The second derivative of H(X) = -*logp + 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 *w1* = *t1*, *w2* = *t2*, …, *wi* = *ti*

The number of documents that satisfy the condition stated in Q3(b)

= *N*! / (*t1! \* t2! \* t3!* …\* *ti*!*)*

**Q3(c)**

Probability of picking a word *wi = P(wi) =* *pi*

Number of ways to get exactly *ti* *= N*! / *ti!*

Probability of getting exactly *ti* *= (N*! / *ti!) / N! = 1/ ti!*

Probability where the occurrence of *wi* is exactly ti for each *wi*

=  *pi \* (1 / ti!)*

= *pi / ti!*

**Q4(a)**

Trigram = *P(wi|ti)P(ti|ti-2,ti-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

*aij* is the transition probability that corresponds to *P(ti|ti-2,ti-1)*

*bjk* is the emission probability that corresponds to *P(wi|ti)*

**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*