# Ling 572 HW1

Daniel Campos dacampos@uw.edu

01/16/2019

## 1 Q1: X and Y be two random variables.

**(a)** $P(X)$ Shown in Table 1

$$P(X = x) = P(X = 1) + P(X = 2) + P(X = 3) = 0.15 + 0.35 + 0.5 = 1 \tag{1}$$

**(b)** $P(Y)$ Shown in Table 1

$$P(Y = x) = P(Y = a) + P(Y = b) = 0.6 + 0.4 = 1 \tag{2}$$

**(c)** $P(X \mid Y)$ Shown in Table 2

$$P(X \mid Y) = P(X|Y = a) + P(X|Y = b) \tag{3}$$

**(d)** $P(Y \mid X)$ Shown in table 3

**(e)** Are X and Y independent? No since for two random variables to be independent P(X—Y) = P(X) and P(Y—X) = P(Y) and that is not the case.

**(f)** $H(X)$

$$H(X) = -\sum_x p(x) \log p(x). \tag{4}$$

$$H(X) = H(X = 1) + H(X = 2) + H(X = 3) = 1.44065 \tag{5}$$

**(g)** $H(Y)$

$$H(Y) = H(Y = a) + H(Y = b) = 0.97095 \tag{6}$$

**(h)** $H(X, Y)$

$$H(X, Y) = H(X = 1, Y = a) + H(X = 2, Y = a) + H(X = 3, Y = a) \tag{7}$$

$$+H(X = 1, Y = b) + H(X = 2, Y = b) + H(X = 3, Y = b) = 2.408694 \tag{8}$$

**(i)** $H(X \mid Y)$

$$H(X|Y) = H(X, Y) - H(Y) = 2.408694 - 0.97095 = 1.4377440000000001 \tag{9}$$

Table 1: P(X) and P(Y)

| | X=1 | X=2 | X=3 | P(Y) |
|---|---|---|---|---|
| Y=a | 0.10 | 0.20 | 0.30 | 0.60 |
| Y=b | 0.05 | 0.15 | 0.20 | 0.40 |
| P(X=x) | 0.15 | 0.35 | 0.50 | 1 |

Table 2: $P(X \mid Y)$

| | X=1 | X=2 | X=3 |
|---|---|---|---|
| Y=a | 1/6 | 1/3 | 1/2 |
| Y=b | 1/8 | 3/8 | 1/2 |

Table 3: $P(Y \mid X)$

| | X=1 | X=2 | X=3 |
|---|---|---|---|
| Y=a | 2/3 | 4/7 | 3/5 |
| Y=b | 1/3 | 3/7 | 2/5 |

**(j)** $H(Y \mid X)$

$$H(Y|X) = H(X,Y) - H(X) = 2.408694 - 1.44065 = 0.9680440000000001 \tag{10}$$

**(k)** $MI(X,Y)$

$$MI(Y|X) = \sum_x \sum_y p(x,y) log((p(x,y)/(p(x)p(y)) = 0.002901074507172899 \tag{11}$$

**(l1)** What is the value for $KL(P(X,Y) \mid\mid Q(X,Y))$?

$$KL(p,q) = \sum_x p(x) log2(p(x)/p(q)) = 0.10212999408564584 \tag{12}$$

**(l2)** What is the value for $KL(Q(X,Y) \mid\mid P(X,Y))$?

$$KL(p,q) = \sum_x p(x) log2(p(x)/p(q)) = 0.07646881528770542 \tag{13}$$

**(l3)** Are they the same? No

## 2 Q2: Random Variable from coin toss

**(a)** Formula for $H(X)$.

$$H(X) = -\sum_x p(x) \log p(x). \tag{14}$$

**(b)** What is $p^*$?
.368 with a entropy of 0.530737816926673

**(c)** Prove that the answer you give in (b) is correct.
To find the maximum value we can take the derivative of the equation and solve for y= 0. This gives us x = .368.

$$H'(X) = \log p(x) + 1)/log(2) = 0 \tag{15}$$

## 3 Q3: Permutations and combinations

**(a)** How many distinct ways are there to form the teams for the class(Including formula)?
For 2, ways = 1, for 4 = 3,

$$Teams(n) = ((n-1)!/(2! * ((n-1)-2)!) \tag{16}$$

**(b)** How many different color sequences are there?
Since each color sequence acts independently the amount of cololor sequences is n! the amount of balls in that color

$$Sequences() = 5! * 3! * 2! = 1440 \tag{17}$$

**(c1)** How many different word sequences are there which contain exactly $t_i$ $w_i$'s for each $w_i$ in $\Sigma$?
To figure this out I started out with a Vocabulary of 0 and 1s of various sizes and tried to understand the document size and the possibilies and how many documents match the desired conditions. Looking at a bunch of examples I was able to confirm that there are N! different word sequences that contain the correct frequency

**(c2)** What is the probability that you will end up with a document where the occurrence of the word $w_i$ (for each $w_i \in \Sigma$) in the document is exactly $t_i$?
Where V is the length of the word array

$$P = \prod_{i=1}^{n} N!/V! \tag{18}$$

## 4 Q4: POS Tagger

**(a)** Write down the formula for calculating $P(w_1, ..., w_n, t_1, ..., t_n)$,

$$P = \prod_{i=1}^{n} P(wi|ti)P(ti|ti-2, ti-1...ti-n) \tag{19}$$

**(b1)** What does each state in HMM correspond to?
In an HMM each state corresponds to a word in a sentence or chain.

**(b2)** How many states are there?
Depending on how large of a model we want to train whatever our look back window is. N states

**(b3)** What probabilities in the formula for (a) the transition probability is the $p_i$ given $t_(i-1)$ and transmison probability is $t_i$ given $w_i$

## 5 Q5: POS Classifier

**(a)** How many unique features are there?
4V + $3V^2$

**(b)** What is x? what is y? The X is a current word while the Y is the predicted POS tag for the current word.

**(c)** For the sentence **Mike/NN likes/VBP cats/NNS**,write down the feature vector for each word in the sentence.
Mike NN PreviousWord BOS CurrentWord Mike NextWord likes surroundingWords (BOS,likes) PreviousTag BOS Prev2tags (BOS,BOS)
likes VBP PreviousWord Mike CurrentWord likes NextWord cats surroundingwords (Mike,cats) PreviousTag NN Prev2tags(NN,BOS)
cats NNS PreviousWord likes CurrentWord cats surroundingwords (likes,EOS) PreviousTag VBP Prev2tags(NN, VBP)

# 6 Q6: Language Identifier

**(a)** How do you plan to build the LangID system?
I would treat this as a classification problem. My input would be a document and the output would be a ID which corresponds to a specific langauge. A good set of features would be characters(as in what unicode characters), word unigrams, word bigrams, word trigrams along with POS unigrams, POS bigrams and POS trigrams.

**(b)** What factors could affect the system performance?
The origin of the training data could really affect the training data because some documents in some langauges may come from new and others may come from literature. The system could also be affected by how similair much of the training data is, if its too similair or has too much overlap making a confident model will be trickly. Finally, since my model would use POS tagging to predict langauge the quality of our POS tagger would be hugely impactful. If the POS tagger is low quality tahn we would just learn noise.