LING 572 HW 1

Devin Johnson

January 24, 2020

Problem 1

a.)

$$P(X = 1) = 0.15$$
$$P(X = 2) = 0.35$$
$$P(X = 3) = 0.50$$

b.)

$$P(Y = a) = 0.60$$
$$P(Y = b) = 0.40$$

c.)

$$P(X = 1|Y = a) = \frac{P(X=1, Y=a)}{P(Y=a)} = 0.17$$
$$P(X = 1|Y = b) = 0.125$$
$$P(X = 2|Y = a) = 0.3$$
$$P(X = 2|Y = b) = 0.38$$
$$P(X = 3|Y = a) = 0.50$$
$$P(X = 3|Y = b) = 0.50$$

d.)

$$P(Y = a|X = 1) = 0.67$$
$$P(Y = a|X = 2) = 0.57$$
$$P(Y = a|X = 3) = 0.60$$
$$P(Y = b|X = 1) = 0.33$$
$$P(Y = b|X = 2) = 0.43$$
$$P(Y = b|X = 3) = 0.40$$

e.)

They are not independent because $p(X = 1|Y = a) \neq p(X = 1)$. This is the only example we need because it serves as a counter example.

f.)

$$
\begin{align}
H(X) &= -\sum_i p(x_i)log_2(p(x_i)) \tag{1}\\
&= (0.15)(-log_2(0.15)) + (0.35)(-log_2(0.35)) + (0.50)(-log_2(0.50)) \tag{2}\\
&= 1.44 \tag{3}
\end{align}
$$

g.)

$$
\begin{align}
H(Y) &= -\sum_i p(y_i)log_2(p(y_i)) \tag{4}\\
&= (0.60)(-log_2(0.60)) + (0.40)(-log_2(0.40)) \tag{5}\\
&= 0.97 \tag{6}
\end{align}
$$

h.)

Assuming this is joint entropy:

$$
\begin{align}
H(X,Y) &= -\sum_{x\in X}\sum_{y\in Y} p(x,y)log_2(p(x,y)) \tag{7}\\
&= P(x=1,y=a)log_2(P(x=1,y=a)+ \tag{8}\\
&\quad p(x=1,y=b)log_2(p(x=1,y=b)+ \tag{9}\\
&\quad p(x=2,y=a)log_2(p(x=2,y=a)+ \tag{10}\\
&\quad p(x=2,y=b)log_2(p(x=2,y=b)+ \tag{11}\\
&\quad p(x=3,y=a)log_2(p(x=3,y=a)+ \tag{12}\\
&\quad p(x=3,y=b)log_2(p(x=3,y=b) \tag{13}\\
&= 2.41 \tag{14}
\end{align}
$$

i.)

$$
\begin{aligned}
H(X|Y) &= -\sum_{y \in Y, x \in X} p(x,y) log_2 \frac{p(x,y)}{p(y)} && (15)\\
&= p(1,a) log_2 \frac{p(1,a)}{p(a)} + && (16)\\
&\quad p(2,a) log_2 \frac{p(2,a)}{p(a)} + && (17)\\
&\quad p(3,a) log_2 \frac{p(3,a)}{p(a)} + && (18)\\
&\quad p(1,b) log_2 \frac{p(1,b)}{p(b)} + && (19)\\
&\quad p(2,b) log_2 \frac{p(2,b)}{p(b)} + && (20)\\
&\quad p(3,b) log_2 \frac{p(3,b)}{p(b)} && (21)\\
&= 1.44 && (22)
\end{aligned}
$$

j.)

$$
\begin{aligned}
H(Y|X) &= -\sum_{x \in X, y \in Y} p(x,y) log_2 \frac{p(x,y)}{p(x)} && (23)\\
&= p(1,a) log_2 \frac{p(1,a)}{p(1)} + && (24)\\
&\quad p(1,b) log_2 \frac{p(1,b)}{p(1)} + && (25)\\
&\quad p(2,a) log_2 \frac{p(2,a)}{p(2)} + && (26)\\
&\quad p(2,b) log_2 \frac{p(2,b)}{p(2)} + && (27)\\
&\quad p(3,a) log_2 \frac{p(3,a)}{p(3)} + && (28)\\
&\quad p(3,b) log_2 \frac{p(3,b)}{p(3)} && (29)\\
&= 0.968 && (30)
\end{aligned}
$$

k.)

$$\begin{aligned}
MI(X,Y) &= H(X) + H(Y) - H(X,Y) && (31)\\
&= 1.4406 + .97095 - 2.40869 && (32)\\
&= 0.00286 && (33)
\end{aligned}$$

l.)

They are not the same.
- x represents an event
- p(x) = probability of x under P distribution
- q(x) = probability of x under Q distribution

$$\begin{aligned}
KL(P||Q) &= H_c(P,Q) - H(P) && (34)\\
&= \sum_{x \in X} p(x) log_2 q(x) + \sum_{x \in X} p(x) log_2 p(x) && (35)\\
&= 2.51 - 2.41 && (36)\\
&= 0.1 && (37)
\end{aligned}$$

$$\begin{aligned}
KL(Q||P) &= H_c(Q,P) - H(Q) && (38)\\
&= \sum_{x \in X} q(x) log_2 p(x) + \sum_{x \in X} q(x) log_2 q(x) && (39)\\
&= 2.25 - 2.17 && (40)\\
&= 0.08 && (41)
\end{aligned}$$

Problem 2

Not assuming fair coin:

a.)

$$\begin{aligned}
H(X) &= -\sum_i p(x_i) log_2(p(x_i)) && (42)\\
&= -(P log_2(P)) - ((1-P) log_2(1-P)) && (43)
\end{aligned}$$

b/c.)

We have a function of P that we would like to maximize. We know P takes values between 0 and 1 since it is a probability. To prove that whatever P* we find is a maximum, we will need to take the first and second derivative. So we want:

$$0 = \frac{d}{dP}(-Plog_2(P) - (1-P)log_2(1-P)) \tag{44}$$

$$0 = \frac{ln(1-P) - ln(P)}{ln(2)} \tag{45}$$

$$P = 0.5 \tag{46}$$

So: $P* = 0.5$ and we've found a single extreme point in our function. Now we need the second derivative to make sure it's a maximum and not a minimum.

$$\frac{d}{dP}(\frac{ln(1-P) - ln(P)}{ln(2)}) \tag{47}$$

$$= \frac{1}{(P)(1-P)(ln(2))} \tag{48}$$

$$\frac{1}{0.5 * 0.5 * ln(2)} = -0.17 \tag{49}$$

Since the second derivative is negative, we know we have a maximum. Therefore we have one and only one extreme and it is a maximum so we know P* is correct. QED

Problem 3

a.)

We know $n \in \{2, 4, 6, 8, 10, ...\}$. For a given n, there are $\frac{n}{2}$ teams of 2. We want to determine how many ways we can pick those teams of 2. We can use the following formula:

$$\frac{n!}{(\frac{n}{2}!) * (2!)^{\frac{n}{2}}}$$

For example: If $n = 4$ then we get:

$$\frac{4!}{2!*2!*2!} = \frac{4*3}{4} = \frac{12}{4} = 3$$

This is exactly what we expect. We are dividing on the bottom to account for overcounting of the positioning of groups and inner position of students. We don't care about those so we divide them out.

b.)

- Numerator = total ways to arrange all balls
- Denominator = divide out balls with same colors in row since we want sequence and don't care about order of duplicate balls next to each other

$$\frac{10!}{5!*3!*2!} = 2520$$

c1.)

- Numerator = total ways to arrange N words
- Denominator = divide out words with repeats since we want sequence and don't care about order of duplicate words next to each other

$$\frac{N!}{\Pi_i(t_i!)}$$

c2.)

- First term = The probability of such a document happening is the product of certain sequences of w_i happening exactly t_i times.
- Second term = Number of such documents.

$$(\Pi_i P(X = w_i)^{t_i})(\frac{N!}{\Pi_i(t_i!)})$$

Problem 4

a.)

- $t_{n+1} = STOP$ (stop symbol)
- $\Pi_{i=1}^{n+1} q(t_i|t_{i-2}, t_{i-1})$ calculates the probability of a given tag sequence $t_1, ..., t_n$.
- $\Pi_{i=1}^{n} e(w_i|t_i)$ calculates the probability of the word sequence $w_1, ..., w_n$ being emitted from the tag sequence $t_1, ..., t_n$.

$$P(w_1...w_n, t_1...t_n) = \Pi_{i=1}^{n+1} q(t_i|t_{i-2}, t_{i-1}) \Pi_{i=1}^{n} e(w_i|t_i)$$

In a trigram HMM model, for a given sequence of words, we want to find the tag sequence that maximizes this probability. We would first estimate the parameters listed above through training data.

b.)

- Each state corresponds to a unique POS tag so the amount of states we have is the number of unique POS tags. If you count emission states as states, then you can add those to the number as well (but I usually don't see such an assumption).
- $\Pi_{i=1}^{n+1} q(t_i|t_{i-2}, t_{i-1})$ calculates the probability of a given tag sequence $t_1, ..., t_n$.
- $\Pi_{i=1}^{n} e(w_i|t_i)$ calculates the probability of the word sequence $w_1, ..., w_n$ being emitted from the tag sequence $t_1, ..., t_n$.

Problem 5

a.)

$$O(V^2 + T^2)$$

b.)

The input x is a word and y is the corresponding tag.

c.)

- Mike/NN: (Mike NN, prevWord None, currWord Mike, nextWord likes, surroundingWords likes, prevTag None, prev2Tags None)
- likes/VBP: (likes VBP, prevWord Mike, currWord likes, nextWord cats, surroundingWords Mike cats, prevTag NN, prev2Tags NN)
- cats/NNS: (cats VBP, prevWord likes, currWord cats, nextWord None, surroundingWords likes, prevTag VBP, prev2Tags NN VBP)

Problem 6

a.)

- x (input): Document
- y (output): Lang ID for document
- Possible features: Features for probabilities of different alphabets (could be hundreds even), unigrams, bigrams, word length, number of spaces, counts of punctuation.

b.)

- Domain of training data (perhaps domains don't match and some language is abbreviated such as in texting)
- Format of training/testing data (should be same format)
- Training data skewed toward one set of languages
- Messy data with irrelevant text