

Natural Language Processing

Assignment 2: Probability and Vector Exercises

Li-Yi Lin, JHED ID: 5F0CE8

Question 1.

(a) Answer:

Since $Y \subseteq Z$, we have $Z = Y \cup (Z \cap \neg Y)$. So the probability of Z is

$$\begin{aligned} p(Z) &= p(Y \cup (Z \cap \neg Y)) \\ &= p(Y) + p(\emptyset) \\ &\geq p(Y) + 0 \\ &= p(Y) \end{aligned} \quad (p(\emptyset) \geq 0)$$

Thus, we proved that if $Y \subseteq Z$, then $p(Y) \leq p(Z)$.

(b) Answer:

We know that $p(X | Z) = \frac{p(X \cap Z)}{p(Z)}$. Since $(X \cap Z) \subseteq Z$, we have the fact that $p(X \cap Z) \leq p(Z)$. Therefore, $p(X | Z) = \frac{p(X \cap Z)}{p(Z)} \leq 1$. Moreover, since probability is larger than or equal to 0, we have $p(X \cap Z) \geq 0$. Because $p(X | Z)$ is conditioned on Z , we have $p(Z) > 0$. Therefore, by the property that $0 \leq p(X \cap Z) \leq p(Z)$, we can prove that $p(X | Z)$ always fall in the range $[0,1]$.

(c) Answer:

Since $E \cap \emptyset = \emptyset$, we have $p(E \cup \emptyset) = p(E) + p(\emptyset)$ by the given axioms. Because we know that $p(E) = 1$ and probability cannot be larger than 1, $p(E \cup \emptyset) = p(E) + p(\emptyset) = 1 + p(\emptyset) \leq 1$. Therefore $p(\emptyset)$ must be 0.

(d) Answer:

Let \bar{X} denote $E - X$. We have $\bar{X} \cup X = E$. Now we are going to prove that $p(X) = 1 - p(\bar{X})$. Since $\bar{X} \cap X = \emptyset$ and $\bar{X} \cup X = E$, we have

$$\begin{aligned} p(E) &= p(\bar{X} \cup X) \\ &= p(\bar{X}) + p(X) - p(\bar{X} \cap X) \\ &= p(\bar{X}) + p(X) \end{aligned} \quad (p(\bar{X} \cap X) = 0 \text{ by (c)})$$

By the equation above, we have $p(E) = p(\bar{X}) + p(X)$. Since $p(E) = 1$, we have $1 = p(\bar{X}) + p(X)$. Therefore we have proven that $p(X) = 1 - p(\bar{X})$.

(e) Answer:

$$\begin{aligned} p(\text{singing} \cap \text{rainy} \mid \text{rainy}) &= \frac{p((\text{singing} \cap \text{rainy}) \cap \text{rainy})}{p(\text{rainy})} \\ &= \frac{p(\text{singing} \cap (\text{rainy} \cap \text{rainy}))}{p(\text{rainy})} && \text{(Intersection is an associative operation)} \\ &= \frac{p(\text{singing} \cap \text{rainy})}{p(\text{rainy})} && (\text{rainy} \cap \text{rainy} = \text{rainy}) \\ &= p(\text{singing} \mid \text{rainy}) \end{aligned}$$

By the equation above, we proved that $p(\text{singling AND rainy} \mid \text{rainy}) = p(\text{singling} \mid \text{rainy})$.

(f) Answer:

$$\begin{aligned} p(Y) &= p((X \cap Y) \cup (\bar{X} \cap Y)) \\ &= p(X \cap Y) + p(\bar{X} \cap Y) - p((X \cap Y) \cap (\bar{X} \cap Y)) \\ &= p(X \cap Y) + p(\bar{X} \cap Y) \quad ((X \cap Y) \cap (\bar{X} \cap Y) = \emptyset) \end{aligned}$$

By dividing both sides by $p(Y)$, we proved that

$$\begin{aligned} 1 &= \frac{p(X \cap Y)}{p(Y)} + \frac{p(\bar{X} \cap Y)}{p(Y)} \\ p(X \mid Y) &= 1 - p(\bar{X} \mid Y) \end{aligned}$$

(g) Answer:

$$\begin{aligned} &(p(X \mid Y) \cdot p(Y) + p(X \mid \bar{Y}) \cdot p(\bar{Y})) \cdot p(\bar{Z} \mid X) / p(\bar{Z}) \\ &= \left(\frac{p(X \cap Y)}{p(Y)} \cdot p(Y) + \frac{p(X \cap \bar{Y})}{p(\bar{Y})} \cdot p(\bar{Y}) \right) \cdot \frac{p(\bar{Z} \cap X)}{p(X)} / p(\bar{Z}) \\ &= (p(X \cap Y) + p(X \cap \bar{Y})) \cdot \frac{p(\bar{Z} \cap X)}{p(X)} / p(\bar{Z}) \\ &= p(X) \cdot \frac{p(\bar{Z} \cap X)}{p(X)} / p(\bar{Z}) \\ &= \frac{p(\bar{Z} \cap X)}{p(\bar{Z})} \\ &= p(X \mid \bar{Z}) \end{aligned}$$

(h) Answer:

We know that

$$p(\text{singling AND rainy}) = p(\text{singling}) + p(\text{rainy}) - p(\text{singling} \cap \text{rainy})$$

So, when $(\text{singling} \cap \text{rainy}) = \emptyset$, then

$$\begin{aligned} p(\text{singling AND rainy}) &= p(\text{singling}) + p(\text{rainy}) - p(\text{singling} \cap \text{rainy}) \\ &= p(\text{singling}) + p(\text{rainy}) - p(\emptyset) \\ &= p(\text{singling}) + p(\text{rainy}) - 0 \\ &= p(\text{singling}) + p(\text{rainy}) \end{aligned} \quad (\text{by (c)})$$

Therefore, when set *singling* and *rainy* are disjoint, the given equation is true.

(i) Answer:

Assume the given equation is true, then

$$p(\text{singling AND rainy}) = p(\text{singling}) \cdot p(\text{rainy})$$

We further assume that $p(\text{rainy}) > 0$ and divide the above equation by $p(\text{rainy})$

$$\frac{p(\text{singing AND rainy})}{p(\text{rainy})} = p(\text{singing})$$

$$p(\text{singing} \mid \text{rainy}) = p(\text{singing})$$

Therefore, for the $p(\text{singing AND rainy}) = p(\text{singing}) \cdot p(\text{rainy})$ to be true, $p(\text{singing} \mid \text{rainy}) = p(\text{singing})$ must also be true. It means that the probability event *rainy* will not affect the probability of event *singing*.

(j) Answer:

We are given that

$$\begin{aligned} p(X \mid Y) &= 0 \\ \frac{p(X \cap Y)}{p(Y)} &= 0 \end{aligned}$$

For the above equation to be 0, $p(X \cap Y)$ must be 0.

Then, we are going to prove that $p(X \mid Y, Z) = 0$

$$\begin{aligned} p(X \mid Y, Z) &= \frac{p(X \cap Y \cap Z)}{p(Y \cap Z)} \\ &\leq \frac{p(X \cap Y)}{p(Y \cap Z)} && (\text{by (a), } (X \cap Y \cap Z) \subseteq (X \cap Y), \text{ so } p(X \cap Y \cap Z) \leq p(X \cap Y)) \\ &= \frac{0}{p(Y \cap Z)} \\ &= 0 \end{aligned}$$

Thus, we have proved that if $p(X \mid Y) = 0$, then $p(X \mid Y, Z) = 0$.

(k) Answer:

By (f), we know

$$\begin{aligned} p(\bar{W} \mid Y) &= 1 - p(W \mid Y) \\ &= 1 - 1 \\ &= 0 \end{aligned}$$

Since $p(\bar{W} \mid Y) = \frac{p(\bar{W} \cap Y)}{p(Y)} = 0$, then $p(\bar{W} \cap Y)$ must be 0.

Then, we are going to prove that $p(W \mid Y, Z) = 1$

$$\begin{aligned} p(W \mid Y, Z) &= 1 - p(\bar{W} \mid Y, Z) && (\text{by (f)}) \\ &= 1 - \frac{p(\bar{W} \cap Y \cap Z)}{p(Y \cap Z)} \\ &= 1 - 0 && (\text{since } (\bar{W} \cap Y \cap Z) \subseteq (\bar{W} \cap Y), p(\bar{W} \cap Y \cap Z) \leq p(\bar{W} \cap Y) \text{ by (a)}) \\ &= 1 \end{aligned}$$

Thus, we have proven that if $p(W \mid Y) = 1$, then $p(W \mid Y, Z) = 1$.

Question 2. (a) Answer:

Denote event "Actual = Blue" as A , "Claimed = blue" as C .

By Bayes' theorem, we have

$$p(A | C)p(C) = p(C | A)p(A)$$

(b) Answer:

Since $p(\text{Actual} = \text{blue}) = 0.1$ is fixed. We further rewrite the equation in (a) as shown below:

$$p(\text{Actual} = \text{blue} | \text{Claimed} = \text{blue}) \propto p(\text{Claimed} = \text{blue} | \text{Actual} = \text{blue})p(\text{Actual} = \text{blue})$$

Then, the prior probability is $p(\text{Actual} = \text{blue})$,

the likelihood of the evidence is $p(\text{Claimed} = \text{blue} | \text{Actual} = \text{blue})$,

and the posterior probability is $p(\text{Actual} = \text{blue} | \text{Claimed} = \text{blue})$.

(c) Answer:

prior probability = $p(\text{Actual} = \text{blue}) = 0.1$

likelihood of the evidence = $p(\text{Claimed} = \text{blue} | \text{Actual} = \text{blue}) = 0.8$

posterior probability = $\frac{p(\text{Claimed}=\text{blue}|\text{Actual}=\text{blue}) \cdot p(\text{Actual}=\text{blue})}{p(\text{Claimed}=\text{blue})} = \frac{0.8 \times 0.1}{p(\text{Claimed}=\text{blue})}$

Now we need to calculate $p(\text{Claimed} = \text{blue})$.

$p(\text{Claimed} = \text{blue})$

$= p(\text{Claimed} = \text{blue} | \text{Actual} = \text{blue}) \cdot p(\text{Actual} = \text{blue}) + p(\text{Claimed} = \text{blue} | \text{Actual} = \text{red}) \cdot p(\text{Actual} = \text{red})$

$= 0.8 \times 0.1 + 0.2 \times 0.9$

$= 0.26$

So, posterior probability = $\frac{0.8 \times 0.1}{p(\text{Claimed}=\text{blue})} = \frac{0.08}{0.26} = 0.3077$

The judge should care about the posterior because it takes not only about the observed events, likelihood, but also the belief about the probability of the event we want to guess. By considering the posterior probability, the judge can avoid the situation that our belief is wrong or we have not enough observed events so that the judge can have more accurate information for judgement.

(d) Answer:

To prove the given equation, we start with the right hand side:

$$\begin{aligned} \frac{p(B | A, Y) \cdot p(A | Y)}{p(B | Y)} &= \frac{\frac{p(B \cap A \cap Y)}{p(A \cap Y)} \cdot \frac{p(A \cap Y)}{p(Y)}}{\frac{p(B \cap Y)}{p(Y)}} \\ &= \frac{p(B \cap A \cap Y)}{p(B \cap Y)} \\ &= p(A | B, Y) \end{aligned}$$

Thus, by we have proven that

$$p(A | B, Y) = \frac{p(B | A, Y) \cdot p(A | Y)}{p(B | Y)}$$

(e) Answer:

By (d), we have

$$p(A | B, Y) = \frac{p(B | A, Y) \cdot p(A | Y)}{p(B | Y)}$$

Now we want to prove that

$$p(A | B, Y) = \frac{p(B | A, Y) \cdot p(A | Y)}{p(B | A, Y) \cdot p(A | Y) + p(B | \bar{A}, Y) \cdot p(\bar{A} | Y)}$$

So we only have to prove that

$$p(B | Y) = p(B | A, Y) \cdot p(A | Y) + p(B | \bar{A}, Y) \cdot p(\bar{A} | Y)$$

We start from the right hand side.

$$\begin{aligned} p(B | A, Y) \cdot p(A | Y) + p(B | \bar{A}, Y) \cdot p(\bar{A} | Y) &= \frac{p(B \cap A \cap Y)}{p(A \cap Y)} \cdot \frac{p(A \cap Y)}{p(Y)} + \frac{p(B \cap \bar{A} \cap Y)}{p(\bar{A} \cap Y)} \cdot \frac{p(\bar{A} \cap Y)}{p(Y)} \\ &= \frac{p(B \cap A \cap Y)}{p(Y)} + \frac{p(B \cap \bar{A} \cap Y)}{p(Y)} \\ &= \frac{p(B \cap Y)}{p(Y)} \\ &= p(B | Y) \end{aligned}$$

By the derivation above, we proved the given equation.

(f) Answer:

Let Y be "Baltimore," A be "car is actually blue," and B be "Claimed the car is blue."

Then the original equation becomes:

$$\begin{aligned} p(\text{car is actually blue} | \text{Claimed the car is blue, Baltimore}) &= \frac{p(B | A, Y) \cdot p(A | Y)}{p(B | A, Y) \cdot p(A | Y) + p(B | \bar{A}, Y) \cdot p(\bar{A} | Y)} \\ &= \frac{0.8 \times 0.1}{0.8 \times 0.1 + 0.2 \times 0.9} \\ &= 0.3077 \end{aligned}$$

Question 3.

(a) Answer:

$$\sum_{c \in \text{cry}} p(c | s) = 1, \text{ where } s \in \text{situation}$$

(b) Answer:

p(cry, situation)	Predator!	Timber!	I need help!	TOTAL
bwa	0	0	0.64	0.64
bwee	0	0	0.08	0.08
kiki	0.2	0	0.08	0.28
TOTAL	0.2	0	0.8	1

(c) Answer:

i. This probability is written as: $p(\text{Predator} | \text{kiki})$

ii. It can be rewritten without the | symbol as: $\frac{p(\text{Predator, kiki})}{p(\text{kiki})}$

iii. Using the above tables, its value is: $\frac{0.2}{0.28} = 0.7143$

iv. Alternatively, Bayes's Theorem allows you to express this probability as:

$$\frac{p(\text{kiki} | \text{Predator}) \cdot p(\text{Predator})}{p(\text{kiki} | \text{Predator}) \cdot p(\text{Predator}) + p(\text{kiki} | \text{Timber}) \cdot p(\text{Timber}) + p(\text{kiki} | \text{I need help}) \cdot p(\text{I need help})}$$

v. Using the above tables, the value of this is:

$$\frac{1 \times 0.2}{1 \times 0.2 + 0 \times 0 + 0.1 \times 0.8} = 0.7143$$

Question 4.

(a) Answer:

$$p(\vec{w}) = \frac{c(BOS\ BOS\ w_1)}{c(BOS\ BOS)} \frac{c(BOS\ w_1\ w_2)}{c(BOS\ w_1)} \frac{c(w_1\ w_2\ w_3)}{c(w_1\ w_2)} \dots \frac{c(w_{n-2}\ w_{n-1}\ w_n)}{c(w_{n-2}\ w_{n-1})} \frac{c(w_{n-1}\ w_n\ EOS)}{c(w_{n-1}\ w_n)}$$

$c(BOS\ BOS)$ counts the total number of sentences in the corpus.

$c(BOS\ BOS\ i)$ counts the total number of sentences that start with a word i in the corpus.

$c(new\ york\ EOS)$ counts the total number of sentences that end with *new york* in the corpus.

(b) Answer:

The probability of $< s >$ do you think the $< /s >$ should be extremely low because "the" has very low probability, or even close to zero, to be the last word in a sentence.

In the trigram model, the parameter that makes this probability low is $c(think\ the\ < /s >)$ because the "think the $< /s >$ " will not happen many times in the corpus. *the*, a determiner, should be followed by a noun or noun phrase. But in the example, *the* is followed by $< /s >$.

(c) Answer:

Expression (A) is matched with description (2) because (A) doesn't have $< s >$ or $< /s >$, we don't know whether it is a sentence. We only know that we hear the three words, "Do you think".

Expression (B) is matched with description (1) since (B) has both the $< s >$ and $< /s >$ to indicate the start and end mark of the sentence "Do you think".

Expression (C) is matched with description (3) because (C) only has $< s >$ indicating the start of a sentence but no $< /s >$. So we only can be sure that "Do you think" is the starting words of a sentence.

$p(\vec{w})$ is to calculate the probability of the sentence using trigram model. Since a sentence needs a $< s >$ and a $< /s >$, quantity (B) is suitable for describing the trigram model probability of a sentence.

(d) Answer:

We first expand the equation for both $p(w)$ and $p_{reversed}(\vec{w})$ and show they are the same.

For $p(\vec{w})$

$$\begin{aligned} p(\vec{w}) &= \frac{p(w_1, w_0, w_{-1})}{p(w_0, w_{-1})} \frac{p(w_2, w_3, w_4)}{p(w_3, w_4)} \dots \frac{p(w_n, w_{n-1}, w_{n-2})}{p(w_{n-1}, w_{n-2})} \frac{p(w_{n+1}, w_n, w_{n-1})}{p(w_n, w_{n-1})} \\ &= \frac{p(w_1, BOS, BOS)}{p(BOS, BOS)} \frac{p(w_2, w_3, w_4)}{p(w_3, w_4)} \dots \frac{p(w_n, w_{n-1}, w_{n-2})}{p(w_{n-1}, w_{n-2})} \frac{p(EOS, w_n, w_{n-1})}{p(w_n, w_{n-1})} \end{aligned}$$

If we use "i love new york" as an example for this model, then it becomes

$$p(i\ love\ new\ york) = \frac{p(i, BOS, BOS)}{p(BOS, BOS)} \frac{p(love, i, BOS)}{p(i, BOS)} \frac{p(new, love, i)}{p(love, i)} \frac{p(york, new, love)}{p(new, love)} \frac{p(EOS, york, new)}{p(york, new)} \quad (4.d.1)$$

For $p_{reversed}(\vec{w})$ (notice that w_1 is the right most word in the origin sentence)

$$\begin{aligned} p_{reversed}(\vec{w}) &= \frac{p(w_0, w_1, w_2)}{p(w_1, w_2)} \frac{p(w_1, w_2, w_3)}{p(w_2, w_3)} \dots \frac{p(w_{n-1}, w_n, w_{n+1})}{P(w_n, w_{n+1})} \frac{p(w_n, w_{n+1}, w_{w+2})}{p(w_{n+1}, w_{n+2})} \\ &= \frac{p(BOS, w_1, w_2)}{p(w_1, w_2)} \frac{p(w_1, w_2, w_3)}{p(w_2, w_3)} \dots \frac{p(w_{n-1}, w_n, w_{n+1})}{P(w_n, w_{n+1})} \frac{p(w_n, EOS, EOS)}{p(EOS, EOS)} \end{aligned}$$

If we use "i love new york" as an example for this model, then it becomes

$$p_{reversed}(\text{i love new york}) = \frac{p(BOS, york, new)}{P(york, new)} \frac{p(york, new, love)}{p(new, love)} \frac{p(new, love, i)}{p(love, i)} \frac{p(love, i, EOS)}{p(i, EOS)} \frac{p(i, EOS, EOS)}{p(EOS, EOS)} \quad (4.d.2)$$

Since the "EOS" in $p(\vec{w})$ equals to the "BOS" in $p_{reversed}(\vec{w})$ and the "BOS" in $p(\vec{w})$ equals to the "EOS" in $p_{reversed}(\vec{w})$, the equation (4.d.1) equals to (4.d.2). If we observe the two equations, they have the same structure. Thus we have proved that $p(\vec{w}) = p_{reversed}(\vec{w})$.

Question 5.

Answer:

We added "a" denoting the topic variable and "h" denoting word history. Then make the language model condition on topic and word history.

$$p(w | h) = \sum_a p(w | a) p(a | h)$$

Next, we rewrite it as a bigram language model.

$$p(w | h) = \sum_a \prod_{i=1}^n p(w_i | w_{i-1}, a) p(a | \text{words before } w_i)$$

Based on the modified model, the formula for $p(w_1 w_2 w_3 w_4)$ can be written as

$$\begin{aligned} p(w | h) &= p(w_1 w_2 w_3 w_4 | h) \\ &= \sum_a p(w_1 | < s >, a) p(a | < s >) p(w_2 | w_1, a) p(a | < s >, w_1) \dots p(< /s > | w_n, a) p(a | < s >, w_1, \dots, w_n) \end{aligned}$$