CSE 250A. Assignment 1

Name: Guanghao Chen

PID: A53276390

Email: guc001@eng.ucsd.edu

1.1 Conditioning on background evidence

(a) $P(X,Y|E) = P(X|Y,E) \cdot P(Y|E)$

Proof:

Since we know that,

$P(X,Y,E) = P(X|Y,E) \cdot P(Y,E) = P(X|Y,E) \cdot P(Y|E) \cdot P(E)$

We also know that,

$P(X,Y,E) = P(X,Y|E) \cdot P(E)$

Therefore,

$P(X|Y,E) \cdot P(Y|E) \cdot P(E) = P(X,Y|E) \cdot P(E)$

Then,

$P(X|Y,E) \cdot P(Y|E) = P(X,Y|E)$

(b)$P(X|Y,E) = \frac{P(Y|X,E)P(X|E)}{P(Y|E)}$

Proof:

Since we know that,

$P(X,Y,E) = P(X|Y,E)P(Y,E) = P(X|Y,E)P(Y|E)P(E)$ (1)

We also have,

$P(X,Y,E) = P(Y|X,E)P(X,E) = P(Y|X,E)P(X|E)P(E)$ (2)

According to formula (1) and (2), it has,

$P(X|Y,E)P(Y|E)P(E) = P(Y|X,E)P(X|E)P(E)$

Then,

$P(X|Y,E)P(Y|E) = P(Y|X,E)P(X|E)$

$P(X|Y,E) = \frac{P(Y|X,E)P(X|E)}{P(Y|E)}$

(c) $P(X|E) = \sum_y P(X, Y=y|E)$

Proof:

$\sum_y P(X, Y = y|E) = \sum_y P(X|Y = y, E)P(Y = y|E)$

$= \sum_y \frac{P(Y=y|X,E)P(X|E)}{P(Y=y|E)} \cdot P(Y = y|E)$

$= \sum_y P(Y = y|X, E)P(X|E)$

$= P(Y = 0|X, E)P(X|E) + P(Y = 1|X, E)P(X|E) + \ldots + P(Y = n|X, E)P(X|E)$

$= P(X|E)$

1.2 Conditional independence

Proof:

Since we have:

$P(X, Y, E) = P(X, Y|E)P(E)$

$= P(X|Y, E)P(Y|E)P(E)$

$= P(Y|X, E)P(X|E)P(E)$

Hence,

$P(X, Y|E)P(E) = P(X|Y, E)P(Y|E)P(E) = P(Y|X, E)P(X|E)P(E)$

$\therefore P(X, Y|E) = P(X|Y, E)P(Y|E) = P(Y|X, E)P(X|E)$

For formula (1), (2) and (3), we can use the above equation to denote the expression in the left hand.

For example, for formula (3). If $P(X, Y|E) = P(X|E)P(Y|E)$

at the same time , $P(X, Y|E) = P(X|Y, E)P(Y|E)$,

then $P(X|E)P(Y|E) = P(X|Y, E)P(Y|E)$

$\therefore P(X|E) = P(X|Y, E)$

Therefore, we can say that under condition E, X and Y are conditionally independent.

1.3 Creative Writing

(a) Cumulative evidence

Z: One person healthy or not

X: One person likes to stay up late or not

Y: One person likes to drink alcohol or not

(b)Explaining away

X: One student is the best student among his classmates;

Z: One student learns diligently;

Y: One student is addict to playing games;

(C) Conditional independence

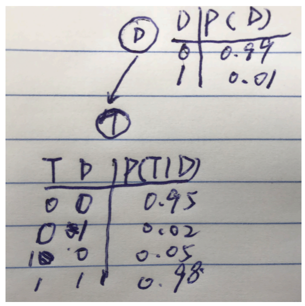
Y: A man has a child;

Z: A man likes to buy diapers;

X: A man has married.

1.4 Bayes Rule

(a)



According to the background,

$$\begin{aligned} P(D=1) &= 0.01 \\ P(T=1|D=0) &= 0.05 \\ P(T=0|D=1) &= 0.02 \end{aligned}$$

Then we have,

$$P(D=0) = 1 - P(D=1) = 0.99$$

$$\begin{aligned} P(T=0|D=0) &= 1 - P(T=1|D=0) \\ &= 1 - 0.05 \\ &= 0.95 \end{aligned}$$

$$\begin{aligned} P(T=1|D=1) &= 1 - P(T=0|D=1) \\ &= 1 - 0.02 \\ &= 0.98 \end{aligned}$$

In conclusion, we can deduce the table $P(D)$ and $P(T|D)$

D	P(D)
0	0.99
1	0.01

T	D	P(T|D)
0	0	0.95
0	1	0.02
1	0	0.05
1	1	0.98

(b)

The question is to calculate $P(D=0|T=1)$

Based on Bayes Rule, it has

$$\begin{aligned} P(T=1|D=0) &= \frac{P(D=0|T=1)P(T=1)}{P(D=0)} \\ => P(D=0|T=1) &= \frac{P(T=1|D=0)P(D=0)}{P(T=1)} \end{aligned}$$

also

$$\begin{aligned} P(T=1|D=1) &= \frac{P(D=1|T=1)P(T=1)}{P(D=1)} \\ => P(D=1|T=1) &= \frac{P(T=1|D=1)P(D=1)}{P(T=1)} \end{aligned}$$

Based on the axiom, it has

$$\begin{aligned} P(D=0|T=1)+P(D=1|T=1) &= 1 \\ \frac{P(T=1|D=0)P(D=0)}{P(T=1)}+\frac{P(T=1|D=1)P(D=1)}{P(T=1)} &= 1 \\ \frac{0.05\times 0.99}{P(T=1)}+\frac{0.98\times 0.01}{P(T=1)} &= 1 \\ \frac{0.0593}{P(T=1)} &= 1 \end{aligned}$$

Therefore,

$P(T=1)=0.0593$

$P(T=0)=1-P(T=1)=1-0.0593=0.9407$

Next we can calculate the $P(D=0|T=1)$

$$\begin{aligned} P(D=0|T=1) &= \frac{P(T=1|D=0)P(D=0)}{P(T=1)} \\ &= \frac{0.05\times 0.99}{0.0593} \\ &= 0.8347 \end{aligned}$$

(c)

$$\begin{aligned}
P(D=1|T=0) &= \frac{P(T=0|D=1)P(D=1)}{P(T=0)} \\
&= \frac{0.02 \times 0.01}{0.9407} \\
&= \frac{2}{9407}
\end{aligned}$$

1.5 Entropy

(a)

In order to show that $H[X] = -\sum p_i log p_i$ can be maximized when $p_i = \frac{1}{n}$

constraint:$\sum p_i = 1$

According to Lagrange Multiplier, we can utilize function F

$max(H[X])$

$= -\sum p_i log p_i + \lambda(\sum p_i - 1)$

Derivative

Derive F with respect to p_i,

Then,

$-(\log p_1 + 1) + \lambda = 0 => \lambda - 1 = \log p_1$

$-(\log p_2 + 1) + \lambda = 0 => \lambda - 1 = \log p_2$

...

$-(\log p_n + 1) + \lambda = 0 => \lambda - 1 = \log p_n$

Derive F with repect to λ,

have $\sum_{i=1}^{n} p_i = 1$

Therefore,

$P_1 = P_2 = \ldots = P_n = \frac{1}{n}$

All in all, $H[X]$will be maximized when $P_1 = P_2 = \ldots = P_n = \frac{1}{n}$.

(b)

$$\begin{aligned}
H(X_1, X_2, \ldots, X_{n-1}) &= -\sum_{x_1}\sum_{x_2}\cdots\sum_{x_{n-1}} P(x_1, x_2, \ldots, x_{n-1}) log P(x_1, x_2, \ldots, x_{n-1}) \\
&= -\sum_{x_1}\sum_{x_2}\cdots\sum_{x_{n-1}} P(x_1, x_2, \ldots, x_{n-1}) log P(x_1, x_2, \ldots, x_{n-1}) \times \sum P(x_n) \\
&= -\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} P(x_1, x_2, \ldots, x_{n-1}, x_n) log P(x_1, x_2, \ldots, x_{n-1})
\end{aligned}$$

$$
\begin{aligned}
H(X_n) &= -\sum x_n logP(x_n) \\
&= -\sum_{x_1}\sum_{x_2}\dots\sum_{x_{n-1}} P(x_1, x_2, \dots, x_{n-1}) \times \sum P(x_n)logP(x_n) \\
&= -\sum_{x_1}\sum_{x_2}\dots\sum_{x_n} P(x_1, x_2, \dots, x_n)logP(x_n)
\end{aligned}
$$

Therefore,

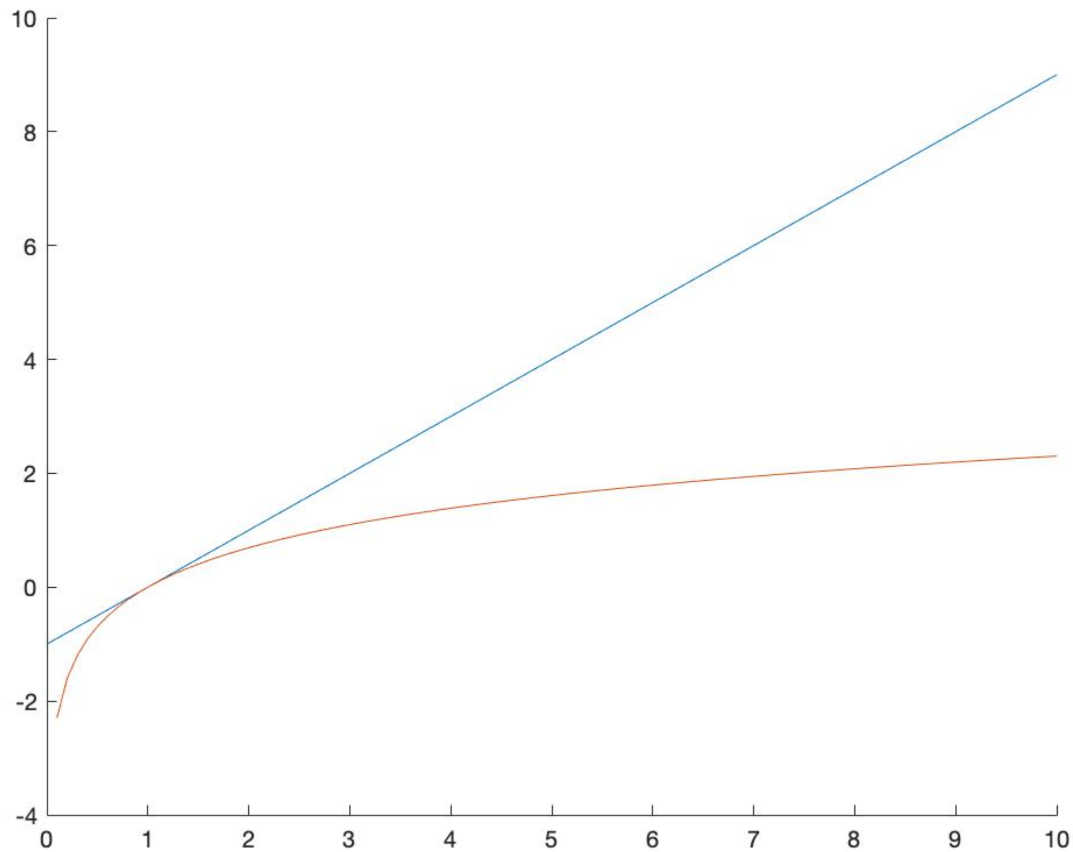
$$
\begin{aligned}
&H(X_1, X_2, \dots, X_{n-1}) + H(X_n) \\
&= (-\sum_{x_1}\sum_{x_2}\dots\sum_{x_n} P(x_1, x_2, \dots, x_{n-1}, x_n)logP(x_1, x_2, \dots, x_{n-1})) + (-\sum_{x_1}\sum_{x_2}\dots\sum_{x_n} P(x_1, x_2, \dots, x_n)logP(x_n)) \\
&= \sum_{x_1}\sum_{x_2}\dots\sum_{x_n} P(x_1, x_2, \dots, x_{n-1}, x_n)logP(x_1, x_2, \dots, x_n) \\
&= H(X_1, X_2, \dots, X_n)
\end{aligned}
$$

Then we can seperate $H(X_1, X_2, \dots, X_n)$ recursively.

$$
\begin{aligned}
&H(X_1, X_2, \dots, X_n) \\
&= H(X_n) + H(X_1, X_2, \dots, X_{n-1}) \\
&= H(X_n) + H(X_{n-1}) + \dots + H(X_{n-2}, X_{n-3}, \dots, H(X_1)) \\
&= H(X_n) + H(X_{n-1}) + \dots + H(X_1) \\
&= \sum H(X_i)
\end{aligned}
$$

1.6 Kullback-Leibler Distance

(a)



Transform the $log(x) \leq (x-1)$ into the following inequality,

have $log(x) - x + 1 \leq 0$

Derive $F(x) = log(x) - x + 1$with repect to x ,

$$\begin{aligned} \frac{dF(x)}{dx} &= \frac{1}{x} - 1 = 0 \\ &=> x = 1 \end{aligned}$$

Then it means there is an extreme point in $x = 1$.

(b)

Let $x = \frac{q_i}{p_i}$

According to inequality in (a), we can input x into it.

And we have,

$$\begin{aligned} log(\frac{q_i}{p_i}) &\leq (\frac{q_i}{p_i} - 1) \\ log(\frac{p_i}{q_i}) &\geq (1 - \frac{q_i}{p_i}) \\ p_i log(\frac{p_i}{q_i}) &\geq p_i(1 - \frac{q_i}{p_i}) \\ &= p_i - q_i \end{aligned}$$

$\therefore$

$$
\begin{aligned}
KL(p,q) &= \sum_i p_i log(\frac{p_i}{q_i}) \\
&\geq \sum p_i (1 - \frac{q_i}{p_i}) \\
&= \sum p_i - \sum q_i \\
&= 1 - 1 \\
&= 0
\end{aligned}
$$

(c)

$$
\begin{aligned}
KL(p,q) &= \sum_i p_i log(\frac{p_i}{q_i}) \\
&= \sum p_i (logp_i - logq_i) \\
&= \sum 2p_i (log\sqrt{p_i} - log\sqrt{q_i}) \\
&= \sum 2p_i (\log \frac{\sqrt{p_i}}{\sqrt{q_i}})
\end{aligned}
$$

According to (a), we have

$\log \frac{p}{q} \geq 1 - \frac{q}{p}$

Then,

$$
\begin{aligned}
KL(p,q) &\geq \sum 2p_i (1 - \frac{\sqrt{q_i}}{\sqrt{p_i}}) \\
&= \sum (2p_i - 2\sqrt{p_i}\sqrt{q_i}) \\
&= \sum (p_i + p_i - 2\sqrt{p_i}\sqrt{q_i}) \\
&= \sum (p_i + q_i - 2\sqrt{p_i}\sqrt{q_i}) (\because \sum p_i = \sum q_i = 1) \\
&= \sum (\sqrt{p_i} - \sqrt{q_i})^2
\end{aligned}
$$

(d)

Let a random variable space $\Omega = \{X\}$and$x \in \{0,1\}$

And assume $P(X=0)=0.7$ $P(X=1)=0.3$ $Q(X=0)=0.5$ $Q(X=1)=0.5$

Then

$KL(P,Q) = P(X=0)log\frac{P(X=0)}{Q(X=0)} + P(X=1)log\frac{P(X=1)}{Q(X=1)} = 0.0823$

$KL(Q,P) = Q(X=0)log\frac{Q(X=0)}{P(X=0)} + Q(X=1)log\frac{Q(X=1)}{P(X=1)} = 0.0872$

Thus,

$KL(P,Q) \neq KL(Q,P)$

1.7 Mutual Information

(a)

Let p_i to be $P(x,y)$ and q_i to be $P(x)P(y)$

then $\sum\sum P(x,y)log(\frac{P(x,y)}{P(x)P(y)}) \geq \sum\sum P(x,y)(1-\frac{P(x)P(y)}{P(x,y)}) = \sum\sum(P(x,y) - \sum P(x)P(y)) \geq 0$

$\because \sum\sum P(x,y) = 1$ and at the same time $\sum\sum P(x)P(y) \leq \sum\sum P(x,y)$

$\therefore I(X,Y) \geq 0$ only when $P(x)P(y) = P(x,y)$ the statement holds equal.

(b)

Let a group of random variablesX_1, X_2

And assume $P(X_1) = 0.7$ $P(X_2) = 0.3$ $Q(X_1) = 0.5$ $Q(X_2) = 0.5$

Then

$KL(P,Q) = P(A)log\frac{P(A)}{Q(A)} + P(B)log\frac{P(B)}{Q(B)} = 0.19$

$KL(Q,P) = Q(A)log\frac{Q(A)}{P(A)} + Q(B)log\frac{Q(B)}{P(B)} = 0.22$

Thus,

$KL(P,Q) \neq KL(Q,P)$

1.8 Compare and Contrast

(a)

$\because P(X,Y,Z) = P(X)P(Y|X)P(Z|X,Y) = P(X)P(Y|X)P(Z|Y)$

under condition X, Y and Z are indepent.

(b)

$\because P(X,Y,Z) = P(X)P(Y|X)P(Z|X,Y) = P(X)P(Y|X)P(Z|Y)$

$\therefore$ under condition Y, X and Z are indepent.

(c)

$\because P(X,Y,Z) = P(Z)P(Y|Z)P(X|Y,Z) = P(Z)P(Y|Z)P(X|Y)$

$\therefore$ under condition Y, X and Z are indepent.

1.9 Hangman

(a)

```
---Fifteen Most Words---
THREE 0.03562714868653127
SEVEN 0.023332724928853858
EIGHT 0.021626496097709325
```

```
WOULD 0.02085818430793947
ABOUT 0.020541544349751077
THEIR 0.018974130893766185
WHICH 0.018545160072784138
AFTER 0.01436452108630337
FIRST 0.014345603577470525
FIFTY 0.013942725872119989
OTHER 0.013836135494765265
FORTY 0.012387837111638222
YEARS 0.011598389898206841
THERE 0.01128553344178502
SIXTY 0.009535207245223231
---Fourteen Least Words---
CCAIR 9.13259047102901e-07
CLEFT 9.13259047102901e-07
FABRI 9.13259047102901e-07
FOAMY 9.13259047102901e-07
NIAID 9.13259047102901e-07
PAXON 9.13259047102901e-07
SERNA 9.13259047102901e-07
TOCOR 9.13259047102901e-07
YALOM 9.13259047102901e-07
BOSAK 7.827934689453437e-07
CAIXA 7.827934689453437e-07
MAPCO 7.827934689453437e-07
OTTIS 7.827934689453437e-07
TROUP 7.827934689453437e-07
```

(b)

correctly guessed	incorrectly guessed	best next guess l	$P(L_i = l \; for \; some \; i \in 1,2,3,4,5 \vert E)$
- - - - -	{}	E	0.5394
- - - - -	{A,I}	E	0.6214
A - - - R	{}	T	0.9816
A - - - R	{E}	O	0.9913
- - U - -	{O, D, L, C}	T	0.7045
- - - - -	{E, O}	I	0.6366
D - - I -	{}	A	0.8207
D - - I -	{A}	E	0.7521
- U - - -	{A, E, I, O, S}	Y	0.6270

(c)

```
import numpy as np
```

```
words = []
values = []
with open('./hw1_word_counts_05.txt') as f:
    data = f.readlines()
for i in range(len(data)):
    (word,count) = data[i].split(' ')
    words.append(word)
    values.append(int(count))
total = sum(values)
for i, item  in  enumerate(values):
    values[i] =values[i]/total
index = range(len(data))
sortIndex = sorted(index, key=lambda i: values[i], reverse=True)
fifteen_most = sortIndex[0:15]
fourteen_least = sortIndex[-14:]
print('---Fifteen Most Words---')
for item in fifteen_most:
    print(words[item],values[item])
print('---Fourteen Least Words---')
for item in fourteen_least:
    print(words[item],values[item])
```

```
"""
Definition Of Functions
"""
def guessNext(words,values,evidence,wrong_guess):
    sum_p = 0
    checks = [0]*len(words)
    #calculate the denominator of P(W|E)
    for i in range(len(words)):
        if isInWord(evidence,wrong_guess,words[i])==1:
            sum_p += values[i]
            checks[i] = 1
    prediction = [0]*26
    chars = 'ABCDEFGHIJKLMNOPQRSTUVWXYZ'
    #calculate pwe
    for i in range(len(chars)):
        if not chars[i] in wrong_guess and not chars[i] in evidence:
            for j in range(len(words)):
                if chars[i] in words[j] and checks[j]==1:
                    prediction[i] += values[j]/sum_p

    maxP = max(prediction)
    best = chars[prediction.index(maxP)]
    return (best, maxP)


def isInWord(evidence,wrong_guess,w):
    """
```

```
    is the evidence match the word
    """
    for i in range(5):
        if(evidence[i]!='$' and w[i]!=evidence[i]) or (evidence[i]=='$' and w[i] in 
evidence):
            return 0
        if w[i] in wrong_guess:
            return 0
    return 1
#Test Stage
print(guessNext(words,values,['$','$','$','$','$'],""))
print(guessNext(words,values,['$','$','$','$','$'],"AI"))
print(guessNext(words,values,['A','$','$','$','R'],""))
print(guessNext(words,values,['A','$','$','$','R'],"E"))
print(guessNext(words,values,['$','$','U','$','$'],"ODLC"))
print(guessNext(words,values,['$','$','$','$','$'],"EO"))
print(guessNext(words,values,['D','$','$','I','$'],""))
print(guessNext(words,values,['D','$','$','I','$'],"A"))
print(guessNext(words,values,['$','U','$','$','$'],"AIEOS"))
```