CSE 250A Hwk 1 – Andrew Ghafari A59020215.

1.1) a) $P(X|Y,E) \times P(Y|E) = \frac{P(X,Y,E)}{P(E)}$ (Chain rule) (RHS)

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

RHS = LHS → done.

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

$= \frac{P(Y|E,X).P(\cancel{E}).P(X/E)}{P(Y|E).P(\cancel{E})} = \frac{P(Y/E,x).P(X/E)}{P(Y/E)}$

using product rule & proof 1.1-a ⇒ done.

c) $P(X|E) = \frac{P(X.E)}{P(E)} = \frac{\sum_{i} P(X, Y=y_i, E)}{P(E)}$ (including all y_is)

(product rule) $:= \sum_{y} P(X, Y=y_i |E)$. proof done.

1.2) a- $P(X.Y|E) = P(X|Y.E) \times P(Y|E)$ (1.1-a)

$= P(X|E)\, P(Y/E)$ (1.2-1)

So (1) implies that $P(X|E) = P(X|Y.E)$ which is (2)

$P(X.Y/E) = \frac{P(X.Y.E)}{P(E)} = \frac{P(Y/X.E).P(X/E).P(E)}{P(E)}$ ✓

$\frac{P(X.Y.E)}{P(X.E)} \cdot \frac{P(X.E)}{P(E)} = P(Y/X.E).P(X/\cancel{E}) = P(\cancel{X/E})P(Y/E)$

$P(Y/X.E) = P(Y/E)$ 1 → 3.

Scanned with CamScanner

b) $P(X.Y|E) = P(X|Y.E) . P(Y/E)$ (proved before)

$= \underbrace{P(X|E)}_{(2)} . P(Y|E)$ (in (2))

So (2) implies that $P(X|E) . P(Y|E) = P(X.Y/E)$ (which is (1)) as proved in (a).

So (2) can imply also (1) & (3) as well.

c) $P(X|Y,E) = \dfrac{P(Y|X,E) . P(X|E)}{P(Y|E)}$ (1.1-b)

So $P(Y|X,E) = \dfrac{P(X|Y,E) \times P(Y|E)}{P(X|E)}$ (in 1.1-b) $= P(Y|E)$ in (3)

So (3) implies that $\dfrac{P(X|Y,E)}{P(X|E)} = 1$ meaning

$P(X|Y,E) = P(X|E)$ which is (2). (3) $\rightarrow$ (2).

Plus, $P(X,Y|E) = P(Y|X.E) . P(X|E)$ (as proved previously)

(3) implies $P(Y|X.E) = P(Y|E)$ (1)

given (1) as proved in (a), we can imply that

$P(X|Y.E) = P(X.E)$ so (3) implies (1) & (2).

Thus (1), (2), (3) are equivalent.

2 → 3:

$$P(X, Y \mid E) = P(Y \mid X, E) \times P(X \mid E)$$

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

using ~~P(Y | X, E)~~ = this in 1, we get.

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

which is 3

Problem 1.3

a) X: Person has covid $X \in \{0,1\}$ $X=1$ has covid

Y: Person coughs. $Y \in \{0,1\}$. $Y=1$: coughs

Z: Person doesn't taste/smell $Z \in \{0,1\}$ $Z=1$ doesn't taste/smell.

$P(X=1) < P(X=1/Y=1) < P(X=1/Y=1, Z=1)$ => Cumulative evidence

b) X: Person has covid $X \in \{0,1\}$ $X=1$ has covid

Y: Person coughs $Y \in \{0,1\}$ $Y=1$ coughs

Z: Person has a negative PCR result $Z \in \{0,1\}$

$Z=1 \rightarrow$ he got a $\ominus$ result

$P(X=1) < P(X=1/Y=1)$ $P(X=1/Y=1) > P(X=1/Y=1, Z=1)$

c) X: John buys an umbrella. X=1 he did
Y: Mary bought an umbrella Y=1 she did
Z: A thunderstorm is arriving on their city (Z=1 arriving)

X, Y are not completely independent if we assume that John & Mary live in same city

$P(X,Y) \neq P(X)P(Y)$.

Now given Z, John's decision to buy an umbrella will no longer have an effect on Mary since her decision is now solely relying on Z=1. and vice versa.

$$P(X=1, Y=1 \mid Z=1) = P(X=1 \mid Z=1)\, P(Y=1 \mid Z=1)$$

1.4) D $\rightarrow$ T D = 1 uses drugs

T = 1 test is (+).

$P(D=1) = 0.01$ ~~P(T=1)~~ =

$P(D=0) = 0.99$

P(T|D)	D	T	
0.90	1	1	
0.95	0	0	
0.05	0	1	F.P.
0.1	1	0	F.N.

complements

b) $P(T=0) = P(T=0 \mid D=0) \cdot P(D=0) + P(T=0 \mid D=1) \cdot P(D=1)$

$= P(T=0, D=0) + P(T=0, D=1)$

$\Rightarrow 0.95 \times 0.99 + 0.1 \times 0.01$

$= 0.9415.$

~~P(D=0) = P(D=0~~

$P(D=0 \mid T=0) = \dfrac{P(D=0, T=0)}{P(T=0)} = \dfrac{0.95 \times 0.99}{0.9415}$

= ~~P(T=0). P(D~~ $= 0.999.$

c) $P(D=1 \mid T=1) = \dfrac{P(D=1, T=1)}{P(T=1)} = \dfrac{P(D=1) \times P(T=1 \mid D=1)}{P(T=1)}$

$= \dfrac{0.01 \times 0.9}{1 - P(T=0)} = \dfrac{0.009}{1 - 0.9415}$

≈ 0.154

1.5) $H(X) = -\sum_{i=1}^{n} p_i \log p_i = -p_1 \log p_1 - p_2 \log p_2 \dots - p_n \log p_n$.

$\nabla_H = \dfrac{d(-p_i \log p_i)}{dp_i} = \dfrac{d(-p_i)}{dp_i} \cdot \log p_i + \dfrac{d(\log p_i)}{dp_i} \cdot -p_i$

$= -\log p_i \times \dfrac{1}{p_i} \times -p_i = -1 - \log p_i$

$\nabla H = \begin{bmatrix} -1-\log p_1 \\ -1-\log p_2 \\ \vdots \\ -1-\log p_n \end{bmatrix}$ $f(x) = \sum_i p_i$ $\dfrac{d(p_i)}{dp_i} = 1$ $\nabla f = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$

$\nabla H = \lambda \nabla f \to \lambda = \nabla H = -1 - \log p_i$ for $i = 1 \to n$.

so $p_1 = p_2 = p_3 \text{———} p_n$ w/ $\sum p_i = 1$ & all p_is are equal. $p_i = \frac{1}{n}$ $\forall i$ in $\in [1, n]$.

b) $H(X_i) = -p(x_i) \log p(x_i)$ w/ $\sum_i p(x_i) = 1$.

$H(X_1, X_2, X_3 \text{—} X_n) = -\sum_{x_1} \sum_{x_2} \sum_{x_n} P(x_1, x_2, x_n) \log(P(x_1, x_2, - x_n))$

w/ $P(x_1, x_2, -- x_n) = P(x_1) \times P(x_2) \text{—} \times P(x_n)$

$(\prod_{i=1}^{n} P_{x_i})$

we get

$H(X_1 \text{—} X_n) = -\sum_{x_1} \sum_{x_2} \text{—} \sum_{x_n} P(x_1) P(X_2) P(X_3) -- P(X_n) \cdot \log(P_{x_1} \cdot P_{x_2} -- P_{x_n})$

w/ $\log(ab) = \log a + \log b$.

$$= -\sum_{x_1}\sum_{x_2}\sum_{x_3}\sum_{x_n} \prod p(x_i) \cdot \left(\sum_i^n \log p(x_i)\right)$$

$$= -\underbrace{\sum_{x_1} p(x_1) \sum_{x_2} p(x_2) \text{ — } \sum_{x_{n-1}} p(x_{n-1})} \cdot \sum p_{x_n}\left(\sum_{i=1}^{n-1} \log p x_i + \log p x_n\right)$$

$$\cdot\left(\sum_{i=1}^{n-1} \log p x_i - H(X_n)\right)$$

1 iteration more

(A)

$$= -\left(\sum p x_1 \sum p x_2 \text{ — } \sum_{n-2} p x_{n-2}\right) \cdot \sum_{n-1} p x_{n-1}\left(\sum_i^{n-1} \log p(x_i) - H(X_n)\right)$$

$$\underbrace{\sum p x_{n-1}\left(\sum_i^{n-2} \log(p(x_i)) + \log p(x_{n-1}) - H(X_n)\right)}$$

and it keeps unfolding.

$$= -H(X_n) - H(X_{n-1}) + \underbrace{\sum_{i=1}^{n-2} \log p(x_i)}_{\text{marginalized}}$$

$\sum_{i=1}^{n} H(X_i) \Leftarrow$ (The minus sign & the ones of $H(X_i)$ will cancel out) to give me $\sum_{i=1}^{n} H(X_i)$

proved.

This will keep happening until all $p(x_i)$s are coupled w/ their respective $\log(p x_i)$ to form an $H(X_i)$ which will lead to $\sum H(X_i)$

I will repeat line (A) just to be a bit clearer:

$$= -\sum_{x_1} p x_1 \sum_{x_2} p x_2 \text{ — } \sum_{n-2} p x_{n-2} \cdot \sum_{x_{n-1}} p x_{n-1} \times \left(\sum_1^{n-1} \log p(x_i) - H(X_n)\right)$$

$$= -\sum_{x_1} p(x_1) \times \text{—} \sum_{n-1} p x_{n-2}\left(-H(X_n) + \log p(x_{i=n-1}) + \sum_{i=1}^{n-2} \log p(x_i)\right)$$

$$= -\sum_{x_1} p(x_1) \text{—} \sum_{n-2} p x_{n-2}\left(-H(X_n) - H(X_{n-1}) + \sum_{n-1} p(x_{n-1}) \sum_1^{n-2} \log p(x_i)\right)$$

$$= + \sum^{n-2} \log p(x_i) \quad \text{marginal}$$

Problem 1.6

$$KL = \sum_i p_i \log\left(p_i / q_i\right)$$

a) let's first prove equality @ 1.

$\log(1) = 0 \quad x - 1 \ (@ x = 1) = 1 - 1 = 0. \implies$ equality.

let $g(x) = \log(x) - (x - 1) \qquad x \in]0 ; +\infty[$

or else $\log(x)$ doesn't exist

$$\frac{dg}{dx} = \frac{1}{x} - 1$$

Sign table:

	0^+		1		$+\infty$
$\frac{dg}{dx}$		+	0	−	
$g(x)$		$\nearrow$	0	$\searrow$	

Using sign of $\frac{dg}{dx}$ we can get that $g(x)$ is monotonically increasing before $x = 1$ then proceeds to decrease after that. Which means that @ $x = 1$ $g(x)$ reaches its maximum $(g(x) = 0) \longrightarrow g(x) \leq 0 \quad \forall x \in]0^+ ; +\infty[$

$$\log(x) - (x - 1) \leq 0 \implies \log(x) \leq x - 1$$

Scanned with CamScanner

b) $KL(p,q) = \sum p_i \log \frac{p_i}{q_i} = -\sum p_i \log \frac{q_i}{p_i}$

given the inequality in a) we can say that.

$$-\sum p_i \log \frac{q_i}{p_i} \geq -\sum p_i \left(\frac{q_i}{p_i} - 1\right)$$

$$-\sum p_i \log \frac{q_i}{p_i} \geq -\sum_i q_i - p_i$$

which is always 0.

$$KL(p,q) \geq \sum(p_i - q_i) \rightarrow \text{so } KL(p,q) \geq 0.$$

when $p_i = q_i$ $\left.\begin{matrix} KL(p,p) = 0 \\ \sum p_i \log(1) = 0 \end{matrix}\right] = \sum p_i - p_i$ (equality proven)

c) $KL(p,q) = \sum p_i \log \frac{p_i}{q_i}$

like we did in (b)

$$KL(q,q) = -\sum p_i \log\left(\frac{q_i}{p_i}\right) = -\sum p_i \left(\frac{q_i}{p_i} - 1\right)$$

but $-\sum p_i \log\left(\frac{q_i}{p_i}\right) = -2\sum p_i \log\sqrt{\frac{q_i}{p_i}}$
(using log formulas)

$$-2\sum p_i \log\sqrt{\frac{q_i}{p_i}} \geq -2\sum p_i \left(\sqrt{\frac{q_i}{p_i}} - 1\right)$$

$$KL(p,q) \geq -2\sum\left(\sqrt{p_i q_i} - p_i\right)$$

$$KL(p,q) \geq -\sum\left(2\sqrt{p_i q_i} - 2p_i\right)$$

$$KL(p,q) \geq \sum 2p_i - 2\sqrt{p_i q_i}$$

given $\sum p_i = \sum q_i = 1$
we can include $\sum q_i$ to make

$$KL(p,q) \geq \sum p_i + q_i - \sqrt{2 p_i q_i}$$

$$\text{or } p_i + q_i - \sqrt{2 q_i p_i} = (\sqrt{p_i} - \sqrt{q_i})^2$$

$$\Rightarrow KL(p,q) \geq (\sqrt{p_i} - \sqrt{q_i})^2 \quad (\text{proven}).$$

d)
Any probabilistic distributions that aren't equal will produce different KL values for (p,q) & (q,p).

example:

$P \in \{0,1\}$ $P(P=0) = P(P=1) = 1/2$
$Q \in \{0,1\}$ $P(Q=0) = 0.1$ $P(Q=1) = 0.9$.

Using formula. $KL(P,Q) = \sum_i p_i \log \frac{p_i}{q_i}$

$$= \sum (p=0) \log\left(\frac{p=0}{q=0}\right) + \sum (p=1) \log\left(\frac{p=1}{q=1}\right)$$

$$= 0.5 \log\left(\frac{0.5}{0.1}\right) + 0.5 \log\left(\frac{0.5}{0.9}\right) = 0.2218.$$

In contrast:

$$KL(q,p) = \sum q=0 \log \frac{q=0}{p=0} + \sum q=1 \log \frac{q=1}{p=1}$$

$$= 0.1 \log\left(\frac{0.1}{0.5}\right) + 0.9 \log\left(\frac{0.9}{0.5}\right) = 0.1598$$

different $\Rightarrow$ proven.

Scanned with CamScanner

1.7)a - Using result of 1.6:

$$I(X;Y) = \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

$= D_{KL}(p,q)$ / given $p = p(x,y)$ $q = p(x)p(y)$

as proved above. $D_{KL}(p,q) \geq 0$.

w/ equality being when $p = q$.

So $I(X;Y) \geq 0$ & equality holds

when $X = Y$, when $p(X,Y) = p(x)p(y)$

so when X, Y are independent.

(which also kind

b) Indep. = $P(x,y) = P(x)P(y)$

if $I(X;Y) = 0$. $\sum_x \sum_y P(x,y) \log \frac{P(x,y)}{P(x)(P(Y)} = 0$.

while $P(x,y) = 0$. doesnt make a lot of sense.

$\log \frac{P(x,y)}{P(x)P(y)} = 0$ means $P(x,y) = P(x)P(y)$

because only $\log(1) = 0$.

which is the independance rule $\Rightarrow X, Y$ are independent.

1.8

a) Yes, BN_1 implies Y and Z are conditionally indep. given X, which is not the case in BN_2

b) No, Both contain same information as per conditional & marginal probabilities.

c) Yes, BN_3 states that X & Z are cond. independent which is not the case for BN_1 (direct edge between X & Z).

Problem 1.9 - Part a

```
import operator
import numpy as np
with open("hw1_word_counts_05.txt", 'r') as f:
  contents = f.readlines()

dicti = {}
for item in contents:
    word, num = item.split()
    dicti[word] = int(num)

sorted_dicti = dict(sorted(dicti.items(), key=operator.itemgetter(1),reverse=True))
# print(dicti)
keysList = list(sorted_dicti.keys())
valsList = list(sorted_dicti.values())
#dictionary holding prob of each word
count = len(keysList)
total = sum(valsList)
prob = {}
for i in range(count):
  prob[keysList[i]] = ( valsList[i]/total)
print(prob)
print("15 Most common words")
print(keysList[:15])
print("14 Least common words")
print(keysList[-14:])
```

```
{'THREE': 0.03562714868653127, 'SEVEN': 0.023332724928853858, 'EIGHT': 0.02162649609770
15 Most common words
['THREE', 'SEVEN', 'EIGHT', 'WOULD', 'ABOUT', 'THEIR', 'WHICH', 'AFTER', 'FIRST', 'FIFTY
14 Least common words
['CCAIR', 'CLEFT', 'FABRI', 'FOAMY', 'NIAID', 'PAXON', 'SERNA', 'TOCOR', 'YALOM', 'BOSAK
```

Yes, results do make sense.

Part b

Removing words that are not compatible with our current setting

```
import string
```

```
def remove_corr(corr_guess):
  temp = {}
  output = []
  alphabet = list(string.ascii_uppercase)
  for elemi in corr_guess:
      output.append(elemi[1])
      if elemi[0] in alphabet:
        alphabet.remove(elemi[0])
  # print(output)
  # print(alphabet)
  for elem in dicti:
    f = 0
    for i in range(5):
      if(i in output):
        for element in corr_guess:
          if element[1] == i:
            x = element[0]
        if(elem[i] != x):
          f = 1
          break
      else:
        if(elem[i] not in alphabet):
          f = 1
          break
    if(f == 0):
      temp[elem] = dicti[elem]
  return temp
```

Removing words with incorrect letters in them

```
  def removeincorr (incorr_guess, temp):
    final = {}
    for elem in temp:
      f = 0
      for i in range(len(incorr_guess)):
        if(incorr_guess[i] in elem):
          f = 1
          break
      if(f == 0):
        final[elem] = temp[elem]
    return final


# temp = remove_corr([["D",0],["I",3]])
# removeincorr(["A"] , temp)
```

Combining these two sub functions into one:

```
def validwords(dicti , corr_guess, incorr_guess):
  if corr_guess == [[]]:
    temp = dicti
  else:
    temp = remove_corr(corr_guess)
  if(incorr_guess == []):
    return temp
  else:
    final = removeincorr( incorr_guess , temp)
  return final


# validwords(dicti,[["U",1]],["A","E","I","O","S"])


import string
def rec_lett(dicti , corr_guess, incorr_guess):
  space = validwords(dicti , corr_guess, incorr_guess)
  rec_dict = {}
  total_count = sum(space.values())
  for elem in space:
    seti = set()
    for i in range(5):
      if(elem[i] not in rec_dict and elem[i] not in seti):
        rec_dict[elem[i]] = space[elem]
        f = 1
      elif(elem[i] in rec_dict and elem[i] not in seti):
        rec_dict[elem[i]] += space[elem]
        f = 1
      seti.add(elem[i])
  if(corr_guess == [[]]):
    answer = dict(sorted(rec_dict.items(), key=operator.itemgetter(1),reverse=True))
    # print("Answer dictionary:", answer)
    # print("total count:", total_count)
    res = list(answer.keys())[0]
    value = answer[str(res)]
    probability = value/total_count
    return (res,probability)
  else:
    final_output = {}
    for element in rec_dict:
      flag = 0
      for i in range(len(corr_guess)):
        if corr_guess[i][0] == element:
          flag = 1
          break
      if (flag == 0):
```

```
            final_output[element] = rec_dict[element]
        answer = dict(sorted(final_output.items(), key=operator.itemgetter(1),reverse=True))
        # print("total count:", total_count)
        res = list(answer.keys())[0]
        value = answer[str(res)]
        probability = value/total_count
        return(res,probability)
```

```
print(rec_lett(dicti, [[]] ,[]))
print(rec_lett(dicti, [[]] ,["E", "A"]))
print(rec_lett(dicti, [["A",0] , ["S" , 4]] ,[]))
print(rec_lett(dicti, [["A",0] , ["S" , 4]] ,["I"]))
print(rec_lett(dicti, [["O",2]] ,["A","E","M","N","T"]))
print(rec_lett(dicti, [[]] ,["E","O"]))
print(rec_lett(dicti, [["D",0] , ["I" , 3]] ,[]))
print(rec_lett(dicti, [["D",0] , ["I" , 3]]  ,["A"]))
print(rec_lett(dicti, [["U",1]] ,["A","E","I","O","S"]))
```

```
('E', 0.5394172389647974)
('O', 0.5340315651557658)
('E', 0.7715371621621622)
('E', 0.7127008416220352)
('R', 0.7453866259829712)
('I', 0.6365554141009611)
('A', 0.8206845238095238)
('E', 0.7520746887966805)
('Y', 0.626965110163053)
```

✓ 0s completed at 11:12 PM