CS 688 Graphic modeling spring 2023

Homework Set 2 Group Members: ByungKwon Moon

1.
```
     x_0            x_1            x_2            x_3
e[-1.0227226e+01 -1.6412308e+01 -2.0317350e+00 -1.9390703e+01]
t[-1.2470186e+01 -9.8640760e+00  9.4265310e+00  1.2294582e+01]
a[ 7.3560660e+00  1.7699176e+01 -2.6818590e+00  6.3192180e+00]
i[-1.8053115e+01 -8.4178340e+00  1.0548553e+01 -1.4567270e+00]
n[ 1.4714884e+01  9.8294460e+00 -6.4720920e+00  2.4403760e+00]
o[-1.2466160e+00 -3.8462030e+00 -9.9398960e+00  2.5276000e-02]
s[-6.0528380e+00 -1.0704701e+01  6.5877370e+00 -1.1689033e+01]
h[ 2.6628975e+01  6.1158580e+00 -8.1612090e+00  4.9050000e-01]
r[-7.0835950e+00  1.0228850e+01  5.5461120e+00  1.2221587e+01]
d[ 6.4336310e+00  5.3717740e+00 -2.8221590e+00 -1.2550800e+00]
```

2.
```
67.58659400000002  test img 1
72.80359200000001  test img 2
96.894979          test img 3
```

3.
```
68.43903558288955   test img1
93.59553578849992   test img2
112.27780334018632  test img3
```

4.
```
label and probability
['h', 'a', 'i', 'r'] 0.42637263590803115
label and probability
['r', 'i', 'a', 'h', 'o'] 0.8724928772929964
label and probability
['s', 'a', 'r', 'i', 'n', 'e'] 0.517546791182391
```

5.
```
x_0             x_1             x_2             x_3
e[[7.74321274e-17 1.17013562e-15 1.98388871e-06 7.00413262e-15]
t[1.32038261e-17 1.12938380e-12 1.73144364e-01 4.59779368e-01]
a [3.23129706e-09 9.99198294e-01 1.03714052e-06 1.60874784e-03]
i[4.67423473e-20 2.50292737e-12 8.16315532e-01 4.35098220e-07]
n [7.17352122e-06 3.23689725e-04 2.15557413e-08 3.16174530e-05]
o [6.31721567e-13 3.42375953e-10 7.56027640e-10 2.45052421e-06]
s [4.76411621e-15 4.22769645e-13 5.96936877e-03 2.23321135e-11]
 h[9.99992822e-01 6.24632206e-06 5.31868344e-09 3.96444598e-06]
 r[2.66391815e-15 4.68843355e-04 4.56691974e-03 5.38572790e-01]
 d[1.72873581e-09 2.92606574e-06 7.66644245e-07 6.26689023e-07]]
```

6.
```
   m1->2(y2)               m2->1(y1)            m4->3(y3)            m3->2(y2)
e [26.901502683957137, [41.29462875861789, [13.157751245692122, [23.788179945699227,
t 26.994461487243363, 41.34959924155791,    12.602158310641181,  23.369117084692817,
a 26.67631917869039,  41.744814612198624,   13.195828473708328,  24.274913362421113,
i 26.551264839495097, 41.65265508688914,    13.091537665643418,  23.732845513829805,
n 26.511586298494358, 41.618278558555794,   12.765968749876375,  23.801951189937387,
o 26.63704053815263,  41.5851287383059,     13.167221989937278, 24.052493344824544,
s 26.619264831144616, 41.05557278761842,    12.602507314335533, 23.604069024142888,
h 26.492810829112617, 41.77087555978143     13.124655254219011,  24.023530690819996,
r 26.538931703710706, 41.82209656018796,    12.92685298949019,  23.921153538504193
d 26.510532325860822] 41.312413831345815]   12.736933487870395] 23.545596050916394]
```

7.
```
 x_0              x_1              x_2              x_3
e[[7.74321274e-17 1.17013562e-15 1.98388871e-06 7.00413262e-15]
t[1.32038261e-17 1.12938380e-12 1.73144364e-01 4.59779368e-01]
a [3.23129706e-09 9.99198294e-01 1.03714052e-06 1.60874784e-03]
i[4.67423473e-20 2.50292737e-12 8.16315532e-01 4.35098220e-07]
n [7.17352122e-06 3.23689725e-04 2.15557413e-08 3.16174530e-05]
o [6.31721567e-13 3.42375953e-10 7.56027640e-10 2.45052421e-06]
s [4.76411621e-15 4.22769645e-13 5.96936877e-03 2.23321135e-11]
 h[9.99992822e-01 6.24632206e-06 5.31868344e-09 3.96444598e-06]
 r[2.66391815e-15 4.68843355e-04 4.56691974e-03 5.38572790e-01]
 d[1.72873581e-09 2.92606574e-06 7.66644245e-07 6.26689023e-07]]

3*3
pair (1,2)
    e              t           r
e[[6.05758711e-32 8.71199022e-29 6.74306327e-20]
t [1.05142518e-32 4.26233596e-30 4.82842055e-21]
r [2.71559306e-30 1.61120779e-27 1.23402616e-18]] 1 2


pair(2,3)
    e              t               r
e[[3.18813562e-21 2.83198789e-16 9.56267753e-18]
t [3.99349515e-18 9.99906428e-14 4.94157096e-15]
r [1.59398984e-09 5.84129721e-05 1.95177784e-06]]
parin (3,4)
    e              t               r
e[[1.00177154e-20 9.44271705e-07 1.03780372e-06]
t [2.49192257e-15 6.62086032e-02 1.06500152e-01]
r [5.61637407e-17 2.18400418e-03 2.37522453e-03]]
```

8.
```
['h', 'a', 'i', 'r']
['r', 'i', 'a', 'h', 'o']
['s', 'a', 'r', 'i', 'n', 'e']
['s', 'i', 'n', 's']
['s', 'e', 'n', 'd']
average = 0.899
```

9.
$$p_\theta(y,x) = \frac{1}{Z(\theta)} \exp\left(\sum_{i=1}^{L}\sum_{f=1}^{F}\theta^F_{y_i f}x_{if} + \sum_{i=1}^{L-1}\theta^T_{y_i y_{i+1}}\right).$$

then average log likliehood is

$$\frac{1}{N}\sum_{n=1}^{N} log p_\theta(y^n, x^n) = \frac{1}{N}\sum_{n=1}^{N} log\left(\frac{1}{Z(\theta,x^n)}\exp\sum_{i=1}^{L}\sum_{f=1}^{F}\theta^F_{y_i^{(n)} f}x_{if}^{(n)} + \sum_{i=1}^{L-1}\theta^T_{y_i^{(n)} y_{i+1}^{(n)}}\right).$$
$$= \frac{1}{N}\sum_{n=1}^{N}\left(-log Z(\theta,x^n) + \sum_{i=1}^{L}\sum_{f=1}^{F}\theta^F_{y_i^{(n)} f}x_{if}^{(n)} + \sum_{i=1}^{L-1}\theta^T_{y_i^{(n)} y_{i+1}^{(n)}}\right).$$

10. first we start with

$$\mathcal{L}(\theta) = \frac{1}{N}\sum_{n=1}^{N}\log p_\theta(y^{(n)}x^{(n)}) \tag{1}$$

$$\frac{d\mathcal{L}}{d\theta_{c,f}} = \frac{d}{d\theta_{c,f}}\frac{1}{N}\sum_{n=1}^{N}\left(-log Z(\theta,x^n) + \sum_{i=1}^{L}\sum_{f=1}^{F}\theta^F_{y_i^{(n)} f}x_{if}^{(n)} + \sum_{i=1}^{L-1}\theta^T_{y_i^{(n)} y_{i+1}^{(n)}}\right)$$
$$= \frac{1}{N}\sum_{n=1}^{N}\left(-\frac{d}{d\theta_{c,f}}log Z(\theta,x^n) + \sum_{i=1}^{L}\sum_{f=1}^{F}\frac{d}{d\theta_{c,f}}\theta^F_{y_i^{(n)} f}x_{if}^{(n)} + \sum_{i=1}^{L-1}\frac{d}{d\theta_{c,f}}\theta^T_{y_i^{(n)} y_{i+1}^{(n)}}\right)$$

we know that θ^T is constant

$$\sum_{i=1}^{L}\sum_{f=1}^{F}\frac{d}{d\theta_{c,f'}}\theta^F_{y_i^{(n)} f}x_{if}^{(n)} = \sum_{i=1}^{L}\sum_{f=1}^{F}1[y_i^{(n)} = c, f = f']x_{if}^n$$

$$-\frac{d}{d\theta_{c,f}}log Z(\theta,x^n) = \frac{d}{d\theta_{c,f}}\frac{1}{Z(\theta,x)}\sum_{y}\exp\left(\sum_{i=1}^{L}\sum_{f=1}^{F}\theta^F_{y_i f}x_{if} + \sum_{i=1}^{L-1}\theta^T_{y_i y_{i+1}}\right).$$
$$= \sum_{y}p(y|x)1[y_i^{(n)} = c, f = f']x_{if}^n = E_{P(y|x)}1[y_i^{(n)} = c, f = f']x_{if}^n$$
$$= \frac{1}{N}\sum_{n=1}^{N}\left(\sum_{i=1}^{L}\sum_{f=1}^{F}1[y_i^{(n)} = c, f = f']x_{if}^n - E_{P(y|x)}1[y_i^{(n)} = c, f = f']x_{if}^n\right)$$

11.

$$\frac{d\mathcal{L}}{d\theta_{c,c'}} = \frac{1}{N}\sum_{n=1}^{N}\left(-\frac{d}{d\theta_{c,c'}}logZ(\theta,x^n) + \sum_{i=1}^{L}\sum_{f=1}^{F}\frac{d}{d\theta_{c,c'}}\theta^F_{y_i^{(n)}f}x_{if}^{(n)} + \sum_{i=1}^{L-1}\frac{d}{d\theta_{c,c'}}\theta^T_{y_i^{(n)}y_{i+1}^{(n)}}\right)$$

we know that θ^F is constant

$$\sum_{i=1}^{L-1}\frac{d}{d\theta_{c,c'}}\theta^T_{y_i^{(n)}y_{i+1}^{(n)}} = \sum_{i=1}^{L-1}1[y_i^{(n)} = c, y_{i+1}^{(n)} = c']$$

$$-\frac{d}{d\theta_{c,c'}}logZ(\theta,x^n) = -\frac{d}{d\theta_{c,c;}}\frac{1}{Z(\theta,x)}\sum_{y}\exp\left(\sum_{i=1}^{L}\sum_{f=1}^{F}\theta^F_{y_if}x_{if} + \sum_{i=1}^{L-1}\theta^T_{y_iy_{i+1}}\right)$$

$$= -\frac{1}{Z(\theta,x)}\sum_{y}\exp\left(\sum_{i=1}^{L-1}\theta^T_{y_iy_{i+1}}1[y_i^{(n)} = c, y_{i+1}^{(n)} = c']\right)$$

$$= -E_{P(y|x)}1[y_i^{(n)} = c, y_{i+1}^{(n)} = c']$$

$$\frac{d\mathcal{L}}{d\theta_{c,c'}} = \frac{1}{N}\sum_{n=1}^{N}\left(\sum_{i=1}^{L-1}1[y_i^{(n)} = c, y_{i+1}^{(n)} = c'] - E_{P(y|x)}1[y_i^{(n)} = c, y_{i+1}^{(n)} = c']\right)$$

12. Explain how, as a byproduct of the sum-product algorithm's computation of the single-variable and pairwise marginal probabilities, you can efficiently compute both the value of the log-likelihood function and the values of the above derivatives
when you use byproduct of sum product it can be make it as
$p(x,y) \rightarrow \sum_y \phi^F(x_i,y_i)m_{i+1\rightarrow i}(y_i)$
this is technically reduced from all the backward sequences which is (i+2,i+3 ... backward edge). Thus, we technically, compute sum over the marginal for each sequence to get log-likelihood.
similar as deriviative, p(y—x) we can compute with using marginal probabilities. so in each iterations we can use mariginal that already computed.

13. -3.6179466372349793 is average

14.

15. $f_\theta(x,y) = -(1-x)^2 - 50(y-x^2)^2$

$\frac{df}{dx} = -(-2(1-x)) - 50(-4x(-x^2+y)) = 2\,(1-x) + 200x\,(y-x^2)$
$\frac{df}{dy} = -100(y-x^2)$

16. x,y max = [-1.0,1.0] maxval = 4.0