In [68]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [6]:

```
id_file = 'hw8_ids.txt'
movie_file = 'hw8_movies.txt'
probR_file = 'hw8_probR_init.txt'
probZ_file = 'hw8_probZ_init.txt'
rating_file = 'hw8_ratings.txt'
```

In [23]:

```
def loadFile(filename, multi_col=False):
    contents = []
    with open(filename) as file:
        for line in file:
            temp = line.strip('\n').strip()
            if multi_col:
                temp = temp.split()
            contents.append(temp)

    return np.asarray([np.array(x) for x in contents])
```

In [167]:

```
id_df = loadFile(id_file)
movie_df = loadFile(movie_file)
probR_df = loadFile(probR_file, True).astype(np.float64)
probZ_df = loadFile(probZ_file).astype(np.float64)
rating_df = loadFile(rating_file, True)
```

In [168]:

```
recommended_df = np.zeros((rating_df.shape[1],))
seen_df = np.zeros((rating_df.shape[1],))
```

In [38]:

```
for row in range(rating_df.shape[0]):
    for col in range(rating_df.shape[1]):
        if rating_df[row][col] == '1':
            recommended_df[col] += 1
        if rating_df[row][col] != '?':
            seen_df[col] += 1
```

In [54]:

```
(recommended_df/seen_df)
```

Out[54]:

```
dtype('float64')
```

In [41]:

```
popularity_df = recommended_df/seen_df
```

In [47]:

```
popularity_df = np.vstack((movie_df, popularity_df)).transpose((1, 0))
```

In [59]:

```
popularity_df = pd.DataFrame(data=popularity_df, columns=["movie", "popularity"])
```

## 8.1 a)

In [67]:

```
with pd.option_context('display.max_rows', None, 'display.max_columns', None):  # more opti
    print(popularity_df.sort_values(by=['popularity'], ascending=False))
```

```
                                         movie          popularity
0                                    Inception  0.9657534246575342
3                                Shutter_Island  0.9444444444444444
21                        The_Dark_Knight_Rises  0.9416342412451362
47                                  The_Martian  0.9393939393939394
39                                 Interstellar  0.9385665529010239
40                     The_Theory_of_Everything  0.9195402298850575
11   Harry_Potter_and_the_Deathly_Hallows:_Part_2  0.9178082191780822
1                            The_Social_Network  0.9016393442622951
5    Harry_Potter_and_the_Deathly_Hallows:_Part_1  0.9006622516556292
37                                    Gone_Girl  0.8944099378881988
65                                The_Lion_King  0.8943661971830986
28                           Wolf_of_Wall_Street  0.8916666666666667
50                       Avengers:_Infinity_War  0.8785714285714286
62                            Avengers:_Endgame  0.8681318681318682
31                               Now_You_See_Me  0.8649789029535865
23                             Django_Unchained  0.8639455782312925
69                                 The_Farewell  0.8604651162790697
20                                 The_Avengers  0.8436482084690554
70                                     Parasite  0.8363636363636363
2                                    Black_Swan  0.8297872340425532
64                     Spiderman:_Far_From_Home  0.8271604938271605
26                               21_Jump_Street  0.8253968253968254
25                               Les_Miserables               0.825
67                                        Joker               0.824
19                The_Perks_of_Being_a_Wallflower  0.8198198198198198
6                                    Iron_Man_2  0.8178694158075601
33                              12_Years_a_Slave  0.8163265306122449
22                              The_Hunger_Games  0.8148148148148148
7                                    Toy_Story_3   0.803347280334728
9                                          Thor  0.8031496062992126
48                            The_Hateful_Eight                 0.8
55                               Hidden_Figures  0.7974683544303798
51                              Ready_Player_One  0.7973856209150327
13                The_Girls_with_the_Dragon_Tattoo  0.7931034482758621
10             Captain_America:_The_First_Avenger  0.7913669064748201
32                                          Her  0.7888888888888889
58                                      Dunkirk  0.7868852459016393
29                              The_Great_Gatsby  0.7864077669902912
61                                Darkest_Hour  0.7848101265822784
59                Three_Billboards_Outside_Ebbing  0.7741935483870968
63                  Once_Upon_a_Time_in_Hollywood   0.768595041322314
56                        Manchester_by_the_Sea  0.7636363636363637
17                            Midnight_in_Paris  0.7560975609756098
54                                   La_La_Land  0.7547169811320755
45                      Avengers:_Age_of_Ultron  0.7424242424242424
38                                   Ex_Machina  0.7391304347826086
49                                 The_Revenant  0.7363636363636363
14                         X-Men:_First_Class  0.7360406091370558
30                                       Frozen  0.7241379310344828
43                              Jurassic_World  0.7209302325581395
24                                Pitch_Perfect                0.72
16                                     The_Help  0.7068965517241379
72                                           Us                 0.7
15                                        Drive  0.6901408450704225
42                           Mad_Max:_Fury_Road  0.6783216783216783
```

```
8                                 Fast_Five   0.6588235294117647
73             Pokemon_Detective_Pikachu   0.6566265060240963
71                                 Good_Boys              0.65625
46                                      Room   0.6515151515151515
75                Terminator:_Dark_Fate   0.6388888888888888
35                        American_Hustle   0.6282051282051282
36                           Man_of_Steel   0.6242774566473989
12                             Bridemaids   0.6190476190476191
53                         Chappaquidick                  0.6
66                               Rocketman   0.5967741935483871
41          Star_Wars:_The_Force_Awakens   0.5935828877005348
34                              World_War_Z   0.5895522388059702
60                           Phantom_Thread   0.5862068965517241
18                              Prometheus    0.584070796460177
57                      The_Shape_of_Water   0.5584415584415584
68        Fast_&_Furious:_Hobbs_&_Shaw   0.5405405405405406
27                               Magic_Mike   0.5151515151515151
4                       The_Last_Airbender  0.47368421052631576
74                                 Hustlers  0.45652173913043476
44                      Fifty_Shades_of_Grey  0.37714285714285717
52                             I_Feel_Pretty    0.358974358974359
```

## 8.1 e)

In [123]:

```
rating_df[rating_df == '?'] = '2'
```

In [125]:

```
rating_df = rating_df.astype(np.int32)
```

In [169]:

```
likelihood_table = np.empty((rating_df.shape[0],))
P_Z_table = np.empty((probZ_df.shape[0]))
P_R_Z_table = np.empty((rating_df.shape[0], probZ_df.shape[0]))
estep_table = np.empty((rating_df.shape[0], probZ_df.shape[0]))
```

In [170]:

```
def get_P_R_Z(t_student, z_i):
    prod = 1
    for p in range(probR_df.shape[0]):
        if rating_df[t_student][p] != '?':
            if rating_df[t_student][p] == '1':
                prod = prod*probR_df[p][z_i]
            else:
                prod = prod*(1 - probR_df[p][z_i])
    return prod
```

In [171]:

```
def get_likelihood(t_student):
    likelihood = 0
    for z in range(probZ_df.shape[0]):
        #likelihood += get_P_R_Z(t_student, z) * probZ_df[z]
        likelihood = likelihood + (P_R_Z_table[t_student][z] * probZ_df[z])
    return likelihood
```

In [172]:

```
def get_estep(t_student, z_i):
    #return (probZ_df[z_i] * get_P_R_Z(t_student, z_i)) / get_likelihood(t_student)
    return (probZ_df[z_i] * P_R_Z_table[t_student][z_i]) / likelihood_table[t_student]
```

In [173]:

```
def update_P_Z(z_i):
    result = 0
    for t in range(rating_df.shape[0]):
        #result += get_estep(t, z_i)
        result += estep_table[t][z_i]
    return (result / rating_df.shape[0])
```

In [174]:

```
def update_P_R_Z(r_j, z_i):
    result = 0
    for t in range(rating_df.shape[0]):
        #rho = get_estep(t, z_i)
        rho = estep_table[t][z_i]
        if rating_df[t][r_j] != '?':
            if rating_df[t][r_j] == '1':
                result += rho
        else:
            #result += rho * get_P_R_Z(t, z_i)
            result += rho * probR_df[r_j][z_i]

    return result / (rating_df.shape[0] * update_P_Z(z_i))
```

In [175]:

```
def get_log_likelihood():
    result = 0
    for t in range(rating_df.shape[0]):
        #result += np.log(get_likelihood(t))
        result += np.log(likelihood_table[t])
    return result / rating_df.shape[0]
```

In [176]:

```
iterations = [0, 1, 2, 4, 8, 16, 32, 64, 128, 256]
for i in range(257):
    for t in range(rating_df.shape[0]):
        for z in range(probZ_df.shape[0]):
            P_R_Z_table[t][z] = get_P_R_Z(t, z)
        likelihood_table[t] = get_likelihood(t)
        for z in range(probZ_df.shape[0]):
            estep_table[t][z] = get_estep(t, z)

    if i in iterations:
        print(i, get_log_likelihood())

    for z in range(probZ_df.shape[0]):
        probZ_df[z] = update_P_Z(z)
    for m in range(probR_df.shape[0]):
        for z in range(probR_df.shape[1]):
            probR_df[m][z] = update_P_R_Z(m, z)
```

```
0 -27.035815003511123
1 -17.5604038243144
2 -16.002362630627825
4 -15.060597317892249
8 -14.501649272825004
16 -14.263788571437592
32 -14.180178075094366
64 -14.170077781591056
128 -14.163960358152188
256 -14.16369243900787
```

# 8.1 f)

In [187]:

```
myrating_index = (id_df == 'A59005342').nonzero()[0][0]
myratings = rating_df[myrating_index]
```

In [194]:

```
for r in range(rating_df.shape[1]):
    if rating_df[myrating_index][r] == '?':
        prob = 0
        for z in range(probZ_df.shape[0]):
            prob = prob + (estep_table[myrating_index][z] * probR_df[r][z])
        print(movie_df[r], prob)
```

```
Black_Swan 0.8727848959223924
Bridemaids 0.8086924623009729
Les_Miserables 0.9164247584086622
Magic_Mike 0.5611254414894956
12_Years_a_Slave 0.874849486233411
Fifty_Shades_of_Grey 0.4905191350999009
I_Feel_Pretty 0.34538457785655813
Chappaquidick 0.9921741329945875
La_La_Land 0.8583660461746279
Hidden_Figures 0.8861577217506736
Phantom_Thread 0.9093209197954123
Darkest_Hour 0.9748884119260098
The_Lion_King 0.9584398213502858
Rocketman 0.6921673559576419
Fast_&_Furious:_Hobbs_&_Shaw 0.7068480787561562
The_Farewell 0.8947347936947422
Hustlers 0.6054760739701981
```

8.1

b) $P(\{R_j = r_j^{(t)}\}_{j\in\Omega_t})$

$$= \sum_{i=1}^{k} P(\{R_j = r_j^{(t)}\}_{j\in\Omega_t}, Z=i) \quad \boxed{\text{marg}}$$

$$= \sum_{i=1}^{k} P(Z=i)\, P(\{R_j = r_j^{(t)}\}_{j\in\Omega_t} \mid Z=i) \quad \boxed{\text{P.R.}}$$

$$= \sum_{i=1}^{k} P(Z=i) \prod_{j\in\Omega_t} P(R_j = r_j^{(t)} \mid Z=i) \quad \boxed{\text{C.I.}}$$

c) $P(Z=i \mid \{R_j = r_j^{(t)}\}_{j\in\Omega_t})$

$$\Rightarrow \frac{P(\{R_j = r_j^{(t)}\}_{j\in\Omega_t} \mid Z=i)\, P(Z=i)}{P(\{R_j = r_j^{(t)}\}_{j\in\Omega_t})} \quad \boxed{\text{Bayes}}$$

$$\Rightarrow \frac{P(Z=i) \prod_{j\in\Omega_t} P(R_j = r_j^{(t)} \mid Z=i)}{\sum_{i'=1}^{k} P(Z=i') \prod_{j\in\Omega_t} P(R_j = r_j^{(t)} \mid Z=i')} \quad \boxed{\text{C.I.}} \quad \boxed{\text{from(b)}}$$

d) $f_{it} = P(Z=i \mid \{R_j = r_j^{(t)}\}_{j \in \Omega_t})$

For root nodes:-

$$P(X_i = x) \leftarrow \frac{1}{T} \sum_{t}^{T} P(X_i = x \mid V_t = v_t)$$

$$\Rightarrow P(Z=i) \leftarrow \frac{1}{T} \sum_{t} f_{it}$$

For nodes with parents

$$P(X_i = x \mid Pa_i = \pi) = \frac{P(X_i = x, Pa_i = \pi \mid V_t = v_t)}{P(Pa_i = \pi \mid V_t = v_t)}$$

$$\Rightarrow P(R_j = 1 \mid Z = i) \leftarrow \frac{\sum_t P(R_j = 1, Z = i \mid \{R_j = r_j^{(t)}\}_{j \in \Omega_t})}{\sum_t P(Z = i \mid \{R_j = r_j^{(t)}\}_{j \in \Omega_t})}$$

$$\boxed{\Rightarrow P(R_j = 1 \mid Z = i) \leftarrow \frac{\sum_{\{t \mid j \in \Omega_t\}} f_{it} I(r_j^{(t)}, 1) + \sum_{\{t \mid j \notin \Omega_t\}} f_{it} P(R_j = 1 \mid Z = i)}{\sum_t f_{it}}}$$

8.2

a) $P(y=i) = \pi_i$

$$P(\vec{x} \mid y=i) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_i|}} \exp\left\{ -\frac{1}{2} (\vec{x}-\vec{\mu}_i)^T \Sigma_i^{-1} (\vec{x}_i-\vec{\mu}_i) \right\}$$

$$P(y=i \mid \vec{x}) = \frac{P(\vec{x} \mid y=i) P(y=i)}{P(\vec{x})} \qquad \boxed{\text{Bayes}}$$

$$= \frac{P(\vec{x} \mid y=i)\,\pi_i}{\sum_j P(\vec{x} \mid y=j) P(y=j)} \qquad \boxed{\text{marg}}$$

$$\Rightarrow P(y=1 \mid \vec{x}) = \frac{P(\vec{x} \mid y=1)\,\pi_1}{\sum P(\vec{x} \mid y=j)\,\pi_j}$$

b) $P(y=1|\vec{x}) = \dfrac{P(\vec{x}|y=1)\,\pi_1}{\sum_j P(\vec{x}|y=j)\,\pi_j}$ from (a)

$$= \frac{P(\vec{x}|y=1)\,\pi_1}{P(\vec{x}|y=0)\,\pi_0 + P(\vec{x}|y=1)\,\pi_1}$$

$$= \frac{1}{1 + \dfrac{P(\vec{x}|y=0)\,\pi_0}{P(\vec{x}|y=1)\,\pi_1}} \quad \ldots (1)$$

$$\Rightarrow \frac{P(\vec{x}|y=0)\,\pi_0}{P(\vec{x}|y=1)\,\pi_1} = \frac{\pi_0 \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu}_0)^T \Sigma^{-1} (\vec{x}-\vec{\mu}_0)\right)}{\pi_1 \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu}_1)^T \Sigma^{-1} (\vec{x}-\vec{\mu}_1)\right)}$$

$$\Rightarrow \frac{\pi_0}{\pi_1} \exp\left\{-\frac{1}{2}\left((\vec{x}-\vec{\mu}_0)^T \Sigma^{-1} (\vec{x}-\vec{\mu}_0) - (\vec{x}-\vec{\mu}_1)^T \Sigma^{-1} (\vec{x}-\vec{\mu}_1)\right)\right\}$$

$$\Rightarrow \frac{\pi_0}{\pi_1} \exp\left\{-\frac{1}{2}\left(\cancel{\vec{x}^T \Sigma^{-1} \vec{x}} - \vec{\mu}_0^T \Sigma^{-1} \vec{x} - \vec{x}^T \Sigma^{-1} \vec{\mu}_0 + \vec{\mu}_0^T \Sigma^{-1} \vec{\mu}_0 \right.\right.$$
$$\left.\left. - \cancel{\vec{x}^T \Sigma^{-1} \vec{x}} + \vec{\mu}_1^T \Sigma^{-1} \vec{x} + \vec{x}^T \Sigma^{-1} \vec{\mu}_1 - \vec{\mu}_1^T \Sigma^{-1} \vec{\mu}_1\right)\right\}$$

Since, $\vec{\mu}^T \Sigma^{-1} \vec{x}$ is symmetric

$$\Rightarrow \vec{\mu}_i^T \Sigma^{-1} \vec{x} = \vec{x}^T \Sigma^{-1} \vec{\mu}_i$$

$$\Rightarrow \frac{\pi_0}{\pi_1} \exp\left\{-\frac{1}{2}\left(2\vec{x}^T \Sigma^{-1} (\vec{\mu}_1 - \vec{\mu}_0) + \vec{\mu}_0^T \Sigma^{-1} \vec{\mu}_0 - \vec{\mu}_1^T \Sigma^{-1} \vec{\mu}_1\right)\right\}$$

$$\Rightarrow \exp\left\{-\left(\vec{x}\Sigma^{-1}(\vec{\mu}_1 - \vec{\mu}_0) + \frac{\vec{\mu}_0^T \Sigma^{-1} \vec{\mu}_0 - \vec{\mu}_1^T \Sigma^{-1} \vec{\mu}_1}{2} - \log\frac{\pi_0}{\pi_1}\right)\right\}$$

$$\Rightarrow \vec{\omega} = \Sigma^{-1}(\vec{\mu}_1 - \vec{\mu}_0)$$

$$\vec{b} = \frac{1}{2}\left(\vec{\mu}_0^T \Sigma^{-1} \vec{\mu}_0 - \vec{\mu}_1^T \Sigma^{-1} \vec{\mu}_1\right) - \log \frac{\pi_0}{\pi_1}$$

c) $$\frac{P(y=1 \mid \vec{x})}{P(y=0 \mid \vec{x})} = \frac{P(\vec{x} \mid y=1)\,\pi_1}{P(\vec{x} \mid y=0)\,\pi_0}$$ $\boxed{\text{Bayes}}$

$$\Rightarrow \frac{\pi_1}{\pi_0} \exp\left\{-\frac{1}{2}\left((\vec{x}-\vec{\mu}_1)^T \Sigma^{-1} (\vec{x}-\vec{\mu}_1) - (\vec{x}-\vec{\mu}_0) \Sigma^{-1} (\vec{x}-\vec{\mu}_0)\right)\right\}$$

$$\Rightarrow \frac{\pi_1}{\pi_0} \exp\left\{-\frac{1}{2}\left(\vec{x}^T \Sigma^{-1} \vec{x} - 2\vec{\mu}_1^T \Sigma^{-1} \vec{x} + \vec{\mu}_1^T \Sigma^{-1} \vec{\mu}_1 - \vec{x}^T \Sigma^{-1} \vec{x} + 2\vec{\mu}_0^T \Sigma^{-1} \vec{x} - \vec{\mu}_0^T \Sigma^{-1} \vec{\mu}_0\right)\right\}$$

$$\Rightarrow \exp\left\{-\left(\vec{x}^T \Sigma^{-1} (\vec{\mu}_0 - \vec{\mu}_1) + \frac{1}{2}\left(\vec{\mu}_1^T \Sigma^{-1} \vec{\mu}_1 - \vec{\mu}_0 \Sigma^{-1} \vec{\mu}_0\right) - \log \frac{\pi_1}{\pi_0}\right)\right\}$$

$$= K$$

$$\Rightarrow \vec{x}^T \Sigma^{-1} (\vec{\mu}_0 - \vec{\mu}_1) + \frac{1}{2}\left(\vec{\mu}_1 \Sigma^{-1} \vec{\mu}_1 - \vec{\mu}_0^T \Sigma^{-1} \vec{\mu}_0\right) - \log \frac{\pi_1}{\pi_0}$$

$$= \log K$$

$$\Rightarrow \vec{\omega} = \Sigma^{-1}(\vec{\mu}_1 - \vec{\mu}_0)$$

$$\vec{b} = \frac{1}{2}\left(\vec{\mu}_0^{T}\Sigma^{-1}\vec{\mu}_0 - \vec{\mu}_1^{T}\Sigma^{-1}\vec{\mu}_1\right) - \log\frac{\pi_1}{\pi_0}$$

$$\Rightarrow \vec{\omega}\cdot\vec{x} + \vec{b} = \log K$$

$\Rightarrow$ when $K = 1$

$$\Rightarrow \vec{\omega}\cdot\vec{x} + \vec{b} = 0$$

8.3

a) $P(Y=1 \mid \vec{X}=\vec{x}) = 1 - \exp(-\vec{v}.\vec{x})$

$$\mathcal{L}(\vec{v}) = \sum_{t=1}^{T} \log P(y_t \mid \vec{x}_t)$$

$$= \sum_{t=1}^{T} \log (1-\exp(-\vec{v}\,\vec{x}))^{y_t} (\exp(-\vec{v}.\vec{x}))^{1-y_t}$$

$$\boxed{\mathcal{L}(\vec{v}) = \sum_{t=1}^{T} y_t \log (1-e^{-\vec{v}.\vec{x}}) + (1-y_t) \log e^{-\vec{v}.\vec{x}}}$$

Further simplifying;

$$= \sum_{t=1}^{T} y_t \log (e^{\vec{v}.\vec{x}}-1) - y_t \log e^{\vec{v}.\vec{x}}$$

$$+ (1-y_t)(-\log e^{\vec{v}.\vec{x}})$$

$$= \sum_{t=1}^{T} y_t \log (e^{\vec{v}.\vec{x}}-1) - \vec{v}.\vec{x}$$

b) $\mathcal{L}(\vec{V}) = \sum_{t=1}^{T} y_t \log(e^{\vec{V}\cdot\vec{x}} - 1) - \vec{V}\cdot\vec{x}$ (from (a))

$$\frac{\partial \mathcal{L}(\vec{V})}{\partial \vec{V}} = \sum_{t=1}^{T} y_t \times \frac{1}{e^{\vec{V}\cdot\vec{x}} - 1} \times e^{\vec{V}\cdot\vec{x}} \times \vec{x} - \vec{x}$$

$$= \sum_{t=1}^{T} \left( y_t \times \frac{1}{1 - e^{-\vec{V}\cdot\vec{x}}} - 1 \right) \vec{x}$$

$$= \sum_{t=1}^{T} \left( \frac{y_t - p_t}{p_t} \right) \vec{x}$$ option (ii)

c) $V_i = -\log(1 - p_i)$

$$-V_i x_i = \log(1 - p_i) x_i$$

$$\Rightarrow \quad e^{-\vec{V}_i \vec{x}_i} = e^{-V_0 x_0 - V_1 x_1 \dots}$$

$$= e^{\sum_i \log(1-p_i) x_i}$$

$$= \prod_i \left( e^{\log(1-p_i)} \right)^{x_i}$$

$$= \prod_i (1 - p_i)^{x_i}$$

$$\Rightarrow \boxed{P(y=1 \mid \vec{x}) = 1 - e^{-\vec{V}_i \cdot \vec{x}} = 1 - \prod_i (1 - p_i)^{x_i}}$$

d) $\mathcal{L}(\vec{V}) = \sum_{t=1}^{T} y_t \log(e^{\vec{V}\cdot\vec{x}_t} - 1) - \vec{V}\cdot\vec{x}_t$

$$\mathcal{L} = \sum_{t=1}^{T} P(y_t \mid \vec{x}_t)$$

$$\mathcal{L}(P_i) = \sum_{t=1}^{T} \log\left(1 - \prod_i (1-P_i)^{x_{it}}\right)^{y_t}\left(\prod_i (1-P_i)^{x_{it}}\right)^{1-y_t}$$

$$\cancel{= \sum_{t=1}^{T} \log}$$

$$= \sum_{t=1} y_t \log\left(1 - \prod_i (1-P_i)^{x_{it}}\right) \cdot + (1-y_t)\log\left(\prod_i (1-P_i)^{x_{it}}\right)$$

$$= \sum_{t=1} y_t \log\left(1 - \prod_i (1-P_i)^{x_{it}}\right) + (1-y_t)\sum_i x_{it} \log(1-P_i)$$

$$\frac{\partial \mathcal{L}}{\partial P_i} = \sum_{t=1}^{T} \frac{y_t\left(\prod_{j\setminus\{i\}} (1-P_j)^{x_{jt}}\right)(1-P_i)^{x_{it}-1} x_{it}}{\left(1 - \prod_i (1-P_i)^{x_{it}}\right)} + (1-y_t)\, x_{it} \times \left(\frac{-1}{1-P_i}\right)$$

$$\Rightarrow \sum_{t=1}^{T} \frac{1}{(1-P_i)} \left( \frac{y_t \prod_j (1-P_j)^{x_{jt}}}{\left(1-\prod_j (1-P_j)^{x_{jt}}\right)} - (1-y_t) \right) x_{it}$$

$$\Rightarrow \sum_{t=1} \frac{1}{(1-P_i)} \left( \frac{y_t \prod_j (1-P_j)^{x_{jt}} - (1-y_t)\left(1-\prod_j (1-P_j)^{x_{jt}}\right)}{\left(1-\prod_j (1-P_j)^{x_{jt}}\right)} \right) x_{it}$$

$$\Rightarrow \sum_{t=1} \frac{1}{(1-P_i)} \left( \frac{y_t - \left(1-\prod_j (1-P_j)^{x_{jt}}\right)}{\left(1-\prod_j (1-P_j)^{x_{jt}}\right)} \right) x_{it}$$

$$\Rightarrow \boxed{\frac{\partial \mathcal{L}}{\partial P_i} = \sum_{t=1}^{T} \frac{1}{(1-P_i)} \left( \frac{y_t - f_t}{f_t} \right) x_{it}} \quad \ldots (i)$$

Now,

$$\mathcal{L}(\vec{v}) = \sum_{t=1}^{T} y_t \log\left(e^{\vec{v}\cdot\vec{x}_t} - 1\right) - \vec{v}\cdot\vec{x}_t$$

$$\frac{\partial \mathcal{L}}{\partial v_i} = \sum_{t=1}^{T} y_t \times \frac{1}{\left(e^{\vec{v}\cdot\vec{x}} - 1\right)} e^{v_i x_i} x_{it} - x_{it}$$

$$\boxed{\frac{\partial \mathcal{L}}{\partial v_i} = \sum_{t=1}^{T} \left( \frac{y_t - f_t}{f_t} \right) x_{it}} \quad \ldots (ii)$$

$\Rightarrow$ From (i) and (ii)

$$\boxed{\frac{\partial L}{\partial P_i} = \frac{1}{(1-P_i)} \frac{\partial L}{\partial V_i}}$$

e) Gradient Ascent :-

$$P_i \leftarrow P_i + \eta \frac{\partial L}{\partial P_i} \qquad (i)$$

EM Update :-

$$P_i \leftarrow \frac{P_i}{T_i} \left[ \sum_{t=1}^{T} \left( \frac{y_t x_{it}}{\rho_t} \right) \right] \qquad (ii)$$

where,

$$\rho_t = P(y_t = 1 \mid \vec{x}_t)$$

From (i) and (d)

$$P_i \leftarrow P_i + \eta \sum_{t=1}^{T} \frac{1}{(1-P_i)} \left( \frac{y_t}{\rho_t} - 1 \right) x_{it}$$

$\Rightarrow$ given $\eta = \dfrac{p_i(1-p_i)}{T_i}$

$$\Rightarrow p_i \leftarrow p_i + \frac{p_i \cancel{(1-p_i)}}{T_i} \sum_{t=1}^{T} \frac{1}{\cancel{(1-p_i)}} \left( \frac{y_t}{f_t} - 1 \right) x_{it}$$

$$p_i \leftarrow \frac{p_i}{T_i} \left( T_i + \sum_{t=1}^{T} \left( \frac{y_t}{f_t} - 1 \right) x_{it} \right)$$

$$\because \; T_i = \sum_{t=1}^{T} x_{it}$$

$$\Rightarrow p_i \leftarrow \frac{p_i}{T_i} \left( \sum_{t=1}^{T} x_{it} + \left( \frac{y_t}{f_t} - 1 \right) x_{it} \right)$$

$$\boxed{p_i \leftarrow \frac{p_i}{T_i} \left( \sum_{t=1}^{T} \frac{x_{it}\, y_t}{f_t} \right)}$$

8.4

a) $P(y=1 \mid \vec{x}) = \sigma(\vec{\omega} \cdot \vec{x})$

$P(y'=1 \mid \vec{x}') = \sigma(\vec{\omega} \cdot \vec{x}')$

$P(S=1 \mid y, y') = I(y, y')$

$$P(y, y' \mid \vec{x}, \vec{x}'; S) = \frac{P(S \mid y, y', \vec{x}, \vec{x}')\, P(y, y' \mid \vec{x}, \vec{x}')}{P(S \mid \vec{x}, \vec{x}')} \quad \boxed{\text{Bayes}}$$

$$\Rightarrow \frac{P(S \mid y, y')\, P(y \mid \vec{x}, \vec{x}')\, P(y' \mid \vec{x}, \vec{x}')}{\sum_i \sum_j P(y=i, y'=j, S \mid \vec{x}, \vec{x}')} \quad \boxed{\text{C.I}} \& \boxed{\text{P.R}} \quad \boxed{\text{marg}}$$

$$\Rightarrow \frac{P(S \mid y, y')\, P(y \mid \vec{x})\, P(y' \mid \vec{x}')}{\sum_i \sum_j P(y=i, y'=j, S \mid \vec{x}, \vec{x}')} \quad \boxed{\text{C.I}}$$

$\because$ we need to calc. $P(y=1, y'=1 \mid \vec{x}, \vec{x}', S=1)$

$$\Rightarrow \frac{P(S=1 \mid y=1, y'=1)\, P(y=1 \mid \vec{x})\, P(y'=1 \mid \vec{x}')}{\sum_i \sum_j P(y=i, y'=j, s=1 \mid \vec{x}, \vec{x}')}$$

$$\Rightarrow \frac{\sigma(\vec{\omega}.\vec{x})\,\sigma(\vec{\omega}.\vec{x}')}{\sum_i \sum_j P(S=1 \mid y=i, y'=j)\, P(y=i \mid \vec{x})\, P(P(y'=j \mid \vec{x}')}$$

similar to numerator

$$\Rightarrow \frac{\sigma(\vec{\omega}.\vec{x})\,\sigma(\vec{\omega}.\vec{x}')}{P(y=0 \mid \vec{x}) * P(y'=0 \mid \vec{x}') + P(y=1 \mid \vec{x}) * P(y'=1 \mid \vec{x}')}$$

$$\Rightarrow \boxed{\frac{\sigma(\vec{\omega}.\vec{x})\,\sigma(\vec{\omega}.\vec{x}')}{\sigma(-\vec{\omega}.\vec{x})\,\sigma(-\vec{\omega}.\vec{x}') + \sigma(\vec{\omega}.\vec{x})\,\sigma(\vec{\omega}.\vec{x}')}}$$

b) $P(y=1, y'=0 \mid \vec{x}, \vec{x}', s=0)$

Similar to (a)

$$\Rightarrow \frac{P(S=0 \mid y=1, y'=0)\, P(y=1 \mid \vec{x}, \vec{x}')\, P(y=0 \mid \vec{x}, \vec{x}')}{P(y=1 \mid \vec{x})\, P(y'=0 \mid \vec{x}') + P(y=0 \mid \vec{x})\, P(y'=1 \mid \vec{x}')}$$

$$\Rightarrow \boxed{\frac{\sigma(\vec{\omega}.\vec{x})\,\sigma(-\vec{\omega}.\vec{x}')}{\sigma(\vec{\omega}.\vec{x})\,\sigma(-\vec{\omega}.\vec{x}') + \sigma(-\vec{\omega}.\vec{x})\,\sigma(\vec{\omega}.\vec{x}')}}$$

c) (i) • $P(y=1 \mid \vec{x}, \vec{x}', S=1)$

$\Rightarrow \sum_i P(y=1, y'=i \mid \vec{x}, \vec{x}', S=1)$ $\boxed{\text{marg}}$

$\Rightarrow \boxed{P(y=1, y'=1 \mid \vec{x}, \vec{x}', S=1) \quad (b)}$

$\because P(y=1, y'=0 \mid \vec{x}, \vec{x}', S=1)$

$= P(S=1 \mid y=1, y'=0)\, P(y=1, y'=0 \mid \vec{x}, \vec{x}')$

$= 0$

(ii) • $P(y'=1 \mid \vec{x}, \vec{x}', S=1)$

$\Rightarrow \sum_i P(y=i, y'=1 \mid \vec{x}, \vec{x}', S=1)$ $\boxed{\text{marg}}$

$\Rightarrow \boxed{P(y=1, y'=1 \mid \vec{x}, \vec{x}', S=1) \quad (b)}$

for similar reason as (i).

$\Rightarrow$

(iii) $P(y=1 \mid \vec{x}, \vec{x}', s=0)$

$$\Rightarrow \sum_i P(y=1, y'=i \mid \vec{x}, \vec{x}', s=0) \quad \boxed{\text{marg}}$$

$$\Rightarrow \boxed{P(y=1, y'=0 \mid \vec{x}, \vec{x}', s=0) \quad (a)}$$

$$\left[ \because P(y=1, y'=1 \mid \vec{x}, \vec{x}', s=0) = 0 \atop \because P(s=0 \mid y=1, y'=1) = 0 \right]$$

(iv) $P(y'=1 \mid \vec{x}, \vec{x}', s=0)$

$$\Rightarrow \sum_i P(y=i, y'=1 \mid \vec{x}, \vec{x}', s=0) \quad \boxed{\text{marg}}$$

$$\Rightarrow P(y=0, y'=1 \mid \vec{x}, \vec{x}', s=0)$$

$$\left[ \because P(s=0 \mid y=1, y'=1) = 0 \right]$$

$$\Rightarrow \frac{P(s=0 \mid y=0, y'=1) P(y=0 \mid \vec{x}) P(y'=1 \mid \vec{x}')}{P(s=0 \mid \vec{x}, \vec{x}')}$$

$\boxed{\text{bayes} \rightarrow \text{P.R.} \rightarrow \text{C.I}}$

$$\Rightarrow \frac{(1 - \sigma(\vec{w} \cdot \vec{x}))\, \sigma(\vec{w} \cdot \vec{x}')}{P(s=0 \mid \vec{x}, \vec{x}')}$$

$$\Rightarrow \boxed{1 - P(y=1, y'=0 \mid \vec{x}, \vec{x}', s=0) \quad (c)}$$

d) $\mathcal{L}(\vec{\omega}) = \sum_t \log P(S_t \mid \vec{x}_t, \vec{x}')$

$\Rightarrow \quad = \sum_t \log P(S_t = 1 \mid \vec{x}_t, \vec{x}_t')^{S_t} P(S=0 \mid \vec{x}_t, \vec{x}_t')^{(1-S_t)}$

$\Rightarrow \quad = \sum_t S_t \log P(S_t = 1 \mid \vec{x}_t, \vec{x}_t') +$

$(1-S_t) \cdot \log P(S_t = 0 \mid \vec{x}_t, \vec{x}_t')$

$\Rightarrow \quad = \sum_t S_t \log \sum_i \sum_j P(S_t = 1, y=i, y'=j \mid \vec{x}_t, \vec{x}_t')$

$+ (1-S_t) \log \cancel{P(S_t} \sum_i \sum_j P(S_t = 0, y=i, y'=j \mid \vec{x}_t, \vec{x}_t')$

$\Rightarrow \quad = \sum_t S_t \log \sum_i \sum_j P(S_t = 1 \mid y=i, y'=j) P(y=i \mid \vec{x}) P(y'=j \mid \vec{x}')$

$+ (1-S_t) \log \sum_i \sum_j P(S_t = 0 \mid y=i, y'=j) P(y=i \mid \vec{x}) P(y'=j \mid \vec{x}')$

$\boxed{\text{marg} \rightarrow \text{P.R.} \rightarrow \text{C.I.}}$

$\Rightarrow \sum_{t=1}^{T} S_t \log \begin{bmatrix} P(S_t = 1 \mid y=0, y'=0) P(y=0 \mid \vec{x}) P(y'=0 \mid \vec{x}') \\ + P(S_t = 1 \mid y=1, y'=1) P(y=1 \mid \vec{x}) P(y'=1 \mid \vec{x}') \end{bmatrix}$

$+ (1-S_t) \log \begin{bmatrix} P(S_t = 0 \mid y=0, y'=1) P(y=0 \mid \vec{x}) P(y'=1 \mid \vec{x}') \\ + P(S_t = 0 \mid y=1, y'=0) P(y=1 \mid \vec{x}) P(y'=0 \mid \vec{x}') \end{bmatrix}$

$$\Rightarrow \boxed{\begin{aligned} &\sum_{t=1}^{T} S_t \log [\sigma(-\vec{\omega} \cdot \vec{x}) \sigma(-\vec{\omega} \cdot \vec{x}') + \sigma(\vec{\omega} \cdot \vec{x}) \sigma(\vec{\omega} \cdot \vec{x}')] \\ &+ (1-S_t) \log [\sigma(-\vec{\omega} \cdot \vec{x}) \sigma(\vec{\omega} \cdot \vec{x}') + \sigma(\vec{\omega} \cdot \vec{x}) \sigma(-\vec{\omega} \cdot \vec{x}')] \end{aligned}}$$

e) $\vec{\omega} \leftarrow \underset{\vec{\omega}}{\operatorname{argmax}} \left\{ \sum_{t} \left[ \bar{y}_t \log \sigma(\vec{\omega} \cdot \vec{x}_t) + \right. \right.$

$(1-\bar{y}_t) \log(\sigma(-\vec{\omega} \cdot \vec{x})) +$

$\bar{y}'_t \log \sigma(\vec{\omega} \cdot \vec{x}'_t) +$

$\left. \left. (1-\bar{y}'_t) \log \sigma(-\vec{\omega} \cdot \vec{x}'_t) \right] \right\}$

$$f(\vec{\omega}) = \bar{y}_t \log \sigma(\vec{\omega} \cdot \vec{x}_t) + (1-\bar{y}_t) \log \sigma(-\vec{\omega} \cdot \vec{x})$$
$$+ \bar{y}'_t \log \sigma(\vec{\omega} \cdot \vec{x}'_t) + (1-\bar{y}'_t) \log \sigma(-\vec{\omega} \cdot \vec{x}'_t)$$

$$\Rightarrow \frac{\partial f(\vec{\omega})}{\partial \vec{\omega}} = \bar{y}_t \frac{\sigma'(\vec{\omega} \cdot \vec{x}_t)}{\sigma(\vec{\omega} \cdot \vec{x}_t)} \vec{x}_t + \frac{(1-\bar{y}_t) \sigma'(-\vec{\omega} \cdot \vec{x}_t)(-\vec{x})}{\sigma(-\vec{\omega} \cdot \vec{x}_t)}$$
$$+ \vec{y}'_t \frac{\sigma'(\vec{\omega} \cdot \vec{x}'_t)}{\sigma(\vec{\omega} \cdot \vec{x}'_t)} \vec{x}'_t - (1-y'_t) \frac{\sigma'(-\vec{\omega} \cdot \vec{x}_t) \vec{x}'}{\sigma(-\vec{\omega} \cdot \vec{x}')}$$

$$\left[ \sigma'(z) = \sigma(z) \sigma(-z) \right]$$

$$\Rightarrow \left[ \bar{y}_t \sigma(-\vec{\omega} \cdot \vec{x}_t) - (1-\bar{y}_t) \sigma(\vec{\omega} \cdot \vec{x}_t) \right] \vec{x}_t$$
$$+ \left[ \bar{y}'_t \sigma(-\vec{\omega} \cdot \vec{x}_t) - (1-y'_t) \sigma(\vec{\omega} \cdot \vec{x}'_t) \right] \vec{x}'_t$$

$$\Rightarrow \sum \left[ (\bar{y}_t - \sigma(\vec{\omega} \cdot \vec{x}_t)) \vec{x}_t + (\bar{y}'_t - \sigma(\vec{\omega} \cdot \vec{x}'_t)) x'_t \right]$$

This is the direction in which we need to ~~appo~~ move the $\vec{\omega}$

$$\Rightarrow \vec{\omega} \leftarrow \vec{\omega} + \eta \sum_t \Big[ y_t \vec{x}_t - \sigma(\vec{\omega}\cdot\vec{x}_t)\vec{x}_t + y'_t \vec{x}'_t - \sigma(\vec{\omega}\cdot\vec{x}'_t)\vec{x}'_t \Big]$$

8.5

a) $\alpha_{it} = P(Y_t = i \mid y_0, \vec{x}_1 \ldots \vec{x}_t)$

$P(Y_t = 1 \mid \vec{x}_t, Y_{t-1} = 0) = \sigma(\vec{\omega}_0 \cdot \vec{x}_t)$

$P(Y_t = 1 \mid \vec{x}_t, Y_{t-1} = 1) = \sigma(\vec{\omega}_1 \cdot \vec{x}_t)$

$\Rightarrow \alpha_{j\,t+1} = P(Y_{t+1} = j \mid y_0, \vec{x}_1 \ldots \vec{x}_{t+1})$

$\Rightarrow \sum_i P(Y_{t+1} = j, Y_t = i \mid y_0, \vec{x}_1 \ldots \vec{x}_{t+1})$ $\boxed{\text{marg}}$

$\Rightarrow \sum_i P(Y_{t+1} = j \mid Y_t = i, \vec{x}_{t+1}) P(Y_t = i \mid y_0, \vec{x}_1 \ldots \vec{x}_{t+1})$

$\boxed{\text{P.R. \& C.I.}}$

$$\Rightarrow \sum_i P(Y_{t+1}=j \mid Y_t=i, \vec{x}_{t+1})\, P(y_t=i \mid y_0, \vec{x}_1 \ldots \vec{x}_t)$$

$\boxed{C.I.}$

$$\Rightarrow \boxed{\sum_i \sigma(\vec{\omega}_i \cdot \vec{x}_{t+1})^{j}\, \sigma(-\vec{\omega}_i \cdot \vec{x}_{t+1})^{(1-j)}\, \alpha_{it}}$$

b) $$l^*_{it} = \max_{y_1 \cdots y_{t-1}} \left[\log P(y_1 \ldots y_t=i \mid y_0, \vec{x}_1 \ldots \vec{x}_t)\right]$$

$$l^*_{j\,t+1} = \max_{y_1 \cdots y_t} \left[\log P(y_1 \ldots y_{t+1}=j \mid y_0, \vec{x}_1 \ldots \vec{x}_{t+1})\right]$$

$$\Rightarrow \max_i \max_{y_1 \cdots y_{t-1}} \left[\log P(y_1 \cdots y_t=i, y_{t+1}=j \mid y_0, \vec{x}_1 \ldots \vec{x}_{t+1})\right] \quad \ldots (i)$$

Now,

$$P(y_1 \ldots y_t=i, y_{t+1}=j \mid y_0, \vec{x}_1 \ldots \vec{x}_{t+1}) =$$

$$\Rightarrow \frac{P(y_1 \ldots y_t=i, y_{t+1}=j, \vec{x}_{t+1} \mid y_0, \vec{x}_1 \ldots \vec{x}_t)}{P(\vec{x}_{t+1} \mid y_0, \vec{x}_1 \ldots \vec{x}_t)} \quad \boxed{P.R.}$$

$$\Rightarrow \frac{P(y_1 \ldots y_t=i \mid y_0, \vec{x}_1 \ldots \vec{x}_t)\, P(y_{t+1}=j, \vec{x}_{t+1} \mid y_0, \vec{x}_1 \ldots \vec{x}_t, y_t=i)}{P(\vec{x}_{t+1} \mid y_0, \vec{x}_1 \ldots \vec{x}_t)}$$

$\boxed{P.R.}$

$\ldots (ii)$

Furthermore;

$$P(y_{t+1}=j, \vec{x}_{t+1} \mid y_0, \vec{x}_1 \ldots \vec{x}_t, y_t)$$

$$= P(y_{t+1}=j \mid \vec{x}_{t+1}, y_t=i)\, P(\vec{x}_{t+1} \mid y_0, \vec{x}_1 \ldots \vec{x}_t)$$

$\boxed{P.R. \,\&\, C.I.}$ (iii)

Substituting in (ii)

$$\Rightarrow P(y_1 \ldots y_t=i \mid y_0, \vec{x}_1 \ldots \vec{x}_t)\, P(y_{t+1}=j \mid \vec{x}_{t+1}, y_t=i)$$

$\Rightarrow$ Substituting in (i)

$$\Rightarrow \max_i \max_{y_1 \ldots y_{t-1}} \Big[\log P(y_1 \ldots y_t=i \mid y_0, \vec{x}_1 \ldots \vec{x}_t) + \log P(y_{t+1}=j \mid \vec{x}_{t+1}, y_t=i)\Big]$$

$$\Rightarrow \max_i \Big[ l^*_{i_t} + \log P(y_{t+1}=j \mid \vec{x}_{t+1}, y_t=i) \Big] \quad \ldots (iv)$$

$\Rightarrow$ We can simplify :-

$$\log P(y_{t+1}=j \mid \vec{x}_{t+1}, y_t=i) =$$

$$\log P(y_{t+1}=0 \mid \vec{x}_{t+1}, y_t=i)^{(1-j)}$$

$$+ \log P(y_{t+1}=1 \mid \vec{x}_{t+1}, y_t=i)^{j}$$

$$\Rightarrow (1-j) \log \sigma(-\vec{\omega}_i \cdot \vec{x}_{t+1}) + j \log \sigma(\vec{\omega}_i \cdot \vec{x}_{t+1}) \quad \ldots (v)$$

Substituting in (iv)

$$\boxed{\Rightarrow l^*_{j\,t+1} = \max_i \left[ l^*_{it} + j \log \sigma(\vec{\omega}_i \cdot \vec{x}_{t+1}) + (1-j) \log \sigma(-\vec{\omega}_i \cdot \vec{x}_{t+1}) \right]}$$

c) $$\phi_{t+1}(j) = \arg\max_i \left[ l^*_{it} + j \log \sigma(\vec{\omega}_i \cdot \vec{x}_{t+1}) + (1-j) \log \sigma(-\vec{\omega}_i \cdot x_{t+1}) \right]$$

$$y^*_t = \phi_{t+1}(y^*_{t+1})$$