8.1 EM algorithm for binary matrix completion

(a)

$$\frac{number\ of\ student\ who\ recommended\ the\ movie}{number\ of\ student\ who\ saw\ the\ movie}$$

I cannot compare these rankings with my individual preferences, because I did not watch most of them

0.9657534246575342 Inception

0.9444444444444444 Shutter_Island

0.9416342412451362 The_Dark_Knight_Rises

0.9393939393939394 The_Martian

0.9385665529010239 Interstellar

0.9195402298850575 The_Theory_of_Everything

0.9178082191780822

Harry_Potter_and_the_Deathly_Hallows:_Part_2

0.9016393442622951 The_Social_Network

0.9006622516556292

Harry_Potter_and_the_Deathly_Hallows:_Part_1

0.8944099378881988 Gone_Girl

0.8943661971830986 The_Lion_King

0.8916666666666667 Wolf_of_Wall_Street

0.8785714285714286 Avengers:_Infinity_War

0.8681318681318682 Avengers:_Endgame

0.8649789029535865 Now_You_See_Me

0.8639455782312925 Django_Unchained

0.8604651162790697 The_Farewell

0.8436482084690554 The_Avengers

0.8363636363636363 Parasite

0.8297872340425532 Black_Swan

0.8271604938271605 Spiderman:_Far_From_Home

0.8253968253968254 21_Jump_Street

0.825 Les_Miserables

0.824 Joker

0.8198198198198198 The_Perks_of_Being_a_Wallflower

0.8178694158075601 Iron_Man_2

0.8163265306122449 12_Years_a_Slave

0.8148148148148148 The_Hunger_Games

0.803347280334728 Toy_Story_3

0.8031496062992126 Thor

0.8 The_Hateful_Eight

0.7974683544303798 Hidden_Figures

0.7973856209150327 Ready_Player_One

0.7931034482758621 The_Girls_with_the_Dragon_Tattoo

0.7913669064748201 Captain_America:_The_First_Avenger

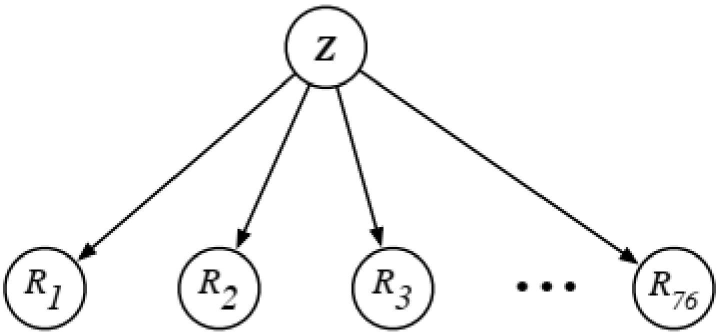
0.7888888888888889 Her

0.7868852459016393 Dunkirk

0.7864077669902912	The_Great_Gatsby
0.7848101265822784	Darkest_Hour
0.7741935483870968	Three_Billboards_Outside_Ebbing
0.768595041322314	Once_Upon_a_Time_in_Hollywood
0.7636363636363637	Manchester_by_the_Sea
0.7560975609756098	Midnight_in_Paris
0.7547169811320755	La_La_Land
0.7424242424242424	Avengers:_Age_of_Ultron
0.7391304347826086	Ex_Machina
0.7363636363636363	The_Revenant
0.7360406091370558	X-Men:_First_Class
0.7241379310344828	Frozen
0.7209302325581395	Jurassic_World
0.72	Pitch_Perfect
0.7068965517241379	The_Help
0.7	Us
0.6901408450704225	Drive
0.6783216783216783	Mad_Max:_Fury_Road
0.6588235294117647	Fast_Five
0.6566265060240963	Pokemon_Detective_Pikachu
0.65625	Good_Boys
0.6515151515151515	Room

0.6388888888888888	Terminator:_Dark_Fate
0.6282051282051282	American_Hustle
0.6242774566473989	Man_of_Steel
0.6190476190476191	Bridemaids
0.6	Chappaquidick
0.5967741935483871	Rocketman
0.5935828877005348	Star_Wars:_The_Force_Awakens
0.5895522388059702	World_War_Z
0.5862068965517241	Phantom_Thread
0.584070796460177	Prometheus
0.5584415584415584	The_Shape_of_Water
0.5405405405405406	Fast_&_Furious:_Hobbs_&_Shaw
0.5151515151515151	Magic_Mike
0.47368421052631576	The_Last_Airbender
0.45652173913043476	Hustlers
0.37714285714285717	Fifty_Shades_of_Grey
0.358974358974359	I_Feel_Pretty

(b)



$$left\ side = P\left(\left\{R_j = r_j^{(t)}\right\}_{j \in \Omega_t}\right) = \sum_i P\left(\left\{R_j = r_j^{(t)}\right\}_{j \in \Omega_t}, Z = i\right)$$

………………………………………………………………………………marginalization

$$= \sum_i^k P(Z = i) P(\left\{R_j = r_j^{(t)}\right\}_{j \in \Omega_t} | Z = i)$$

………………………………………………………………………………product rule

$$= \sum_i^k P(Z = i) \prod_{j \in \Omega_t} P(R_j = r_j^{(t)} | Z = i) = right\ side$$

………………………………………………………………………………d-separation #2

(c)

$$left\ side = P\left(Z = i \middle| \left\{R_j = r_j^{(t)}\right\}_{j \in \Omega_t}\right)$$

$$= \frac{P\left(\left\{R_j = r_j^{(t)}\right\}_{j \in \Omega_t} | z = i\right) P(z = i)}{P\left(\left\{R_j = r_j^{(t)}\right\}_{j \in \Omega_t}\right)}$$

……………………………………………………………………………………bayes rule

Where

$$P\left(\left\{R_j = r_j^{(t)}\right\}_{j\in\Omega_t} | z = i\right) = \prod_{j\in\Omega_t} P(R_j = r_j^{(t)} | Z = i)$$

……………………………………………………………………………………d-separation #2

And from(b), we already know that

$$P\left(\left\{R_j = r_j^{(t)}\right\}_{j\in\Omega_t}\right) = \sum_{i}^{k} P(Z = i) \prod_{j\in\Omega_t} P(R_j = r_j^{(t)} | Z = i)$$

So we can get

$$left\ side = P\left(Z = i \middle| \left\{R_j = r_j^{(t)}\right\}_{j\in\Omega_t}\right)$$

$$= \frac{P\left(\left\{R_j = r_j^{(t)}\right\}_{j\in\Omega_t} | z = i\right) P(z = i)}{P\left(\left\{R_j = r_j^{(t)}\right\}_{j\in\Omega_t}\right)}$$

$$= \frac{P(z = i) \prod_{j\in\Omega_t} P\left(R_j = r_j^{(t)} \middle| Z = i\right)}{\sum_{i'=1}^{k} P(Z = i') \prod_{j\in\Omega_t} P\left(R_j = r_j^{(t)} \middle| Z = i'\right)} = right\ side$$

(d)

According to the general EM algorithm update formula, we have

For the root node

$$P(X = x) \leftarrow \frac{\sum_t P(X = x|V)}{T}$$

For the son node

$$P(X = x|pa = \pi) \leftarrow \frac{\sum_t P(X = x, pa = \pi|V)}{\sum_t P(pa = \pi|V)}$$

In this case

$$V = \left\{R_j = r_j^{(t)}\right\}_{j \in \Omega_t}$$

$$\rho_{it} = P\left(Z = i \middle| \left\{R_j = r_j^{(t)}\right\}_{j \in \Omega_t}\right)$$

so we can get

$$P(Z = i) \leftarrow \frac{1}{T}\sum_{t=1}^{T} P\left(Z = i \middle| \left\{R_j = r_j^{(t)}\right\}_{j \in \Omega_t}\right) = \frac{1}{T}\sum_{t=1}^{T} \rho_{it}$$

$$P(R_j = 1|Z = i) \leftarrow \frac{\sum_t P\left(R_j = 1, Z = i \middle| \left\{R_j = r_j^{(t)}\right\}_{j \in \Omega_t}\right)}{\sum_t P\left(Z = i \middle| \left\{R_j = r_j^{(t)}\right\}_{j \in \Omega_t}\right)}$$

$$= \frac{\sum_t P\left(R_j = 1, Z = i \middle| \left\{R_j = r_j^{(t)}\right\}_{j \in \Omega_t}\right)}{\sum_{t=1}^{T} \rho_{it}}$$

Where

$$P\left(R_j = 1, Z = i \middle| \left\{R_j = r_j^{(t)}\right\}_{j \in \Omega_t}\right)$$

$$= P\left(Z = i \middle| \left\{R_j = r_j^{(t)}\right\}_{j \in \Omega_t}\right) P\left(R_j = 1 \middle| Z = i, \left\{R_j = r_j^{(t)}\right\}_{j \in \Omega_t}\right)$$

………………………………………………………………………………………product rule

$$= \rho_{it} P\left(R_j = 1 \middle| Z = i, \left\{R_j = r_j^{(t)}\right\}_{j \in \Omega_t}\right)$$

So we can get

$$\sum_t P\left(R_j = 1, Z = i \middle| \left\{R_j = r_j^{(t)}\right\}_{j \in \Omega_t}\right)$$

$$= \sum_t \rho_{it} P\left(R_j = 1 \middle| Z = i, \left\{R_j = r_j^{(t)}\right\}_{j \in \Omega_t}\right)$$

$$= \sum_{\{t | j \in \Omega_t\}} \rho_{it} P\left(R_j = 1 \middle| Z = i, \left\{R_j = r_j^{(t)}\right\}_{j \in \Omega_t}\right)$$

$$+ \sum_{\{t | j \notin \Omega_t\}} \rho_{it} P\left(R_j = 1 \middle| Z = i, \left\{R_j = r_j^{(t)}\right\}_{j \in \Omega_t}\right)$$

$$\sum_t P\left(R_j = 1, Z = i \middle| \left\{R_j = r_j^{(t)}\right\}_{j \in \Omega_t}\right)$$

$$= \sum_{\{t | j \in \Omega_t\}} \rho_{it} I(r_j^{(t)}, 1) + \sum_{\{t | j \notin \Omega_t\}} \rho_{it} P\left(R_j = 1 \middle| Z = i\right)$$

………………………………………………………………………………………d-separation #2

In all, we can get

$$P\left(R_j = 1 | Z = i\right) \leftarrow \frac{\sum_t P\left(R_j = 1, Z = i \middle| \left\{R_j = r_j^{(t)}\right\}_{j \in \Omega_t}\right)}{\sum_{t=1}^{T} \rho_{it}}$$

$$= \frac{\sum_{\{t | j \in \Omega_t\}} \rho_{it} I(r_j^{(t)}, 1) + \sum_{\{t | j \notin \Omega_t\}} \rho_{it} P\left(R_j = 1 \middle| Z = i\right)}{\sum_{t=1}^{T} \rho_{it}}$$

Note that since in this update rule $P\left(R_j = 1 | Z = i\right)$ only involve $P\left(R_j = 1 \middle| Z = i\right)$ in the right side, no other term(e.g. $P\left(R_{j'} = 1 \middle| Z = i'\right)$) we do not need to create another array

(e)

From(b), we already know that

$$P\left(\left\{R_j = r_j^{(t)}\right\}_{j\in\Omega_t}\right) = \sum_i^k P(Z=i)\prod_{j\in\Omega_t} P(R_j = r_j^{(t)}|Z=i)$$

So we can get

$$L = \frac{1}{T}\sum_{t=1}^{T}\log P\left(\left\{R_j = r_j^{(t)}\right\}_{j\in\Omega_t}\right)$$

$$= \frac{1}{T}\sum_{t=1}^{T}\log\left[\sum_i^k P(Z=i)\prod_{j\in\Omega_t} P(R_j = r_j^{(t)}|Z=i)\right]$$

In conclusion, we have

E-Step

$$\rho_{it} = P\left(Z=i\middle|\left\{R_j = r_j^{(t)}\right\}_{j\in\Omega_t}\right) = \frac{P(z=i)\prod_{j\in\Omega_t} P\left(R_j = r_j^{(t)}\middle|Z=i\right)}{\sum_{i'=1}^{k} P(Z=i')\prod_{j\in\Omega_t} P\left(R_j = r_j^{(t)}\middle|Z=i'\right)}$$

M-Step

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

$$P(R_j = 1|Z=i)\leftarrow\frac{\sum_{\{t|j\in\Omega_t\}}\rho_{it} I(r_j^{(t)},1) + \sum_{\{t|j\notin\Omega_t\}}\rho_{it} P\left(R_j = 1\middle|Z=i\right)}{\sum_{t=1}^{T}\rho_{it}}$$

$$= \frac{\sum_{\{t|j\in\Omega_t\}}\rho_{it} I(r_j^{(t)},1) + \sum_{\{t|j\notin\Omega_t\}}\rho_{it} P\left(R_j = 1\middle|Z=i\right)}{T * P(Z=i)}$$

Program result:

Iteration	Log-likelihood L
0	-27.03581500351123
1	-17.5604038243144
2	-16.00236263062783
4	-15.060597317892249
8	-14.501649272824999
16	-14.263788571437594
32	-14.180178075094377
64	-14.170077781591058
128	-14.163960358152186
256	-14.163692439007864

Of course, it increases at each iteration since auxiliary function guarantee that the log likelihood will be equal to or more than the previous one at each iteration.

```
CSE250A08
C:\Users\HP\AppData\Local\Programs\Python\Python37\python.exe
0	-27.03581500351123
1	-17.5604038243144
2	-16.00236263062783
4	-15.060597317892249
8	-14.501649272824999
16	-14.263788571437594
32	-14.180178075094377
64	-14.170077781591058
128	-14.163960358152186
256	-14.163692439007864

Process finished with exit code 0
```

(f)

$$P\left(R_l = 1 \middle| \left\{R_j = r_j^{(t)}\right\}_{j\in\Omega_t}\right) = \sum_{i=1}^{k} P\left(R_l = 1, Z = i \middle| \left\{R_j = r_j^{(t)}\right\}_{j\in\Omega_t}\right)$$

……………………………………………………………………………………marginalization

$$= \sum_{i=1}^{k} P\left(Z = i \middle| \left\{R_j = r_j^{(t)}\right\}_{j \in \Omega_t}\right) P\left(R_l = 1 \middle| Z = i, \left\{R_j = r_j^{(t)}\right\}_{j \in \Omega_t}\right)$$

……………………………………………………………………………………product rule

For $l \notin \Omega_t$, we can get

$$P\left(R_l = 1 \middle| \left\{R_j = r_j^{(t)}\right\}_{j \in \Omega_t}\right) = \sum_{i=1}^{k} P\left(Z = i \middle| \left\{R_j = r_j^{(t)}\right\}_{j \in \Omega_t}\right) P(R_l = 1 | Z = i)$$

From (c) and (e), we already get

$$\rho_{it} = P\left(Z = i \middle| \left\{R_j = r_j^{(t)}\right\}_{j \in \Omega_t}\right)$$

$$P(R_l = 1 | Z = i)$$

Program result for unseen movies:

0.9822638475036042	Inception
0.9787347806930398	Shutter_Island
0.9760742422425076	The_Dark_Knight_Rises
0.9611054183583134	The_Martian
0.9603155501551057	Interstellar
0.9532430101040604	The_Theory_of_Everything
0.9472529971957862	Avengers:_Infinity_War
0.9437806576835228	Gone_Girl
0.9384744247194854	Avengers:_Endgame
0.925516136493239	The_Social_Network
0.9241943603926942	The_Avengers
0.9042553518893099	Django_Unchained

0.8997468080954295	Now_You_See_Me
0.8913923867886028	Her
0.8883720635743731	21_Jump_Street
0.8840344102564319	Spiderman:_Far_From_Home
0.8805978987824168	Parasite
0.8750845098896503	Joker
0.8735765031739757	Toy_Story_3
0.8695427910650063	Ready_Player_One
0.8662703868828496	Les_Miserables
0.864680416620664	The_Farewell
0.8588349214983563	The_Hunger_Games
0.8527059384441631	Hidden_Figures
0.8526563490945032	Three_Billboards_Outside_Ebbing
0.8496688205775254	Manchester_by_the_Sea
0.8492519455814379	Black_Swan
0.8482974449335975	Iron_Man_2
0.848053053533284	Dunkirk
0.8469881568346	The_Perks_of_Being_a_Wallflower
0.8464343774019171	Darkest_Hour
0.8458414543422115	Captain_America:_The_First_Avenger
0.8418399522618606	Thor
0.8347995847816359	The_Revenant

0.8327094553965656	12_Years_a_Slave
0.8298949988123878	The_Hateful_Eight
0.8212573513855634	Ex_Machina
0.815256402726195	Jurassic_World
0.8124139810436862	Avengers:_Age_of_Ultron
0.8100846696924235	X-Men:_First_Class
0.8095517107595849	Once_Upon_a_Time_in_Hollywood
0.8063918625602785	La_La_Land
0.7918602656252836	Frozen
0.7887667950190065	Pitch_Perfect
0.7730172541835454	The_Girls_with_the_Dragon_Tattoo
0.7555483092725866	Good_Boys
0.7538517142356382	The_Help
0.7511939253594344	Mad_Max:_Fury_Road
0.7506828285452684	Us
0.7477295284441052	Fast_Five
0.7399508210219788	Drive
0.7126617472853463	Pokemon_Detective_Pikachu
0.7074691172594335	Midnight_in_Paris
0.7069865424434713	Room
0.7067184359070894	Terminator:_Dark_Fate
0.7058509544890318	American_Hustle

0.6765287080777059	Man_of_Steel
0.6671146614677038	World_War_Z
0.6549065428546698	Rocketman
0.6474300341645679	Star_Wars:_The_Force_Awakens
0.6295381732490605	Prometheus
0.6258179331135564	Fast_&_Furious:_Hobbs_&_Shaw
0.6200495624226724	The_Shape_of_Water
0.5988275491056453	I_Feel_Pretty
0.5357239938991715	Chappaquidick
0.5113088551619838	Phantom_Thread
0.4789044500884727	The_Last_Airbender
0.47832220141069975	Magic_Mike
0.4676020302594148	Hustlers
0.40992555087408605	Fifty_Shades_of_Grey