CSE 250A. Assignment 8 Solutions

Out: *Tue Nov 16*
Due: *Tue Nov 23*

8.1 EM algorithm for binary matrix completion

(a) **Sanity check**

You may paste screenshot of output to cover these tables (instead of filling them).

Rec (%)	Title	
	Movie	Rating
52	I_Feel_Pretty	0.358974
44	Fifty_Shades_of_Grey	0.377143
74	Hustlers	0.456522
4	The_Last_Airbender	0.473684
27	Magic_Mike	0.515152
68	Fast_&_Furious:_Hobbs_&_Shaw	0.540541
57	The_Shape_of_Water	0.558442
18	Prometheus	0.584071
60	Phantom_Thread	0.586207
34	World_War_Z	0.589552
41	Star_Wars:_The_Force_Awakens	0.593583
66	Rocketman	0.596774
53	Chappaquidick	0.600000
12	Bridemaids	0.619048
36	Man_of_Steel	0.624277
35	American_Hustle	0.628205
75	Terminator:_Dark_Fate	0.638889
46	Room	0.651515
71	Good_Boys	0.656250
73	Pokemon_Detective_Pikachu	0.656627
8	Fast_Five	0.658824
42	Mad_Max:_Fury_Road	0.678322
15	Drive	0.690141
72	Us	0.700000
16	The_Help	0.706897
24	Pitch_Perfect	0.720000

Rec (%)	Title	
	Movie	Rating
43	Jurassic_World	0.720930
30	Frozen	0.724138
14	X-Men:_First_Class	0.736041
49	The_Revenant	0.736364
38	Ex_Machina	0.739130
45	Avengers:_Age_of_Ultron	0.742424
54	La_La_Land	0.754717
17	Midnight_in_Paris	0.756098
56	Manchester_by_the_Sea	0.763636
63	Once_Upon_a_Time_in_Hollywood	0.768595
59	Three_Billboards_Outside_Ebbing	0.774194
61	Darkest_Hour	0.784810
29	The_Great_Gatsby	0.786408
58	Dunkirk	0.786885
32	Her	0.788889
10	Captain_America:_The_First_Avenger	0.791367
13	The_Girls_with_the_Dragon_Tattoo	0.793103
51	Ready_Player_One	0.797386
55	Hidden_Figures	0.797468
48	The_Hateful_Eight	0.800000
9	Thor	0.803150
7	Toy_Story_3	0.803347
22	The_Hunger_Games	0.814815
33	12_Years_a_Slave	0.816327
6	Iron_Man_2	0.817869
19	The_Perks_of_Being_a_Wallflower	0.819820

Rec (%)	Title	
	Movie	Rating
67	Joker	0.824000
25	Les_Miserables	0.825000
26	21_Jump_Street	0.825397
64	Spiderman:_Far_From_Home	0.827160
2	Black_Swan	0.829787
70	Parasite	0.836364
20	The_Avengers	0.843648
69	The_Farewell	0.860465
23	Django_Unchained	0.863946
31	Now_You_See_Me	0.864979
62	Avengers:_Endgame	0.868132
50	Avengers:_Infinity_War	0.878571
28	Wolf_of_Wall_Street	0.891667
65	The_Lion_King	0.894366
37	Gone_Girl	0.894410
5	Harry_Potter_and_the_Deathly_Hallows:_Part_1	0.900662
1	The_Social_Network	0.901639
11	Harry_Potter_and_the_Deathly_Hallows:_Part_2	0.917808
40	The_Theory_of_Everything	0.919540
39	Interstellar	0.938567
47	The_Martian	0.939394
21	The_Dark_Knight_Rises	0.941634
3	Shutter_Island	0.944444
0	Inception	0.965753

(b) **Likelihood**

$$
\begin{aligned}
P\left(\left\{R_j = r_j^{(t)}\right\}_{j\in\Omega_t}\right) &\overset{\text{marg}}{=} \sum_{i=1}^{k} P\left(\left\{R_j = r_j^{(t)}\right\}_{j\in\Omega_t}, Z=i\right) \\
&\overset{\text{P.R}}{=} \sum_{i=1}^{k} P(Z=i)\, P\left(\left\{R_j = r_j^{(t)}\right\}_{j\in\Omega_t} \,\middle|\, Z=i\right) \\
&\overset{\text{CI}}{=} \sum_{i=1}^{k} P(Z=i) \prod_{j\in\Omega_t} P\left(\left\{R_j = r_j^{(t)}\right. \,\middle|\, Z=i\right) \qquad \text{QED}
\end{aligned}
$$

(c) **E-step**

$$
\begin{aligned}
P\left(Z=i \,\middle|\, \left\{R_j = r_j^{(t)}\right\}_{j\in\Omega_t}\right) &= \frac{P\left(Z=i, \left\{R_j = r_j^{(t)}\right\}_{j\in\Omega_t}\right)}{P\left(\left\{R_j = r_j^{(t)}\right\}_{j\in\Omega_t}\right)} \\
&\overset{\text{answer b)}}{=} \frac{P\left(Z=i, \left\{R_j = r_j^{(t)}\right\}_{j\in\Omega_t}\right)}{\sum_{i=1}^{k} P(Z=i) \prod_{j\in\Omega_t} P\left(\left\{R_j = r_j^{(t)}\right. \,\middle|\, Z=i\right)} \\
&\overset{\text{P.R.}}{=} \frac{P(Z=i)\, P\left(\left\{R_j = r_j^{(t)}\right\}_{j\in\Omega_t} \,\middle|\, Z=i\right)}{\sum_{i'=1}^{k} P(Z=i') \prod_{j\in\Omega_t} P\left(\left\{R_j = r_j^{(t)}\right. \,\middle|\, Z=i'\right)} = \frac{P(Z=i) \prod_{j\in\Omega_t} P\left(\left\{R_j = r_j^{(t)}\right. \,\middle|\, Z=i\right)}{\sum_{i=1}^{k} P(Z=i) \prod_{j\in\Omega_t} P\left(\left\{R_j = r_j^{(t)}\right. \,\middle|\, Z=i\right)} \qquad \text{QED}
\end{aligned}
$$

(d) **M-step**

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

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

$$P(R_j=1 \mid Z=i) \longleftarrow \frac{\sum_t P\left(R_j=1, Z=i \,\middle|\, \{R_k=r_k^{(t)}\}_{k\in\Omega_t}\right)}{\sum_t P\left(Z=i \,\middle|\, \{R_k=r_k^{(t)}\}_{k\in\Omega_t}\right)} \leftarrow \rho_{it}$$

Numerator: $$\sum_t P\left(R_j=1, Z=i \,\middle|\, \{R_k=r_k^{(t)}\}_{k\in\Omega_t}\right) \overset{\text{P.R.}}{=} \sum_{\{t \mid j\in\Omega_t\}} I(r_j^{(t)},1)\, P\left(Z=i \,\middle|\, \{R_k=r_k^{(t)}\}_{k\in\Omega_t}\right) + \sum_{\{t \mid j\notin\Omega_t\}} P\left(Z=i \,\middle|\, \{R_k=r_k^{(t)}\}_{k\in\Omega_t}\right) P\left(R_j=1 \,\middle|\, Z=i, \{R_k=r_k^{(t)}\}_{k\in\Omega_t}\right)$$

$$\overset{\text{CI}}{=} \sum_{\{t \mid j\in\Omega_t\}} I(r_j^{(t)},1)\, \overbrace{P\left(Z=i \,\middle|\, \{R_k=r_k^{(t)}\}_{k\in\Omega_t}\right)}^{\rho_{it}} + \sum_{\{t \mid j\notin\Omega_t\}} \overbrace{P\left(Z=i \,\middle|\, \{R_k=r_k^{(t)}\}_{k\in\Omega_t}\right)}^{\rho_{it}} P(R_j=1 \mid Z=i)$$

$$\Rightarrow \quad P(R_j=1 \mid Z=i) \longleftarrow \frac{\sum_{\{t \mid j\in\Omega_t\}} I(r_j^{(t)},1)\rho_{it} + \sum_{\{t \mid j\notin\Omega_t\}} P(R_j=1 \mid Z=i)\rho_{it}}{\sum_{t=1}^{T}\rho_{it}} \qquad \text{QED}$$

(e) Implementation

iteration	log-likelihood $\mathcal{L}$
Iteration: 0	Log: -27.0358
Iteration: 1	Log: -17.5604
Iteration: 2	Log: -16.0024
Iteration: 4	Log: -15.0606
Iteration: 8	Log: -14.5016
Iteration: 16	Log: -14.2638
Iteration: 32	Log: -14.1802
Iteration: 64	Log: -14.1701
Iteration: 128	Log: -14.164
Iteration: 256	Log: -14.1637

(f) Personal movie recommendations

	Unseen movie	Expected rating
17	The_Martian	0.957777
3	Captain_America:_The_First_Avenger	0.919661
15	Gone_Girl	0.917336
2	Thor	0.908369
29	The_Farewell	0.899042
23	Manchester_by_the_Sea	0.865086
21	Chappaquidick	0.860513
1	Toy_Story_3	0.850382
10	The_Perks_of_Being_a_Wallflower	0.833299
27	Darkest_Hour	0.832298
30	Parasite	0.820570
5	The_Girls_with_the_Dragon_Tattoo	0.810251
6	X-Men:_First_Class	0.805592
22	Hidden_Figures	0.793803
25	Three_Billboards_Outside_Ebbing	0.780099
11	Frozen	0.778526
19	The_Revenant	0.774300
12	Her	0.771187
13	American_Hustle	0.764935
18	The_Hateful_Eight	0.763712
8	The_Help	0.748773
35	Terminator:_Dark_Fate	0.731152
26	Phantom_Thread	0.727953
32	Us	0.724827
31	Good_Boys	0.708575
7	Drive	0.697637
33	Pokemon_Detective_Pikachu	0.696073
16	Mad_Max:_Fury_Road	0.682776
14	Man_of_Steel	0.680190
4	Bridemaids	0.678331
28	Rocketman	0.646191
24	The_Shape_of_Water	0.583863
9	Prometheus	0.576349
0	The_Last_Airbender	0.540432
34	Hustlers	0.508773
20	I_Feel_Pretty	0.308193

These recommendations are fairly accurate!

(g) Source code

Source code is appended to the end of the document.

8.2 Mixture model decision boundary

(a) Posterior probability

$$P(y=1|\vec{x}) \overset{\text{Bayes}}{=} \frac{P(y=1)P(\vec{x}\,|\,y=1)}{P(\vec{x})}$$

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

$$= \frac{\pi_1 (2\pi)^{-\frac{d}{2}} |\Sigma_1|^{-\frac{1}{2}} e^{-\frac{1}{2}(\vec{x}-\mu_1)^T\Sigma_1^{-1}(\vec{x}-\mu_1)}}{\pi_1 (2\pi)^{-\frac{d}{2}} |\Sigma_1|^{-\frac{1}{2}} e^{-\frac{1}{2}(\vec{x}-\mu_1)^T\Sigma_1^{-1}(\vec{x}-\mu_1)} + \pi_0 (2\pi)^{-\frac{d}{2}} |\Sigma_0|^{-\frac{1}{2}} e^{-\frac{1}{2}(\vec{x}-\mu_0)^T\Sigma_0^{-1}(\vec{x}-\mu_0)}}$$

(b) **Decision boundary** $P(y=1\,|\,\vec{x}) = \dfrac{\pi_1 \cancel{(2\pi)^{-\frac{d}{2}}} \cancel{|\Sigma|^{-\frac{1}{2}}} e^{-\frac{1}{2}(\vec{x}-\mu_1)^T\Sigma^{-1}(\vec{x}-\mu_1)}}{\pi_1 \cancel{(2\pi)^{-\frac{d}{2}}} \cancel{|\Sigma|^{-\frac{1}{2}}} e^{-\frac{1}{2}(\vec{x}-\mu_1)^T\Sigma^{-1}(\vec{x}-\mu_1)} + \pi_0 \cancel{(2\pi)^{-\frac{d}{2}}} \cancel{|\Sigma|^{-\frac{1}{2}}} e^{-\frac{1}{2}(\vec{x}-\mu_0)^T\Sigma^{-1}(\vec{x}-\mu_0)}}$

$$= \frac{\pi_1 e^{-\frac{1}{2}(\vec{x}-\mu_1)^T\Sigma^{-1}(\vec{x}-\mu_1)}}{\pi_1 e^{-\frac{1}{2}(\vec{x}-\mu_1)^T\Sigma^{-1}(\vec{x}-\mu_1)} + \pi_0 e^{-\frac{1}{2}(\vec{x}-\mu_0)^T\Sigma^{-1}(\vec{x}-\mu_0)}} = \frac{1}{1 + \frac{\pi_0}{\pi_1} e^{-\frac{1}{2}(\vec{x}-\mu_0)^T\Sigma^{-1}(\vec{x}-\mu_0) + \frac{1}{2}(\vec{x}-\mu_1)^T\Sigma^{-1}(\vec{x}-\mu_1)}}$$

$\frac{\pi_0}{\pi_1} = e^{\log\frac{\pi_0}{\pi_1}}$

$$= \frac{1}{1 + e^{\log\frac{\pi_0}{\pi_1} + \left(-\frac{1}{2}(\vec{x}-\mu_0)^T\Sigma^{-1}(\vec{x}-\mu_0) + \frac{1}{2}(\vec{x}-\mu_1)^T\Sigma^{-1}(\vec{x}-\mu_1)\right)}} = \frac{1}{1 + e^{\log\frac{\pi_0}{\pi_1} + \left(-\frac{1}{2}(\vec{x}-\mu_0)^T\Sigma^{-1}(\vec{x}-\mu_0) + \frac{1}{2}(\vec{x}-\mu_1)^T\Sigma^{-1}(\vec{x}-\mu_1)\right)}}$$

$$= \frac{1}{1 + e^{\log\frac{\pi_0}{\pi_1} - \frac{1}{2}\left(\cancel{\vec{x}\Sigma^{-1}\vec{x}} - 2\vec{x}\Sigma^{-1}\mu_0 + \vec{\mu}_0\Sigma^{-1}\vec{\mu}_0\right) + \frac{1}{2}\left(\cancel{\vec{x}\Sigma^{-1}\vec{x}} - 2\vec{x}\Sigma^{-1}\vec{\mu}_1 + \vec{\mu}_1\Sigma^{-1}\vec{\mu}_1\right)}}$$

$$= \frac{1}{1 + e^{\log\frac{\pi_0}{\pi_1} + \vec{x}\Sigma^{-1}\mu_0 - \frac{1}{2}\vec{\mu}_0\Sigma^{-1}\vec{\mu}_0 - \vec{x}\Sigma^{-1}\vec{\mu}_1 + \frac{1}{2}\vec{\mu}_1\Sigma^{-1}\vec{\mu}_1}}$$

$$= \frac{1}{1 + e^{\log\frac{\pi_0}{\pi_1} + \vec{x}\Sigma^{-1}(\vec{\mu}_0 - \vec{\mu}_1) + \frac{1}{2}(\vec{\mu}_1\Sigma^{-1}\vec{\mu}_1 - \vec{\mu}_0\Sigma^{-1}\vec{\mu}_0)}}$$

$$= \frac{1}{1 + e^{-z}}, \quad \text{where } z = \vec{\omega}\vec{x} + b \quad \text{with } \vec{\omega} = \Sigma^{-1}(\mu_1 - \mu_0) \text{ and } b = -\log\frac{\pi_0}{\pi_1} - \frac{1}{2}(\vec{\mu}_1\Sigma^{-1}\vec{\mu}_1 - \vec{\mu}_0\Sigma^{-1}\vec{\mu}_0)$$

(c) **Shifting hyperplane**

$$\frac{P(y=1\,|\,\vec{x})}{P(y=0\,|\,\vec{x})} = \frac{\sigma(\omega\vec{x}+b)}{1-\sigma(\omega\vec{x}+b)} = \frac{\sigma(\omega\vec{x}+b)}{\sigma(-\omega\vec{x}-b)} = \frac{\frac{1}{1+e^{-(\vec{\omega}\vec{x}+b)}}}{\frac{1}{1+e^{\vec{\omega}\vec{x}+b}}} = \frac{1+e^{\vec{\omega}\cdot\vec{x}+b}}{1+e^{-(\vec{\omega}\cdot\vec{x}+b)}} = k$$

We let $z = \vec{\omega}\vec{x} + b$.

$$\Rightarrow \frac{1+e^{z}}{1+e^{-z}} = \frac{e^{z}e^{-z} + e^{z}}{1+e^{-z}} = \frac{e^{z}\cancel{(1+e^{-z})}}{\cancel{(1+e^{-z})}} = k$$

$$\Rightarrow e^{\vec{\omega}\vec{x}+b} = k$$

$\underline{\underline{\vec{\omega}\vec{x} + b = \log k}}$ is the hyperplane.

Boundary: $\vec{\omega}\vec{x} + b = \log 1 = 0$. OK.

8.3 Gradient ascent versus EM

(a) **Log likelihood**

$$\mathcal{L}(\vec{v}) = \sum_{t=1}^{T} \log P(y_t|\vec{x}_t)$$
$$= \sum_{t=1}^{T} y_t \log P(y_t=1|x_t) + (1-y_t)\log P(y_t=0|\vec{x}_t)$$
$$= \sum_{t=1}^{T} y_t \log(1-e^{-\vec{v}\cdot\vec{x}}) + (1-y_t)\log(1-(1-e^{-\vec{v}\cdot\vec{x}}))$$
$$= \sum_{t=1}^{T} y_t \log(1-e^{-\vec{v}\cdot\vec{x}}) + (1-y_t)\log e^{-\vec{v}\cdot\vec{x}}$$
$$= \underline{\underline{\sum_{t=1}^{T} y_t \log(1-e^{-\vec{v}\cdot\vec{x}}) - (1-y_t)\vec{v}\cdot\vec{x}}}$$

(b) **Gradient**

$$\frac{\partial\mathcal{L}}{\partial\vec{v}} = \sum_{t=1}^{T} \frac{y_t(-e^{-\vec{v}\cdot\vec{x}})(-\vec{x})}{1-e^{-\vec{v}\cdot\vec{x}}} - (1-y_t)\cdot x_t$$
$$= \sum_{t=1}^{T} x_t\left[\frac{y_t e^{-\vec{v}\cdot\vec{x}}}{\rho_t} - \frac{\rho_t}{\rho_t} + \frac{y_t\rho_t}{\rho_t}\right]$$
$$= \sum_{t=1}^{T} x_t\left[y_t\left(\frac{e^{-\vec{v}\cdot\vec{x}} + (1-e^{-\vec{v}\cdot\vec{x}})}{\rho_t}\right) - \frac{\rho_t}{\rho_t}\right]$$
$$= \underline{\underline{\sum_{t=1}^{T} x_t\left[\frac{y_t-\rho_t}{\rho_t}\right]}} \qquad \boxed{\text{Expression (ii)}}$$

(c) **Noisy-OR**

$$P(y=1|\vec{x}) = 1 - e^{-\vec{v}\cdot\vec{x}}$$

$$= 1 - e^{-(-\vec{x}\log(1-\vec{p}))}$$

$$= 1 - e^{\log(1-\vec{p})\vec{x}}$$

$$= 1 - e^{\sum_{i=1}^{k}\log(1-p_i)\cdot x_i}$$

$$= 1 - e^{\log\prod_{i=1}^{k}(1-p_i)^{x_i}}$$

$$= 1 - \prod_{i=1}^{k}(1-p_i)^{x_i}$$

(d) **Chain rule**

$$\frac{\partial \mathcal{L}}{\partial p_i} = \frac{\partial \mathcal{L}}{\partial v_i}\frac{\partial v_i}{\partial p_i}$$

$$\frac{\partial v_i}{\partial p_i} = \frac{\partial}{\partial p_i}\left(-\log(1-p_i)\right) = \frac{1}{1-p_i}$$

$$\Rightarrow \frac{\partial \mathcal{L}}{\partial p_i} = \frac{1}{1-p_i}\frac{\partial \mathcal{L}}{\partial v_i} \qquad \text{QED}$$

(e) **Gradient ascent versus EM**

$$p_i \longleftarrow p_i + \eta_i\left(\frac{\partial \mathcal{L}}{\partial p_i}\right)$$

$$p_i + \eta_i\left(\frac{\partial \mathcal{L}}{\partial p_i}\right) = p_i + \frac{p_i\cancel{(1-p_i)}}{T_i}\cdot\frac{1}{\cancel{1-p_i}}\frac{\partial \mathcal{L}}{\partial v_i}$$

$$= p_i + \frac{p_i}{T_i}\frac{\partial \mathcal{L}}{\partial v_i}$$

$$= p_i + \frac{p_i}{T_i}\sum_{t=1}^{T}x_{it}\left[\frac{y_t-\rho_t}{\rho_t}\right]$$

$$= p_i + \frac{p_i}{T_i}\sum_{t=1}^{T}x_{it}\left[\frac{y_t-\rho_t}{\rho_t}\right]$$

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

$$= p_i + \frac{p_i}{T_i}\sum_{t=1}^{T}\frac{x_{it}y_t}{\rho_t} - \frac{p_i}{T_i}\sum_{t=1}^{T}x_{it}$$

$$= p_i + \frac{p_i}{T_i}\sum_{t=1}^{T}\frac{x_{it}y_t}{\rho_t} - \frac{p_i}{\cancel{T_i}}\cancel{T_i}$$

$$= \frac{p_i}{T_i}\sum_{t=1}^{T}\frac{x_{it}y_t}{\rho_t} \qquad \text{QED} \qquad \text{(Comparing the updates)}$$

8.4 Similarity learning with logistic regression

(a) **Inference for similar examples**

$$
\begin{aligned}
P(y=1,y'=1\,|\,\vec{x},\vec{x}',s=1) &\overset{\text{Bayes}}{=} \frac{P(s=1\,|\,y=1,y'=1,\vec{x},\vec{x}')\,P(y=1,y'=1\,|\,\vec{x},\vec{x}')}{P(s=1\,|\,\vec{x},\vec{x}')} \\
&\overset{\text{CI \& P.R.}}{=} \frac{P(s=1\,|\,y=1,y'=1)\,P(y=1\,|\,\vec{x},\vec{x}')\,P(y'=1\,|\,\vec{x},\vec{x}',y=1)}{P(s=1\,|\,\vec{x},\vec{x}')} \\
&\overset{\text{CI}}{=} \frac{P(s=1\,|\,y=1,y'=1)\,P(y=1\,|\,\vec{x})\,P(y'=1\,|\,\vec{x}')}{P(s=1\,|\,\vec{x},\vec{x}')} \\
&\overset{\text{CPT}}{=} \frac{1\cdot\sigma(\vec{\omega}\cdot\vec{x})\cdot\sigma(\vec{\omega}\cdot\vec{x}')}{P(s=1\,|\,\vec{x},\vec{x}')} \\
&\overset{\text{marg}}{=} \frac{1\cdot\sigma(\vec{\omega}\cdot\vec{x})\cdot\sigma(\vec{\omega}\cdot\vec{x}')}{\sum_{y,y'} P(s=1,y,y'\,|\,\vec{x},\vec{x}')} \\
&= \frac{1\cdot\sigma(\vec{\omega}\cdot\vec{x})\cdot\sigma(\vec{\omega}\cdot\vec{x}')}{\sum_{y,y'} P(s=1\,|\,y,y')\,P(y\,|\,\vec{x})\,P(y'\,|\,\vec{x}')} \\
&= \frac{1\cdot\sigma(\vec{\omega}\cdot\vec{x})\cdot\sigma(\vec{\omega}\cdot\vec{x}')}{\sum_{y,y'} I(y,y')\,P(y\,|\,\vec{x})\,P(y'\,|\,\vec{x}')} \\
&= \frac{\sigma(\vec{\omega}\cdot\vec{x})\cdot\sigma(\vec{\omega}\cdot\vec{x}')}{P(y=0\,|\,\vec{x})P(y'=0\,|\,\vec{x}')+P(y=1\,|\,\vec{x})P(y'=1\,|\,\vec{x}')} \\
&= \frac{\sigma(\vec{\omega}\cdot\vec{x})\cdot\sigma(\vec{\omega}\cdot\vec{x}')}{(1-\sigma(\vec{\omega}\cdot\vec{x}))(1-\sigma(\vec{\omega}\cdot\vec{x}'))+\sigma(\vec{\omega}\cdot\vec{x})\sigma(\vec{\omega}\cdot\vec{x}')} \\
&= \underline{\underline{\frac{\sigma(\vec{\omega}\cdot\vec{x})\cdot\sigma(\vec{\omega}\cdot\vec{x}')}{\sigma(-\vec{\omega}\cdot\vec{x})\sigma(-\vec{\omega}\cdot\vec{x}')+\sigma(\vec{\omega}\cdot\vec{x})\sigma(\vec{\omega}\cdot\vec{x}')}}}
\end{aligned}
$$

(b) **Inference for dissimilar examples**

$$
\begin{aligned}
P(y=1,y'=0\,|\,\vec{x},\vec{x}',s=0) &\overset{\text{Bayes}}{=} \frac{P(s=0\,|\,y=1,y'=0,\vec{x},\vec{x}')\,P(y=1,y'=0\,|\,\vec{x},\vec{x}')}{P(s=0\,|\,\vec{x},\vec{x}')} \\
&\overset{\text{CI \& P.R.}}{=} \frac{P(s=0\,|\,y=1,y'=0)\,P(y=1\,|\,\vec{x},\vec{x}')\,P(y'=0\,|\,\vec{x},\vec{x}',y=1)}{P(s=0\,|\,\vec{x},\vec{x}')} \\
&\overset{\text{CI}}{=} \frac{P(s=0\,|\,y=1,y'=0)\,P(y=1\,|\,\vec{x})\,P(y'=0\,|\,\vec{x}')}{P(s=0\,|\,\vec{x},\vec{x}')} \\
&\overset{\text{CPT}}{=} \frac{1\cdot\sigma(\vec{\omega}\cdot\vec{x})\cdot\sigma(-\vec{\omega}\cdot\vec{x}')}{P(s=1\,|\,\vec{x},\vec{x}')} \\
&\overset{\text{marg}}{=} \frac{1\cdot\sigma(\vec{\omega}\cdot\vec{x})\cdot\sigma(-\vec{\omega}\cdot\vec{x}')}{\sum_{y,y'} P(s=0,y,y'\,|\,\vec{x},\vec{x}')} \\
&= \frac{1\cdot\sigma(\vec{\omega}\cdot\vec{x})\cdot\sigma(-\vec{\omega}\cdot\vec{x}')}{\sum_{y,y'} P(s=0\,|\,y,y')\,P(y\,|\,\vec{x})\,P(y'\,|\,\vec{x}')} \\
&\overset{\substack{\text{dissimilar}\\ y\,\&\,y'}}{=} \frac{\sigma(\vec{\omega}\cdot\vec{x})\cdot\sigma(-\vec{\omega}\cdot\vec{x}')}{P(y=0\,|\,\vec{x})P(y'=1\,|\,\vec{x}')+P(y=1\,|\,\vec{x})P(y'=0\,|\,\vec{x}')} \\
&= \underline{\underline{\frac{\sigma(\vec{\omega}\cdot\vec{x})\cdot\sigma(-\vec{\omega}\cdot\vec{x}')}{\sigma(-\vec{\omega}\cdot\vec{x})\sigma(\vec{\omega}\cdot\vec{x}')+\sigma(\vec{\omega}\cdot\vec{x})\sigma(-\vec{\omega}\cdot\vec{x}')}}}
\end{aligned}
$$

(c) **E-Step**

$$(\mathbf{b}) \quad P(y=1|\vec{x},\vec{x}',s=1) \;=\; P(y=1, y'=1 \,|\, \vec{x}, \vec{x}', s=1)$$

$$(\mathbf{b}) \quad P(y'=1|\vec{x},\vec{x}',s=1) \;=\; P(y=1, y'=1 \,|\, \vec{x}, \vec{x}', s=1)$$

$$(\mathbf{a}) \quad P(y=1|\vec{x},\vec{x}',s=0) \;=\; P(y=1, y'=0 \,|\, \vec{x}, \vec{x}', s=0)$$

$$(\mathbf{c}) \quad P(y'=1|\vec{x},\vec{x}',s=0) \;=\; 1 - P(y=1, y'=0 \,|\, \vec{x}, \vec{x}', s=0)$$

(d) **Log-likelihood**

$$\mathcal{L}(\vec{\omega}) = \sum_t \log P(s_t \,|\, \vec{x}_t, \vec{x}_t')$$

$$= \sum_t s_t \log P(s=1 \,|\, \vec{x}_t, \vec{x}_t') + (1-s_t) \log P(s=0 \,|\, \vec{x}_t, \vec{x}_t')$$

from a/b)

$$\underline{\underline{= \sum_t s_t \log\Big(\sigma(-\vec{\omega}\cdot\vec{x})\,\sigma(-\vec{\omega}\cdot\vec{x}') + \sigma(\vec{\omega}\cdot\vec{x})\,\sigma(\vec{\omega}\cdot\vec{x}')\Big) + (1-s_t)\log\Big(\sigma(-\vec{\omega}\cdot\vec{x})\,\sigma(\vec{\omega}\cdot\vec{x}') + \sigma(\vec{\omega}\cdot\vec{x})\,\sigma(-\vec{\omega}\cdot\vec{x}')\Big)}}$$

(e) **M-step**

$$\boxed{\vec{w} \leftarrow \vec{w} + \eta \left\{ \sum_t \Big[\big(\bar{y}_t - \sigma(\vec{\omega}\cdot\vec{x}_t)\big)\vec{x}_t + \big(\bar{y}_t' - \sigma(\vec{\omega}\cdot\vec{x}_t')\big)\vec{x}_t' \Big] \right\}}$$

8.5 Logistic regression across time

(a) **Likelihood**

$$\begin{aligned}
\alpha_{j,t+1} &= P(Y_{t+1}=j \mid y_0, \vec{x}_1, \dots, \vec{x}_{t+1}) \\
&\overset{\text{marg}}{=} \sum_{k\in\{0,1\}} P(y_{t+1}=j, y_t=k \mid y_0, \vec{x}_1, \dots, \vec{x}_t, \vec{x}_{t+1}) \\
&\overset{\text{P.R.}}{=} \sum_{k\in\{0,1\}} P(y_t=k \mid y_0, \vec{x}_1, \dots, \vec{x}_t, \vec{x}_{t+1}) P(y_{t+1}=j \mid y_t=k, y_0, \vec{x}_1, \dots, \vec{x}_t, \vec{x}_{t+1}) \\
&\overset{\text{CI}}{=} \sum_{k\in\{0,1\}} P(y_t=k \mid y_0, \vec{x}_1, \dots, \vec{x}_t) P(y_{t+1}=j \mid y_t=k, \vec{x}_{t+1}) \\
&= \sum_{k\in\{0,1\}} \alpha_{it} \cdot P(y_{t+1}=j \mid y_t=k, \vec{x}_{t+1}) \\
&= \underline{\underline{\sum_{k\in\{0,1\}} \alpha_{it} \left(j \cdot \sigma(\vec{\omega}_k \cdot x_{t+1}) + (1-j)\sigma(\vec{\omega}_k \cdot x_{t+1})\right)}}
\end{aligned}$$

$$\cancel{= \sum_{k\in\{0,1\}} \alpha_{it} \cdot \sigma(\vec{\omega}_k \cdot \vec{x}_{t+1}) \quad = \alpha_{it}\left(\sigma(\vec{\omega}_0 \cdot \vec{x}_{t+1}) + \sigma(\vec{\omega}_1 \cdot \vec{x}_{t+1})\right)}$$

(b) **Most likely hidden states**

$$\begin{aligned}
\ell^*_{j,t+1} &= \max_{y_1,\dots,y_t} \left[\log P(y_1, \dots, y_t, y_{t+1}=i \mid y_0, \vec{x}_1, \dots, \vec{x}_{t+1})\right] \\
&= \max_{y_1,\dots,y_{t-1}} \max_j \left[\log P(y_1, \dots, y_{t-1}, y_t=j, y_{t+1}=i \mid y_0, \vec{x}_1, \dots, x_{t+1})\right] \\
&= \max_{y_1,\dots,y_{t-1}} \max_j \left[\log P(y_1, \dots, y_{t-1}, y_t=j \mid y_0, \vec{x}_1, \dots, x_{t+1}) + \log P(y_{t+1}=i \mid y_0, \vec{x}_1, \dots, \vec{x}_{t+1}, y_1, \dots, y_{t-1}, y_t=j)\right] \\
&= \max_{y_1,\dots,y_{t-1}} \max_j \left[\log P(y_1, \dots, y_{t-1}, y_t=j \mid y_0, \vec{x}_1, \dots, x_t) + \log P(y_{t+1}=i \mid \vec{x}_{t+1}, y_t=j)\right] \\
&= \underline{\underline{\max_j \left[\ell^*_{jt} + \log\left(i\left(j\sigma(\vec{\omega}_1 \cdot x_{t+1}) + (1-j)\sigma(\vec{\omega}_1 \cdot x_{t+1})\right) + (1-i)\left(j(1-\sigma(\vec{\omega}_1 \cdot x_{t+1})) + (1-j)(1-\sigma(\vec{\omega}_1 \cdot x_{t+1}))\right)\right)\right]}}
\end{aligned}$$

(c) **Prediction**

for $t = 1$ to $T-1$

for $j = 0$ to 1

$$\Phi_{t+1}(j) = \operatorname{argmax}_{i\in\{0,1\}} \left[\frac{1}{\alpha_{it+1}}\left(\ell^*_{it} + \log\left(j\left(i\,\sigma(\vec{\omega}_1 \cdot x_{t+1}) + (1-i)\sigma(\vec{\omega}_1 \cdot x_{t+1})\right) + (1-j)\left(i(1-\sigma(\vec{\omega}_1 \cdot x_{t+1})) + (1-i)(1-\sigma(\vec{\omega}_1 \cdot x_{t+1}))\right)\right)\right)\right]$$

for $t = T-1$ to 1

$$y^*_t = \left[\ \phi_{t+1}(y^*_{t+1}) \qquad\qquad \right]$$

Source code 1g)

a)

```python
import pandas as pd
import numpy as np
```

[1] ✓ 2.1s Python

```python
movies = open('hw8_movies.txt').read().splitlines()
ids = open('hw8_ids.txt').read().splitlines()
ratings = np.loadtxt('hw8_ratings.txt', dtype='str')
```

[2] ✓ 0.1s Python

```python
movie_with_ratings = []
for i in range(len(movies)):
    rating = ratings[:,i]
    recommended = sum(rating == "1")
    seen = sum(rating != "?")
    movie_with_ratings.append((movies[i], recommended/seen))

df = pd.DataFrame(movie_with_ratings, columns =['Movie', 'Rating'])
df = df.sort_values(by="Rating")
# Print first 1-26, then 27-52, then 53-77
print(df[52:76])
```

[22] ✓ 0.4s Python

e) **e) Implementation**

```python
probZ = np.loadtxt('hw8_probZ_init.txt', dtype='float32')
probR = np.loadtxt('hw8_probR_init.txt', dtype='float32')
```

[23] ✓ 0.6s Python

```python
k = 4
n_iter = 256
T = len(ids)
```

[26] ✓ 0.6s Python

```python
prob_z = np.copy(probZ)
prob_r = np.copy(probR)
posteriors = np.empty([k,T], dtype='float32')
log_likelihoods = []
```

[29] ✓ 0.6s Python

```python
def e_step(i, t, prob_z, prob_r):
    recommended, = np.where(ratings[t,:] == "1")
    not_recommended, = np.where(ratings[t,:] == "0")
    numerator = prob_z[i]*np.prod(prob_r[recommended,i])*np.prod(1-prob_r[not_recommended,i])
    denominator = 0
    for x in range(k):
        denominator += prob_z[x]*np.prod(prob_r[recommended,x])*np.prod(1-prob_r[not_recommended,x])
    return numerator/denominator
```

✓ 0.3s Python

```python
def m_step(i, j, posteriors, prob_r):
    seen, = np.where(ratings[:,j] == "1")
    not_seen, = np.where(ratings[:,j] == "?")
    return (np.sum(posteriors[i,seen])+prob_r[j,i]*np.sum(posteriors[i,not_seen]))/np.sum(posteriors[i,:])
```

✓ 0.4s Python

```python
def log_likelihood(t, prob_z, prob_r):
    sum = 0
    for i in range(k):
        recommended, = np.where(ratings[t,:] == "1")
        not_recommended, = np.where(ratings[t,:] == "0")
        sum += prob_z[i]*np.prod(prob_r[recommended,i])*np.prod(1-prob_r[not_recommended,i])
    return np.log(sum)
```

✓ 0.5s Python

```python
def em(prob_r, prob_z, log_likelihoods, posteriors, n_iter):
    for iter in range(n_iter+1):
        loglike = 0
        prob_z_temp = np.empty(k)
        prob_r_temp = np.empty([len(movies),k])
        for t in range(T):
            loglike += log_likelihood(t, prob_z, prob_r)
            for i in range(k):
                posteriors[i,t] = e_step(i, t, prob_z, prob_r)
        for i in range(k):
            prob_z_temp[i] = np.sum(posteriors[i,:])/T
            for j in range(len(movies)):
                prob_r_temp[j,i] = m_step(i, j, posteriors, prob_r)
        log_likelihoods.append(loglike/T)
        prob_z = prob_z_temp
        prob_r = prob_r_temp
    return log_likelihoods, posteriors, prob_r, prob_z
```

✓ 0.3s Python

```python
log_likelihoods, post, probr, probz = em(prob_r, prob_z, log_likelihoods, posteriors, n_iter)
```

✓ 35.2s Python

```python
print(f"Iteration: 0 \t Log: {round(log_likelihoods[0],4)}")
print(f"Iteration: 1 \t Log: {round(log_likelihoods[1],4)}")
print(f"Iteration: 2 \t Log: {round(log_likelihoods[2],4)}")
print(f"Iteration: 4 \t Log: {round(log_likelihoods[4],4)}")
print(f"Iteration: 8 \t Log: {round(log_likelihoods[8],4)}")
print(f"Iteration: 16 \t Log: {round(log_likelihoods[16],4)}")
print(f"Iteration: 32 \t Log: {round(log_likelihoods[32],4)}")
print(f"Iteration: 64 \t Log: {round(log_likelihoods[64],4)}")
print(f"Iteration: 128 \t Log: {round(log_likelihoods[128],4)}")
print(f"Iteration: 256 \t Log: {round(log_likelihoods[256],4)}")
```

✓ 0.5s Python

f) Personal movie recommendations

f)

```python
idx = ids.index("U09085243")
not_seen, = np.where(ratings[idx,:] == "?")
unseen_with_expected = []

for movie in not_seen:
    expected = 0
    for i in range(k):
        estep = e_step(i, idx, probz, probr)
        mstep = m_step(i, movie, post, probr)
        expected += estep*mstep
    unseen_with_expected.append((movies[movie], expected))

df2 = pd.DataFrame(unseen_with_expected, columns =['Unseen movie', 'Expected rating'])
df2 = df2.sort_values(by="Expected rating", ascending=False)
```

✓ 0.1s Python

```python
df2
```

✓ 0.1s Python

	Unseen movie	Expected rating
17	The_Martian	0.957777
3	Captain_America:_The_First_Avenger	0.919661