## 7.1 Two-state MDP *(8 pts)*

(a) **Policy evaluation**

*(2 pts)* **From the Bellman equation:**

$$\begin{aligned}
V^\pi(0) &= R(0) + \gamma\left[P(s'\!=\!0|s\!=\!0,a\!=\!\downarrow)V^\pi(0) + P(s'\!=\!1|s\!=\!0,a\!=\!\downarrow)V^\pi(1)\right] \\
&= -1 + \tfrac{1}{2}\left[\tfrac{3}{4}V^\pi(0) + \tfrac{1}{4}V^\pi(1)\right] \\
&= -1 + \tfrac{3}{8}V^\pi(0) + \tfrac{1}{8}V^\pi(1)
\end{aligned}$$

$$\begin{aligned}
V^\pi(1) &= R(1) + \gamma\left[P(s'\!=\!0|s\!=\!1,a\!=\!\downarrow)V^\pi(0) + P(s'\!=\!1|s\!=\!1,a\!=\!\downarrow)V^\pi(1)\right] \\
&= 2 + \tfrac{1}{2}\left[\tfrac{1}{4}V^\pi(0) + \tfrac{3}{4}V^\pi(1)\right] \\
&= 2 + \tfrac{1}{8}V^\pi(0) + \tfrac{3}{8}V^\pi(1)
\end{aligned}$$

*(2 pts)* **Rearranging the above and solving:**

$$\begin{aligned}
\tfrac{5}{8}V^\pi(0) \,-\, \tfrac{1}{8}V^\pi(1) &= -1 \\
-\tfrac{1}{8}V^\pi(0) \,+\, \tfrac{5}{8}V^\pi(1) &= 2
\end{aligned}$$

**A little algebra gives** $V^\pi(0) = -1$ and $V^\pi(1) = 3$.

(b) **Greedy policy**

*(1 pt)* **Action-value function for action** $a\!=\!\downarrow$**:**

$$\begin{aligned}
Q^\pi(s\!=\!0,a\!=\!\downarrow) &= V^\pi(0) \;=\; -1 \\
Q^\pi(s\!=\!1,a\!=\!\downarrow) &= V^\pi(1) \;=\; 3
\end{aligned}$$

*(2 pts)* **Action-value function for action** $a\!=\!\uparrow$**:**

$$\begin{aligned}
Q^\pi(s\!=\!0,a\!=\!\uparrow) &= R(0) + \gamma\left[P(s'\!=\!0|s\!=\!0,a\!=\!\uparrow)V^\pi(0) + P(s'\!=\!1|s\!=\!0,a\!=\!\uparrow)V^\pi(1)\right] \\
&= -1 + \tfrac{1}{2}\left[\tfrac{1}{2}(-1) + \tfrac{1}{2}(3)\right] \\
&= -0.5
\end{aligned}$$

$$\begin{aligned}
Q^\pi(s\!=\!1,a\!=\!\uparrow) &= R(1) + \gamma\left[P(s'\!=\!0|s\!=\!1,a\!=\!\uparrow)V^\pi(0) + P(s'\!=\!1|s\!=\!1,a\!=\!\uparrow)V^\pi(1)\right] \\
&= 2 + \tfrac{1}{2}\left[\tfrac{1}{2}(-1) + \tfrac{1}{2}(3)\right] \\
&= 2.5
\end{aligned}$$

*(1 pt)* **Greedy policy:**

$$\begin{aligned}
\pi'(0) &= \operatorname*{argmax}_a Q^\pi(s\!=\!0,a) \;=\uparrow \\
\pi'(1) &= \operatorname*{argmax}_a Q^\pi(s\!=\!1,a) \;=\downarrow
\end{aligned}$$

## 7.2 Three-state MDP *(12 pts)*

(a) **Policy evaluation**

*(3 pts)* **From the Bellman equation:**

$$\begin{aligned}
V^\pi(1) &= R(1) + \gamma \left[P(s'\!=\!1|s\!=\!1, a=\uparrow)V^\pi(1) + P(s'\!=\!2|s\!=\!1, a=\uparrow)V^\pi(2) + P(s'\!=\!3|s\!=\!1, a=\uparrow)V^\pi(3)\right] \\
&= -15 + \tfrac{2}{3}\left[\tfrac{3}{4}V^\pi(1) + \tfrac{1}{4}V^\pi(2) + (0)V^\pi(3)\right] \\
&= -15 + \tfrac{1}{2}V^\pi(1) + \tfrac{1}{6}V^\pi(2)
\end{aligned}$$

$$\begin{aligned}
V^\pi(2) &= R(2) + \gamma \left[P(s'\!=\!1|s\!=\!2, a=\uparrow)V^\pi(1) + P(s'\!=\!2|s\!=\!2, a=\uparrow)V^\pi(2) + P(s'\!=\!3|s\!=\!2, a=\uparrow)V^\pi(3)\right] \\
&= 30 + \tfrac{2}{3}\left[\tfrac{1}{2}V^\pi(1) + \tfrac{1}{2}V^\pi(2) + (0)V^\pi(3)\right] \\
&= 30 + \tfrac{1}{3}V^\pi(1) + \tfrac{1}{3}V^\pi(2)
\end{aligned}$$

$$\begin{aligned}
V^\pi(3) &= R(3) + \gamma \left[P(s'\!=\!1|s\!=\!3, a=\downarrow)V^\pi(1) + P(s'\!=\!2|s\!=\!3, a=\downarrow)V^\pi(2) + P(s'\!=\!3|s\!=\!3, a=\downarrow)V^\pi(3)\right] \\
&= -25 + \tfrac{2}{3}\left[(0)V^\pi(1) + \tfrac{1}{4}V^\pi(2) + \tfrac{3}{4}V^\pi(3)\right] \\
&= -25 + \tfrac{1}{6}V^\pi(2) + \tfrac{1}{2}V^\pi(3)
\end{aligned}$$

*(1 pt)* **Rearranging the above:**

$$\begin{aligned}
15 &= -\tfrac{1}{2}V^\pi(1) \;+\; \tfrac{1}{6}V^\pi(2), \\
30 &= -\tfrac{1}{3}V^\pi(1) \;+\; \tfrac{2}{3}V^\pi(2), \\
25 &= \tfrac{1}{6}V^\pi(2) \;-\; \tfrac{1}{2}V^\pi(3).
\end{aligned}$$

*(1 pt)* **A little algebra gives:**

$$\begin{aligned}
V^\pi(1) &= -18, \\
V^\pi(2) &= +36, \\
V^\pi(3) &= -38.
\end{aligned}$$

(b) **Greedy policy**

*(3 pts)* **Action-value function for action $a=\uparrow$:**

$$\begin{aligned}
Q^\pi(s\!=\!1, a=\uparrow) &= V^\pi(1) \;=\; -18, \\
Q^\pi(s\!=\!2, a=\uparrow) &= V^\pi(2) \;=\; 36, \\
Q^\pi(s\!=\!3, a=\uparrow) &= R(3) + \gamma\left[P(s'\!=\!1|s\!=\!3, a=\uparrow)V^\pi(1) + P(s'\!=\!2|s\!=\!3, a=\uparrow)V^\pi(2) + P(s'\!=\!3|s\!=\!3, a=\uparrow)V^\pi(3)\right], \\
&= -25 + \tfrac{2}{3}\left[(0)(-18) + \tfrac{3}{4}(36) + \tfrac{1}{4}(-38)\right], \\
&= -\tfrac{40}{3}.
\end{aligned}$$

*(3 pts)* **Action-value function for action $a = \downarrow$:**

$$\begin{aligned}
Q^\pi(s{=}1, a{=}\downarrow) &= R(1) + \gamma \left[P(s'{=}1|s{=}1, a{=}\downarrow)V^\pi(1) + P(s'{=}2|s{=}1, a{=}\downarrow)V^\pi(2) + P(s'{=}3|s{=}1, a{=}\downarrow)V^\pi(3)\right], \\
&= -15 + \tfrac{2}{3}\left[\tfrac{1}{4}(-18) + \tfrac{3}{4}(36) + (0)(-38)\right], \\
&= 0, \\
Q^\pi(s{=}2, a{=}\downarrow) &= R(2) + \gamma \left[P(s'{=}1|s{=}2, a{=}\downarrow)V^\pi(1) + P(s'{=}2|s{=}2, a{=}\downarrow)V^\pi(2) + P(s'{=}3|s{=}2, a{=}\downarrow)V^\pi(3)\right], \\
&= 30 + \tfrac{2}{3}\left[(0)(-18) + \tfrac{1}{2}(36) + \tfrac{1}{2}(-38)\right], \\
&= \tfrac{88}{3}, \\
Q^\pi(s{=}3, a{=}\downarrow) &= V^\pi(3) \; = \; -38.
\end{aligned}$$

*(1 pt)* **Greedy policy:**

$$\begin{aligned}
\pi'(1) &= \operatorname*{argmax}_a Q^\pi(s{=}1, a) \; = \downarrow \\
\pi'(2) &= \operatorname*{argmax}_a Q^\pi(s{=}2, a) \; = \uparrow, \\
\pi'(3) &= \operatorname*{argmax}_a Q^\pi(s{=}3, a) \; = \uparrow.
\end{aligned}$$

## 7.3 EM algorithm for binary matrix completion *(25 pts)*

(a) **Sanity check** *(2 pts)*

| Title | Rec (%) |
|---|---|
| Solo | 0.37755102040816324 |
| Justice League | 0.3900709219858156 |
| The Shape of Water | 0.3956043956043956 |
| Ex Machina | 0.4631578947368421 |
| Star Trek Beyond | 0.46534653465346537 |
| Batman v Superman: Dawn of Justice | 0.47904191616766467 |
| Star Wars: The Last Jedi | 0.4857142857142857 |
| Terminator Genisys | 0.48717948717948717 |
| Tron | 0.5075757575757576 |
| Suicide Squad | 0.5111111111111111 |
| Mad Max: Fury Road | 0.5227272727272727 |
| The Last Airbender | 0.5246913580246914 |
| Wonder Woman | 0.536723163841808 |
| Ant-Man and the Wasp | 0.5440414507772021 |
| It | 0.5496688741721855 |
| World War Z | 0.5555555555555556 |
| Oceans 8 | 0.5725190839694656 |
| Man of Steel | 0.577639751552795 |
| Jumanji: Welcome to the Jungle | 0.5857988165680473 |
| 2001: A Space Odyssey | 0.5871559633027523 |
| Get Out | 0.5970149253731343 |
| Furious 7 | 0.5973154362416108 |
| Star Wars: The Phantom Menace | 0.6024844720496895 |
| Moana | 0.6116504854368932 |
| Rogue One | 0.6183206106870229 |
| Logan | 0.6277372262773723 |
| Terminator 2 | 0.6370370370370371 |
| The Greatest Showman | 0.6462585034013606 |
| The Lego Movie | 0.6467661691542289 |
| Fantastic Beasts and Where To Find Them | 0.6628571428571428 |
| Blade Runner 2049 | 0.6692913385826772 |
| Venom | 0.6974358974358974 |
| Frozen | 0.7 |
| Thor: Ragnarok | 0.7070707070707071 |
| Deadpool 2 | 0.7142857142857143 |
| Guardians of the Galaxy Vol. 2 | 0.7202072538860104 |
| Jurassic World | 0.7251184834123223 |
| Mission: Impossible - Fallout | 0.7483443708609272 |
| Captain America: Civil War | 0.7534883720930232 |
| Guardians of the Galaxy | 0.7610619469026548 |
| La La Land | 0.7627118644067796 |
| The Lord of the Rings: The Fellowship of the Ring | 0.7705882352941177 |
| Coco | 0.78 |
| The Hunger Games | 0.7802690582959642 |
| Iron Man 3 | 0.780373831775701 |
| Zootopia | 0.7905138339920948 |
| Black Panther | 0.7991071428571429 |
| The Martian | 0.8011049723756906 |
| Avengers: Infinity War | 0.8073770491803278 |
| The Wolf of Wall Street | 0.8082901554404145 |
| The Imitation Game | 0.8106060606060606 |
| Harry Potter and the Deathly Hallows: Part 2 | 0.8108108108108109 |
| The Avengers | 0.8285714285714286 |
| The Matrix | 0.8350515463917526 |
| Jurassic Park (1993) | 0.839572192513369 |
| Doctor Strange | 0.8401826484018264 |
| The Dark Knight | 0.8412698412698413 |
| WALL-E | 0.8434782608695652 |
| Inception | 0.8723404255319149 |
| Interstellar | 0.8858447488584474 |

(b) **Likelihood** *(3 pts)*

$$\begin{aligned}
P\left(\left\{R_j = r_j^{(t)}\right\}_{j\in\Omega_t}\right) &= \sum_{i=1}^{k} P\left(Z=i, \left\{R_j = r_j^{(t)}\right\}_{j\in\Omega_t}\right) \quad \boxed{\textbf{marginalization}} \\
&= \sum_{i=1}^{k} P(Z=i)\, P\left(\left\{R_j = r_j^{(t)}\right\}_{j\in\Omega_t} \middle| Z=i\right) \quad \boxed{\textbf{product rule}} \\
&= \sum_{i=1}^{k} P(Z=i) \prod_{j\in\Omega_t} P\left(R_j = r_j^{(t)} \middle| Z=i\right) \quad \boxed{\textbf{conditional independence}}
\end{aligned}$$

(c) **E-step** *(2 pts)*

$$\begin{aligned}
P\left(Z=i \middle| \left\{R_j = r_j^{(t)}\right\}_{j\in\Omega_t}\right) &= \frac{P(Z=i)\ P\left(\left\{R_j = r_j^{(t)}\right\}_{j\in\Omega_t} \middle| Z=i\right)}{P\left(\left\{R_j = r_j^{(t)}\right\}_{j\in\Omega_t}\right)} \quad \boxed{\textbf{Bayes rule}} \\
&= \frac{P(Z=i)\ \prod_{j\in\Omega_t} P\left(R_j = r_j^{(t)} \middle| Z=i\right)}{\sum_{i'=1}^{k} P(Z=i')\ \prod_{j\in\Omega_t} P\left(R_j = r_j^{(t)} \middle| Z=i'\right)} \quad \boxed{\textbf{substitute from part (b)}}
\end{aligned}$$

(d) **M-step** *(3 pts)*

The visible data $V^{(t)}$ for the $t^{\text{th}}$ student are the observed movie ratings $\left\{R_j = r_j^{(t)}\right\}_{j\in\Omega_t}$.
Thus the EM updates are given by:

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

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

Using the suggested shorthand, we can write these as

$$P(Z=i) \longleftarrow \frac{1}{T}\sum_{t=1}^{T} \rho_{it}, \qquad \boxed{\textbf{+1 pt}}$$

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

Finally consider the terms in the numerator of the second update. If the $t^{\text{th}}$ student saw the $\ell^{\text{th}}$ movie, then $R_\ell$ is an **observed rating**. Thus if $\ell \in \Omega_t$ we have:

$$\begin{aligned} P\left(Z\!=\!i, R_\ell\!=\!1 \,\middle|\, \left\{R_j\!=\!r_j^{(t)}\right\}_{j\in\Omega_t}\right) &= I\left(r_\ell^{(t)},1\right) P\left(Z\!=\!i \,\middle|\, \left\{R_j\!=\!r_j^{(t)}\right\}_{j\in\Omega_t}\right) \\ &= I\left(r_\ell^{(t)},1\right) \rho_{it} \end{aligned}$$

On the other hand, if the $t^{\text{th}}$ student did not see the $\ell^{\text{th}}$ movie, then $R_\ell$ is a **hidden variable**. Thus if $\ell \notin \Omega_t$ we have:

$$\begin{aligned} & P\left(Z\!=\!i, R_\ell\!=\!1 \,\middle|\, \left\{R_j\!=\!r_j^{(t)}\right\}_{j\in\Omega_t}\right) \\ &\quad = P\left(Z\!=\!i \,\middle|\, \left\{R_j\!=\!r_j^{(t)}\right\}_{j\in\Omega_t}\right) P\left(R_\ell\!=\!1 \,\middle|\, Z\!=\!i, \left\{R_j\!=\!r_j^{(t)}\right\}_{j\in\Omega_t}\right) \quad \boxed{\textbf{product rule}} \\ &\quad = P\left(Z\!=\!i \,\middle|\, \left\{R_j\!=\!r_j^{(t)}\right\}_{j\in\Omega_t}\right) P(R_\ell\!=\!1|Z\!=\!i) \quad \boxed{\textbf{conditional independence}} \\ &\quad = \rho_{it}\, P(R_\ell\!=\!1|Z\!=\!i) \end{aligned}$$

Substituting these last two results into the numerator of the second update, we find that

$$P(R_\ell\!=\!1|Z\!=\!i) \longleftarrow \frac{\sum_{\{t|\ell\in\Omega_t\}} \rho_{it}\, I\left(r_\ell^{(t)},1\right) + \sum_{\{t|\ell\notin\Omega_t\}} \rho_{it}\, P(R_\ell\!=\!1|Z\!=\!i)}{\sum_{t=1}^{T} \rho_{it}} \quad \boxed{\textbf{+2 pts}}$$

(e) **Implementation** *(3 pts)*

| iteration | log-likelihood $\mathcal{L}$ |
|---|---|
| 0 | -33.4145 |
| 1 | -21.3919 |
| **2** | **-19.6171** |
| **4** | **-18.7595** |
| **8** | **-18.3982** |
| 16 | -18.2229 |
| **32** | **-18.1025** |
| **64** | **-18.0486** |
| **128** | **-18.0472** |
| **256** | **-18.0471** |

(f) **Personal movie recommendations** *(2 pts)*

(g) **Source code** *(10 pts)*