## CSE 150A/250A. Assignment 7

**Out:** *Tue, May 27*
**Due:** *Mon June 2* (by 11:59 PM, Pacific Time, via gradescope)
**Grace period:** 24 hours

**NOTICE: the final will be in class (WLH 2001) on Thursday, June 5, 2025 during class time 3:30 to 4:50pm.**

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

Consider the Markov decision process (MDP) with two states $s \in \{0, 1\}$, two actions $a \in \{\downarrow, \uparrow\}$, discount factor $\gamma = \frac{1}{2}$, and rewards and transition matrices as shown below:

| $s$ | $R(s)$ |
|---|---|
| 0 | -1 |
| 1 | 2 |

| $s$ | $s'$ | $P(s'\|s, a = \downarrow)$ |
|---|---|---|
| 0 | 0 | $\frac{3}{4}$ |
| 0 | 1 | $\frac{1}{4}$ |
| 1 | 0 | $\frac{1}{4}$ |
| 1 | 1 | $\frac{3}{4}$ |

| $s$ | $s'$ | $P(s'\|s, a = \uparrow)$ |
|---|---|---|
| 0 | 0 | $\frac{1}{2}$ |
| 0 | 1 | $\frac{1}{2}$ |
| 1 | 0 | $\frac{1}{2}$ |
| 1 | 1 | $\frac{1}{2}$ |

(a) **Policy evaluation (4 pts)**

Consider the policy $\pi$ that chooses the action shown in each state. For this policy, solve the linear system of Bellman equations (by hand) to compute the state-value function $V^\pi(s)$ for $s \in \{0, 1\}$. Your answers should complete the following table. (*Hint:* the missing entries are whole numbers.) **Show your work for full credit.**

| $s$ | $\pi(s)$ | $V^\pi(s)$ |
|---|---|---|
| 0 | $\downarrow$ | |
| 1 | $\downarrow$ | |

(b) **Policy improvement (4 pts)**

Compute the greedy policy $\pi'(s)$ with respect to the state-value function $V^\pi(s)$ from part (a). Your answers should complete the following table. **Show your work for full credit.**

| $s$ | $\pi(s)$ | $\pi'(s)$ |
|---|---|---|
| 0 | $\downarrow$ | |
| 1 | $\downarrow$ | |

## 7.2 Three-state MDP (12 pts)

Consider the Markov decision process (MDP) with three states $s \in \{1, 2, 3\}$, two actions $a \in \{\uparrow, \downarrow\}$, discount factor $\gamma = \frac{2}{3}$, and rewards and transition matrices as shown below:

| $s$ | $R(s)$ |
|---|---|
| 1 | -15 |
| 2 | 30 |
| 3 | -25 |

| $s$ | $s'$ | $P(s'\|s, a = \uparrow)$ |
|---|---|---|
| 1 | 1 | $\frac{3}{4}$ |
| 1 | 2 | $\frac{1}{4}$ |
| 1 | 3 | 0 |
| 2 | 1 | $\frac{1}{2}$ |
| 2 | 2 | $\frac{1}{2}$ |
| 2 | 3 | 0 |
| 3 | 1 | 0 |
| 3 | 2 | $\frac{3}{4}$ |
| 3 | 3 | $\frac{1}{4}$ |

| $s$ | $s'$ | $P(s'\|s, a = \downarrow)$ |
|---|---|---|
| 1 | 1 | $\frac{1}{4}$ |
| 1 | 2 | $\frac{3}{4}$ |
| 1 | 3 | 0 |
| 2 | 1 | 0 |
| 2 | 2 | $\frac{1}{2}$ |
| 2 | 3 | $\frac{1}{2}$ |
| 3 | 1 | 0 |
| 3 | 2 | $\frac{1}{4}$ |
| 3 | 3 | $\frac{3}{4}$ |

(a) **Policy evaluation (6 pts)**

Consider the policy $\pi$ that chooses the action shown in each state. For this policy, solve the linear system of Bellman equations (by hand) to compute the state-value function $V^\pi(s)$ for $s \in \{1, 2, 3\}$. Your answers should complete the following table. (*Hint:* the missing entries are whole numbers.) **Show your work for full credit.**

| $s$ | $\pi(s)$ | $V^\pi(s)$ |
|---|---|---|
| 1 | $\uparrow$ | |
| 2 | $\uparrow$ | |
| 3 | $\downarrow$ | |

(b) **Policy improvement (6 pts)**

Compute the greedy policy $\pi'(s)$ with respect to the state-value function $V^\pi(s)$ from part (a). Your answers should complete the following table. **Show your work for full credit.**

| $s$ | $\pi(s)$ | $\pi'(s)$ |
|---|---|---|
| 1 | $\uparrow$ | |
| 2 | $\uparrow$ | |
| 3 | $\downarrow$ | |

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

In this problem you will use the EM algorithm to build a simple movie recommendation system. Download the files *hw7_movies.txt*, *hw7_ids.txt*, and *hw7_ratings.txt*. We also provide *hw7_code_skeleton.py* as a starter. The last of these files contains a matrix of zeros, ones, and missing elements denoted by question marks. The $\langle i, j \rangle^{\text{th}}$ element in this matrix contains the $i^{\text{th}}$ student's rating of the $j^{\text{th}}$ movie, according to the following key:

| | |
|---|---|
| 1 | recommended, |
| 0 | not recommend, |
| ? | not seen. |

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

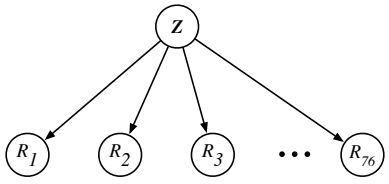
Compute the mean popularity rating of each movie, given by the simple ratio

$$\frac{\text{number of students who recommended the movie}}{\text{number of students who saw the movie}},$$

and sort the movies by this ratio. Print out the movie titles from least popular to most popular along with their mean popularity ratings. Note how well these rankings do or do not corresponding to your individual preferences.

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

Now you will learn a naive Bayes model of these movie ratings, represented by the belief network shown below, with hidden variable $Z \in \{1, 2, \ldots, k\}$ and partially observed binary variables $R_1, R_2, \ldots, R_{60}$ (corresponding to movie ratings).




This model assumes that there are $k$ different types of movie-goers, and that the $i^{\text{th}}$ type of movie-goer—who represents a fraction $P(Z=i)$ of the overall population—likes the $j^{\text{th}}$ movie with conditional probability $P(R_j=1|Z=i)$. Let $\Omega_t$ denote the set of movies seen (and hence rated) by the $t^{\text{th}}$ student. Show that the likelihood of the $t^{\text{th}}$ student's ratings is given by

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

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

The E-step of this model is to compute, for each student, the posterior probability that he or she corresponds to a particular type of movie-goer. Show that

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

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

The M-step of the model is to re-estimate the probabilities $P(Z=i)$ and $P(R_j=1|Z=i)$ that define the CPTs of the belief network. As shorthand, let

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

denote the probabilities computed in the E-step of the algorithm. Also, let $T$ denote the number of students. Show that the EM updates are given by

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

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

Download the files *hw7_probZ_init.txt* and *hw7_probR_init.txt*, and use them to initialize the probabilities $P(Z=i)$ and $P(R_j=1|Z=i)$ for a model with $k=4$ types[1] of movie-goers. Run 256 iterations of the EM algorithm, computing the (normalized) log-likelihood

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

at each iteration. Does your log-likelihood increase (i.e., become less negative) at each iteration? Fill in a completed version of the following table, using the already provided entries to check your work. Note, there will be some small variance across correct implementations. We have reported four significant figures – a precision we have determined to be mostly reproducible. However, if you're getting only three significant figures of agreement, that is not necessarily indicative of a problem.

[1] There is nothing special about these initial values or the choice of $k=4$; feel free to experiment with other choices.

| iteration | log-likelihood $\mathcal{L}$ |
|---|---|
| 0 | -33.4145 |
| 1 | -21.3919 |
| 2 | |
| 4 | |
| 8 | |
| 16 | -18.2229 |
| 32 | |
| 64 | |
| 128 | |
| 256 | |

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

Find your student PID in *hw7_ids.txt* to determine the row of the ratings matrix that stores your personal data. Compute the posterior probability in part (c) for this row from your trained model, and then compute your *expected* ratings on the movies *you haven't yet seen*:

$$P\left(R_\ell = 1 \,\middle|\, \left\{R_j = r_j^{(t)}\right\}_{j \in \Omega_t}\right) \;=\; \sum_{i=1}^{k} P\left(Z = i \,\middle|\, \left\{R_j = r_j^{(t)}\right\}_{j \in \Omega_t}\right) P(R_\ell = 1 | Z = i) \quad \text{for } \ell \notin \Omega_t.$$

Print out the list of these (unseen) movie sorted by their expected ratings. Does this list seem to reflect your personal tastes better than the list in part (a)? Hopefully it does (although our data set is obviously *far* smaller and more incomplete than the data sets at companies like Netflix or Amazon).

*Note:* if you didn't complete the survey in time, then you will need to hard-code your ratings in order to answer this question.

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

Turn in a copy of your source code for all parts of this problem. As usual, you may program in the language of your choice.