ECE59500RL HW1

Robert (Cars) Chandler – chandl71@purdue.edu

Table of contents

Problem 1

1.1

Property 1

Proving the first property of a norm over a vector space for $\|\mathbf{x}\|_\infty$, we first recall that the L-∞ norm of a vector is defined as the element in the vector with the maximum absolute value. As such,

$$\|\mathbf{x}\|_\infty \geq 0, \quad \forall \mathbf{x} \in \mathbb{R}^n$$

since the norm must be an absolute value of some real element, so it must be at least zero:

$$|r| \geq 0, \quad \forall r \in \mathbb{R}$$

Additionally,

$$\|\mathbf{x}\|_\infty = 0 \iff \mathbf{x} = \mathbf{0}$$

Since the only number in $\mathbb{R}$ that has an absolute value of 0 is 0, so for the maximum absolute value (the L-∞ norm) to be 0, all elements of $\mathbf{x}$ must be 0.

Property 2

To prove the second property, we begin with $\| \mathbf{x}\|_\infty$, which is the element of the vector $\alpha\mathbf{x}$ with the maximum absolute value. Each element of $\alpha\mathbf{x}$ is αx_i, and it holds that

$$|\alpha r| = |\alpha||r|, \quad \forall r \in \mathbb{R}$$

So we can say that the maximum absolute value of $\alpha\mathbf{x}$ is equal to the maximum absolute value of $\mathbf{x}$ multiplied by $|\alpha|$:

$$\| \mathbf{x}\|_\infty = |\alpha|\|\mathbf{x}\|_\infty, \quad \forall \mathbf{x} \in \mathbb{R}^n, \alpha \in \mathbb{R}$$

Property 3

To prove the third property, we begin with $\|\mathbf{x}_1 + \mathbf{x}_2\|_\infty$, which is the maximum absolute value of the vector $\mathbf{x}_1 + \mathbf{x}_2$. The maximum absolute value of the element-wise sum of two vectors is maximized when the two elements with the largest magnitude in each vector share the same index:

$$\max\left(\|\mathbf{x}_1 + \mathbf{x}_2\|_\infty\right) = \|\mathbf{x}_1\|_\infty + \|\mathbf{x}_2\|_\infty$$

Since this is the maximum value of the LHS, in any other case, we can say that the sum on the right-hand side must be greater than $\|\mathbf{x}_1 + \mathbf{x}_2\|_\infty$.

So we can say that the maximum absolute value of $\mathbf{x}_1 + \mathbf{x}_2$ is less than or equal to the sum of the maximum absolute values of $\mathbf{x}_1$ and $\mathbf{x}_2$:

$$\|\mathbf{x}_1 + \mathbf{x}_2\|_\infty \leq \|\mathbf{x}_1\|_\infty + \|\mathbf{x}_2\|_\infty, \quad \forall \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$$

1.2

Comparing the L1 and L2 norms, we can square each one to compare them more directly:

$$\|\mathbf{x}\|_2^2 = \left(\sqrt{\sum_{i=1}^{N} |x_i|^2}\right)^2 = \sum_{i=1}^{N} |x_i|^2$$

and

$$\|\mathbf{x}\|_1^2 = \left(\sum_{i=1}^{N} |x_i|\right)^2 = \sum_{i=1}^{N} |x_i|^2 + 2 * \sum_{i<j} |x_i||x_j|$$

We can compare the results of these equations directly, and it is apparent that the L1 norm squared has an additional, nonnegative term added, so it must be greater than or equal to the L2 norm squared:

$$\|\mathbf{x}\|_1^2 = \sum_{i=1}^{N} |x_i|^2 + 2 * \sum_{i<j} |x_i||x_j| \geq \sum_{i=1}^{N} |x_i|^2 = \|\mathbf{x}\|_2^2$$

This implies that

$$\|\mathbf{x}\|_1 \geq \|\mathbf{x}\|_2$$

If we then compare the square L2 norm to the square of the L-∞ norm:

$$\|\mathbf{x}\|_\infty^2 = \left(\max\{|x_1|, ..., |x_n|\}\right)^2 \leq \sum_{i=1}^{N} |x_i|^2 = \|\mathbf{x}\|_2^2$$

since the square of any one element of a vector of positive values must necessarily be less than or equal to a sum of the squares all of those values. This once again implies that

$$\|\mathbf{x}\|_2 \geq \|\mathbf{x}\|_\infty$$

Combining this with the earlier inequality:

$$\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1$$

1.3

Given vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, the absolute value of the inner product of these vectors is:

$$|\langle \mathbf{x}, \mathbf{y} \rangle| = \left| \sum_{i=1}^{N} x_i y_i \right|$$

which, due to the properties of multiplication and absolute value, is also equivalent to

$$\sum_{i=1}^{N} |x_i||y_i|$$

while the product of the L-∞ and L1 norms is:

$$\|\mathbf{x}\|_\infty \|\mathbf{y}\|_1 = \max\{|x_1|, ..., |x_n|\} \sum_{i=1}^{N} |y_i|$$

If we break these down into an index-by-index basis, then for each $i \geq 1 \in \mathbb{N}$, it holds that

$$|x_i||y_i| \leq x_{\text{absmax}}|y_i|$$

and therefore this also holds across the aggregation of these indices into a sum:

$$\sum_{i=1}^{N} |x_i||y_i| \leq \max\{|x_1|, ..., |x_n|\} \sum_{i=1}^{N} |y_i|$$

which is equivalent to

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\|_\infty \|\mathbf{y}\|_1$$

Problem 2

2.1

We can start with the definition of $\mathbb{E}_X[X]$, introduce a marginal sum over Y, turn this into a conditional probability sum, rearrange the sums to form an expectation over Y, and finally form a marginal expectation over $X|Y$:

$$\begin{aligned}
\mathbb{E}_X[X] &= \sum_{x \in \mathcal{X}} x \mathbb{P}(X = x) \\
&= \sum_{x \in \mathcal{X}} x \sum_{y \in \mathcal{Y}} \mathbb{P}(X = x, Y = y) \\
&= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} x \mathbb{P}(X = x, Y = y) \\
&= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} x \mathbb{P}(X = x | Y = y) \mathbb{P}(Y = y) \\
&= \sum_{y \in \mathcal{Y}} \left(\sum_{x \in \mathcal{X}} x \mathbb{P}(X = x | Y = y) \right) \mathbb{P}(Y = y) \\
&= \mathbb{E}_Y \left[\sum_{x \in \mathcal{X}} x \mathbb{P}(X = x | Y) \right] \\
&= \mathbb{E}_Y[\mathbb{E}_X[X|Y]]
\end{aligned}$$

2.2

We can follow the same pattern as before, but with the continuous equivalents of the discrete operations, i.e. integrals instead of sums, pdfs instead of pmfs:

$$
\begin{aligned}
\mathbb{E}_X[X] &= \int_X x f_X(x)\, dx \\
&= \int_X x \left(\int_Y f_{XY}(x,y)\, dy \right) dx \\
&= \int_Y \int_X x f_{XY}(x,y)\, dx\, dy \\
&= \int_Y \int_X x f_{X|Y}(x|y) f_Y(y)\, dx\, dy \\
&= \int_Y \left(\int_X x f_{X|Y}(x|y)\, dx \right) f_Y(y)\, dy \\
&= \mathbb{E}_Y \left[\int_X x f_{X|Y}(x|y)\, dx \right] \\
&= \mathbb{E}_Y[\mathbb{E}_X[X|Y]]
\end{aligned}
$$

2.3

$\mathbb{E}_R\left[R_t | S_t = s, A_t = a\right]$ with a stationary policy $\pi(a|s)$ applied specifies which action will be taken, thus lifting the need to specify $\mathcal{A}$ from the model, so we are left with:

$$
\begin{aligned}
\mathbb{E}_R\left[R_t | S_t = s\right] &= \mathbb{E}_{\mathcal{A}}\left[\mathbb{E}_R\left[R_t | A_t = a, S_t = s\right] | S_t = s\right] \\
&= \mathbb{E}_{\mathcal{A}}\left[\bar{R}(s,a) | S_t = s\right] \\
&= \sum_{a \in \mathcal{A}} \pi(a|s) \bar{R}(s,a)
\end{aligned}
$$

Problem 3

3.1

The state space is

$$\mathcal{S} = \{\text{Country}, \text{Jazz}, \text{Rock}\}$$

The initial distribution is the following discrete probability distribution vector, where the indices are in the order Country, Jazz, Rock:

$$\mu_0 = \{0, 1, 0\}$$

The transition matrix, with rows and columns in the order Country, Jazz, Rock, is:

$$P = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0.3 & 0.1 & 0.6 \\ 0.2 & 0 & 0.8 \end{bmatrix}$$

3.2

Let C, J, R be the states corresponding to the respective genres. The probability of starting with Jazz each time is 1, so we just represent our initial state in probability space with a 1 and we multiply the probability of each transition after that.

3.2.a

The probability of transitioning from Country to Country $\mathbb{P}(C|C) = 0$, so the probability of the sequence is 0.

3.2.b

$$1 \cdot \mathbb{P}(J|R)\mathbb{P}(R|R)\mathbb{P}(C|R)\mathbb{P}(J|C) = 1 \cdot 0.6 \cdot 0.8 \cdot 0.2 \cdot 0.5 = 0.048$$

3.2.c

$$\begin{aligned} & 1 \cdot \mathbb{P}(J|R)\mathbb{P}(C|R)\mathbb{P}(R|C)\mathbb{P}(C|R) \ldots \mathbb{P}(R|C)\mathbb{P}(C|R) \\ =& \mathbb{P}(J|R) \prod_{i=1}^{m} \mathbb{P}(C|R) \prod_{i=1}^{m-1} \mathbb{P}(R|C) \\ =& \mathbb{P}(J|R)\mathbb{P}(C|R)^m \mathbb{P}(R|C)^{m-1} \\ =& 0.6 \cdot 0.2^m 0.5^{m-1} \end{aligned}$$

3.2.d

As the final product repeats more and more as m approaches infinity, the product will approach 0 since the fractions raised to the m and $m-1$ power will both approach 0 since they are less than one and raised to an infinite power.

3.3

We can find this using the transition matrix raised to the $t = 2$ power:

```
import numpy as np

mu0 = np.array([0, 1, 0])
trans_matrix = np.array([
  [0 , 0.5 , 0.5],
  [0.3 , 0.1 , 0.6],
  [0.2 , 0 , 0.8 ],
])

mu2 = mu0 @ np.linalg.matrix_power(trans_matrix, 2)
```

$$\mu_0 P^2 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0.3 & 0.1 & 0.6 \\ 0.2 & 0 & 0.8 \end{bmatrix}^2 = [0.15, 0.16, 0.69]$$

Where the entries in the vector correspond to the probability of playing Country, Jazz, and Rock at $t = 2$, respectively.

3.4

We can calculate the steady-state distribution $\bar{\mu}$ using a similar code as above, repeatedly applying the transition matrix until some convergence threshold is met. We will set the steady-state condition to be reached when a difference of less than 1e-8 is reached for all values in the distribution:

```
state_distribution = mu0
t = 0
difference_criterion = 1e-8

while(True):
  t += 1
  old_state_distribution = state_distribution.copy()
  state_distribution = state_distribution @ trans_matrix

  if np.all(np.abs(state_distribution - old_state_distribution) < difference_criterion):
    break
```

The steady-state distribution is approximately

$$\bar{\mu} = [0.17475728268372312, 0.09708737729711539, 0.7281553400191619]$$

and the convergence criterion is met at time $t = 17$.

3.5

We first re-define our stationary distribution since the new requirement is to extend to 100 timesteps, which is further than the previous convergence condition reached. We set it to be the distribution at $t = 110$ since this is past the final value required for plotting and the distribution should be effectively unchanging at this point.

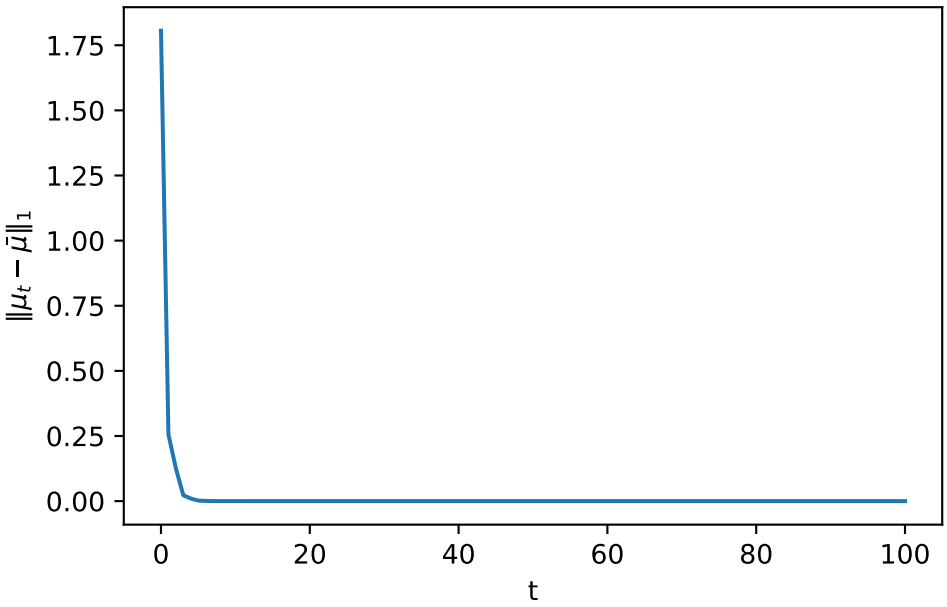
```
import matplotlib.pyplot as plt

steady_state = mu0 @ np.linalg.matrix_power(trans_matrix, 110)

trans_matrices = np.array([np.linalg.matrix_power(trans_matrix, i) for i in range(101)])

l1_norms = np.linalg.norm(mu0 @ trans_matrices - steady_state, ord=1, axis=1)

fig, ax = plt.subplots()
ax.plot(l1_norms)
ax.set_xlabel("t")
ax.set_ylabel(r"$\Vert \mu_t - \bar{\mu} \Vert_1$");
```



3.6

Based on the steady-state distribution, we expect Rock to be played most often, because given any sufficiently large t-value, the probability of the state being Rock is fixed and is much higher than the other two states. This makes sense given that Rock has such a high probability of transitioning to itself.

Problem 4

4.1

We can prove this by induction.

The *induction hypothesis* is:

$$\begin{aligned}&\mathbb{P}[X_{t+1}, X_{t+2}, \dots, X_{t+m} | X_0, X_1, \dots, X_t] \\ =&\mathbb{P}[X_{t+1}, X_{t+2}, \dots, X_{t+m} | X_t], \quad \forall t \in \mathbb{N}_0, \forall m \in \mathbb{N}\end{aligned}$$

We assume this hypothesis to be correct for some value $m \in \mathbb{N}$. Given that the MC satisfies the Markov property, the following *base case* holds for $m = 1$:

$$\mathbb{P}[X_{t+1} | X_0, X_1, \dots, X_t] = \mathbb{P}[X_{t+1} | X_t]$$

In the *induction step*, we will show that the hypothesis is also correct for $m+1$:

$$\begin{aligned}&\mathbb{P}[X_{t+1}, X_{t+2}, \dots, X_{t+m}, X_{t+m+1} | X_0, X_1, \dots, X_t] \\ =&\mathbb{P}[X_{t+1}, X_{t+2}, \dots, X_{t+m}, X_{t+m+1} | X_t], \quad \forall t \in \mathbb{N}_0, \forall m \in \mathbb{N}\end{aligned}$$

We can start with the LHS of this equation and use the definition of conditional probability (or the "chain rule" or "product rule" if you prefer) to write it as the product of two conditional probabilities, "pulling out" the last event X_{t+m+1}. If we take the form $P(A, B) = P(B|A)P(A)$, then for our scenario $A = X_{t+1}, ..., X_{t+m}$ and $B = X_{t+m+1}$ and $P = P(\cdot | X_0, ..., X_t)$:

$$\begin{aligned}&\mathbb{P}[X_{t+1}, X_{t+2}, \dots, X_{t+m}, X_{t+m+1} | X_0, X_1, \dots, X_t] \\ =&\mathbb{P}[X_{t+m+1} | X_0, X_1, \dots, X_t, X_{t+1}, \dots, X_{t+m}] \cdot \mathbb{P}[X_{t+1}, X_{t+2}, \dots, X_{t+m} | X_0, X_1, \dots, X_t] \\ =&\mathbb{P}[X_{t+m+1} | X_0, X_1, \dots, X_{t+m}] \cdot \mathbb{P}[X_{t+1}, X_{t+2}, \dots, X_{t+m} | X_0, X_1, \dots, X_t]\end{aligned}$$

We can use the induction hypothesis, which is already assumed to be true, to rewrite the second term:

$$\begin{aligned}&\mathbb{P}[X_{t+m+1} | X_0, X_1, \dots, X_{t+m}] \cdot \mathbb{P}[X_{t+1}, X_{t+2}, \dots, X_{t+m} | X_0, X_1, \dots, X_t] \\ =&\mathbb{P}[X_{t+m+1} | X_0, X_1, \dots, X_{t+m}] \cdot \mathbb{P}[X_{t+1}, X_{t+2}, \dots, X_{t+m} | X_t]\end{aligned}$$

We want to use the definition of conditional probability once more to pull these two terms back together into a single conditional probability, but the condition of the first term is not quite right. However, because of the Markov property, it is true that:

$$\mathbb{P}[X_{t+m+1} | X_0, X_1 \dots, X_{t+m}] = \mathbb{P}[X_{t+m+1} | X_{t+m}]$$

and it is also true that

$$\mathbb{P}[X_{t+m+1} | X_t, X_{t+1} \dots, X_{t+m}] = \mathbb{P}[X_{t+m+1} | X_{t+m}]$$

so we can equate these two by the transitive property:

$$\mathbb{P}[X_{t+m+1} | X_0, X_1 \dots, X_{t+m}] = \mathbb{P}[X_{t+m+1} | X_t, X_{t+1} \dots, X_{t+m}]$$

So we can substitute this term in the original equation and use the definition of conditional probability to join the two terms back together:

$$
\begin{aligned}
&\mathbb{P}[X_{t+m+1}|X_0, X_1, \dots, X_{t+m}] \cdot \mathbb{P}[X_{t+1}, X_{t+2}, \dots, X_{t+m}|X_t] \\
=&\mathbb{P}[X_{t+m+1}|X_t, X_{t+1}, \dots, X_{t+m}] \cdot \mathbb{P}[X_{t+1}, X_{t+2}, \dots, X_{t+m}|X_t] \\
=&\frac{\mathbb{P}[X_t, X_{t+1}, \dots, X_{t+m+1}]}{X_t, X_{t+1}, \dots, X_{t+m}} \frac{X_t, X_{t+1}, \dots, X_{t+m}}{X_t} \\
=&\frac{\mathbb{P}[X_t, X_{t+1}, \dots, X_{t+m+1}]}{X_t} \\
=&\mathbb{P}[X_{t+1}, X_{t+2}, \dots, X_{t+m+1}|X_t]
\end{aligned}
$$

So, we have proven the induction step, therefore the hypothesis holds for all $m \in \mathbb{N}$, and the induction hypothesis is proven:

$$
\begin{aligned}
&\mathbb{P}[X_{t+1}, X_{t+2}, \dots, X_{t+m}|X_0, X_1, \dots, X_t] \\
=&\mathbb{P}[X_{t+1}, X_{t+2}, \dots, X_{t+m}|X_t], \quad \forall t \in \mathbb{N}_0, \forall m \in \mathbb{N} \quad \blacksquare
\end{aligned}
$$

4.2

We wish to prove that the following holds:

$$
\begin{aligned}
&\mathbb{P}[X_{t+k}, X_{t+k+1}, \dots, X_{t+m}|X_0, X_1, \dots, X_t] \\
=&\mathbb{P}[X_{t+k}, X_{t+k+1}, \dots, X_{t+m}|X_t], \quad \forall t \in \mathbb{N}_0, \forall k, m \in \mathbb{N}, m \geq k
\end{aligned}
$$

In the case where $k = 1$, the equation is equivalent to the one we proved in 4.1, so it holds.

In all other cases where $1 < k \leq m$, the probability of the next state X_{t+k} is dependent only on the state before it X_{t+k-1} due to the Markov property. However, if $1 < k \leq m$, then the previous state X_{t+k-1} cannot be equal to X_t. This is to say that if X_{t+k} is the next "future" state, then it is only dependent on X_{t+k-1}, which is the "current" state, but X_t is necessarily a "past" state. X_t is the most current state on which the conditional probability in the equation is defined, so all we have is past information, which is all the same in that it does not inform us about the probability of the next state. So for all valid k other than $k = 1$, the probability of future states conditioned on past information from $t = 0$ to $t = t$ is equivalent to the probability of those future states conditioned only on X_t since they are all past information and therefore are all equally uninformative. So it holds that

$$
\begin{aligned}
&\mathbb{P}[X_{t+k}, X_{t+k+1}, \dots, X_{t+m}|X_0, X_1, \dots, X_t] \\
=&\mathbb{P}[X_{t+k}, X_{t+k+1}, \dots, X_{t+m}|X_t], \quad \forall t \in \mathbb{N}_0, \forall k, m \in \mathbb{N}, m \geq k
\end{aligned}
$$

Problem 5

5.1

We assume that the state (i, j) represents the previous two states in X_{t-2}, X_{t-1} order, such that a sequence $2, 1, 2$ in the second-order model can be represented by a sequence $(2, 1), (1, 2)$ in the simple MC.

Using the transition from $(1, 1)$ to $(1, 1)$ as an example, the probability of this transition is found from the given probability:

$$\mathcal{P}((1,1)|(1,1)) = \mathcal{P}(X_t = 1|X_{t-2} = 1, X_{t-1} = 1) = 0.8$$

and we can use the complement to define the probability of $(1, 1)$ to $(1, 2)$:

$$\begin{aligned}\mathcal{P}((1,2)|(1,1)) &= \mathcal{P}(X_t = 2|X_{t-2} = 1, X_{t-1} = 1) \\ &= 1 - \mathcal{P}(X_t = 1|X_{t-2} = 1, X_{t-1} = 1) \\ &= 0.2\end{aligned}$$

We repeat this across all combinations:

$$\begin{aligned}\mathcal{P}((1,1)|(1,1)) &= \mathcal{P}(X_t = 1|X_{t-2} = 1, X_{t-1} = 1) = 0.8 \\ \mathcal{P}((1,2)|(1,1)) &= 1 - \mathcal{P}(X_t = 2|X_{t-2} = 1, X_{t-1} = 1) = 0.2 \\ \mathcal{P}((2,1)|(1,2)) &= \mathcal{P}(X_t = 1|X_{t-2} = 1, X_{t-1} = 2) = 0.1 \\ \mathcal{P}((2,2)|(1,2)) &= 1 - \mathcal{P}(X_t = 2|X_{t-2} = 1, X_{t-1} = 1) = 0.9 \\ \mathcal{P}((1,1)|(2,1)) &= \mathcal{P}(X_t = 1|X_{t-2} = 2, X_{t-1} = 1) = 0.3 \\ \mathcal{P}((1,2)|(2,1)) &= 1 - \mathcal{P}(X_t = 2|X_{t-2} = 2, X_{t-1} = 1) = 0.7 \\ \mathcal{P}((2,1)|(2,2)) &= \mathcal{P}(X_t = 1|X_{t-2} = 2, X_{t-1} = 2) = 0.7 \\ \mathcal{P}((2,2)|(2,2)) &= 1 - \mathcal{P}(X_t = 2|X_{t-2} = 2, X_{t-1} = 2) = 0.3\end{aligned}$$

Which, in graph form, looks like:

To find the initial state distribution, we take the probabilities of getting all $2 \cdot 2 = 4$ possibilities of the first two states using the probabilities in the problem statement.

The first-order initial distribution is:

$$\begin{aligned}\mathbb{P}_1(X_0 = 1) &= 0.5 \\ \mathbb{P}_1(X_0 = 2) &= 0.5\end{aligned}$$

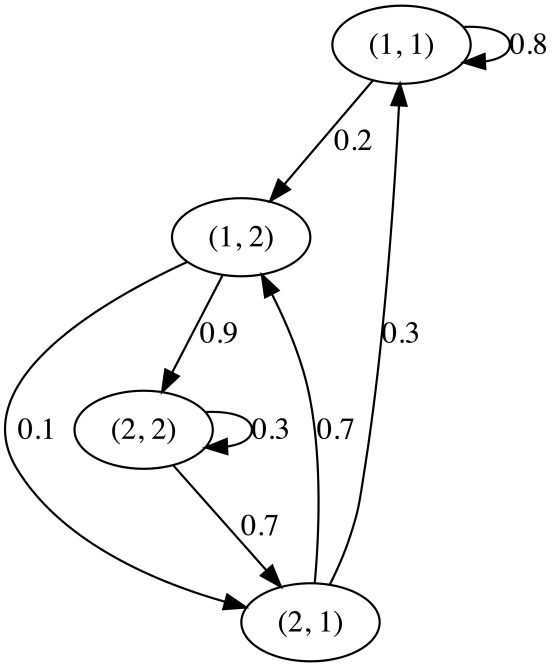


Figure 1: Model Transition Graph

Where $\mathbb{P}_1$ indicates that the probability measure is acting on the first-order model's state space.

Using this, the second-order initial distribution is:

$$\begin{aligned}
\mathbb{P}_2(X_0 = (1,1)) &= \mathbb{P}_1(X_1 = 1 | X_0 = 1)\mathbb{P}_1(X_0 = 1) = 0.2 \cdot 0.5 = 0.1 \\
\mathbb{P}_2(X_0 = (2,1)) &= \mathbb{P}_1(X_1 = 1 | X_0 = 2)\mathbb{P}_1(X_0 = 2) = 0.4 \cdot 0.5 = 0.2 \\
\mathbb{P}_2(X_0 = (1,2)) &= \mathbb{P}_1(X_1 = 2 | X_0 = 1)\mathbb{P}_1(X_0 = 1) = 0.8 \cdot 0.5 = 0.4 \\
\mathbb{P}_2(X_0 = (2,2)) &= \mathbb{P}_1(X_1 = 2 | X_0 = 2)\mathbb{P}_1(X_0 = 2) = 0.6 \cdot 0.5 = 0.3
\end{aligned}$$

Where $\mathbb{P}_2$ indicates that the probability measure is acting on the second-order model's state space.

Therefore:

$$\mu_0 = [0.1, 0.2, 0.4, 0.3]$$

for the state-space

$$\mathcal{S} = \{(1,1), (1,2), (2,1), (2,2)\}$$

5.2

To create a simple Markov chain from an MC of order k, we first need to form a new state-space $\mathcal{S}'$ from the original space $\mathcal{S}$ where each element of $\mathcal{S}'$ is a k-tuple such that all $|\mathcal{S}|^k$ combinations of k values from $\mathcal{S}$ are included. The values in the tuple represent a sequence of k states in the k-order space from oldest to newest.

To form the new state transition probabilities, we take an existing k-tuple from $\mathcal{S}'$ and consider a state from $\mathcal{S}$ as the next state in the sequence. We require each of the state values in the k-tuple sequence to determine the probability of the next state. This probability should be given as a conditional probability in the original model, conditioned on the k previous values. We assign the probability whose condition matches the order of the sequence in the tuple. We repeat this for each tuple in $\mathcal{S}'$.

To form the initial distribution, we consider a tuple from $\mathcal{S}'$ and multiply out the probabilities given in the original model of obtaining each step such that the first probability is not conditioned on any former states, and each successive probability will be conditional on one additional previous state. For example, if the tuple is (S_0, S_0, S_1), the first term in the product should be $\mathbb{P}(X_0 = S_0)$, the the probability of obtaining S_0 on the first step. This will be multiplied by $\mathbb{P}(X_1 = S_0 | X_0 = S_0)$, and will finally be multiplied by $\mathbb{P}(X_2 = S_1 | X_0 = S_0, X_1 = S_0)$. This process is repeated for each tuple in $\mathcal{S}'$.

Problem 6

6.1

The scenario can be described as an MDP with reward defined by the tuple $\mu = (\mu_0, \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma)$

Additionally, we define the following "helper" sets and a function to map between the movies available and the genres of those movies. We could describe all elements of the MDP without these by writing out all possible combinations, but these help to minimize the amount of writing necessary:

$$G = \{A, H, C\}$$
$$M = \{A, B, C, D\}$$
$$\mathcal{G} : M \to G^n, \quad n \in \{1, 2, 3\}$$

G is the set of genres available, M is the set of movies available, and $\mathcal{G}$ describes which genres apply for a given movie.

$\mathcal{S} = \{s_A, s_H, s_C\}$ where each state corresponds to the user's desired genre, which is either Action, Horror, or Comedy, respectively.

$\mu_0 = [\frac{1}{3}, \frac{1}{3}, \frac{1}{3}]$ where the order of the indices is Action, Horror, Comedy; the same as for $\mathcal{S}$.

$\mathcal{A} = \{a_A, a_B, a_C, a_D\}$ which are actions corresponding to which movie the system recommends to the user.

$\mathcal{P} = \mathcal{P}(s, a) : \mathcal{S} \times \mathcal{A} \to \Delta(\mathcal{S})$ where the function can be described piecewise as:

$$\mathcal{P}(s,a) = \begin{cases} [1,0,0], s = s_A, A \notin \mathcal{G}(a) \\ [0,1,0], s = s_H, H \notin \mathcal{G}(a) \\ [0,0,1], s = s_C, C \notin \mathcal{G}(a) \\ [0,0.5,0.5], s = s_A, A \in \mathcal{G}(a) \\ [0.5,0,0.5], s = s_H, H \in \mathcal{G}(a) \\ [0.5,0.5,0], s = s_C, C \in \mathcal{G}(a) \end{cases}$$

When the user's genre preference is not one of the genres of the movie chosen by the previous action, the user's genre preference remains the same with probability 1. If one of the movie's genres is the user's preference, then the next genre preference will be one of the other two genres with equal probability (0.5) for each genre.

$\mathcal{R} = \mathcal{R}(s, a) : \mathcal{S} \times \mathcal{A} \to [-1, 1]$

Where the reward function can be defined piecewise as:

$$\mathcal{R}(s_i, a_j) = \begin{cases} 1, & i \in \mathcal{G}(j), i \in G, j \in M \\ -1, & i \notin \mathcal{G}(j), i \in G, j \in M \end{cases}$$

Which is to say that if the genre preference defined by the input state s_i (where i is the genre in the set of genres G) is in the set of genres of the movie chosen in action a_j (which is the output of the function $\mathcal{G}(j)$ where j is the movie in the set of movies M), then the reward will be 1, which corresponds to the user submitting a Like, and if the desired genre is *not* in the set of genres for the movie chosen, the reward is -1, corrresponding to the user submitting a Dislike.

$\gamma = 0.95$

6.2

The transition-based graph of the recommendation strategy is shown below where the nodes correspond to the different states and the edges are labeled with the actions taken in the transitions they represent along with the probability of taking that respective action/transition:

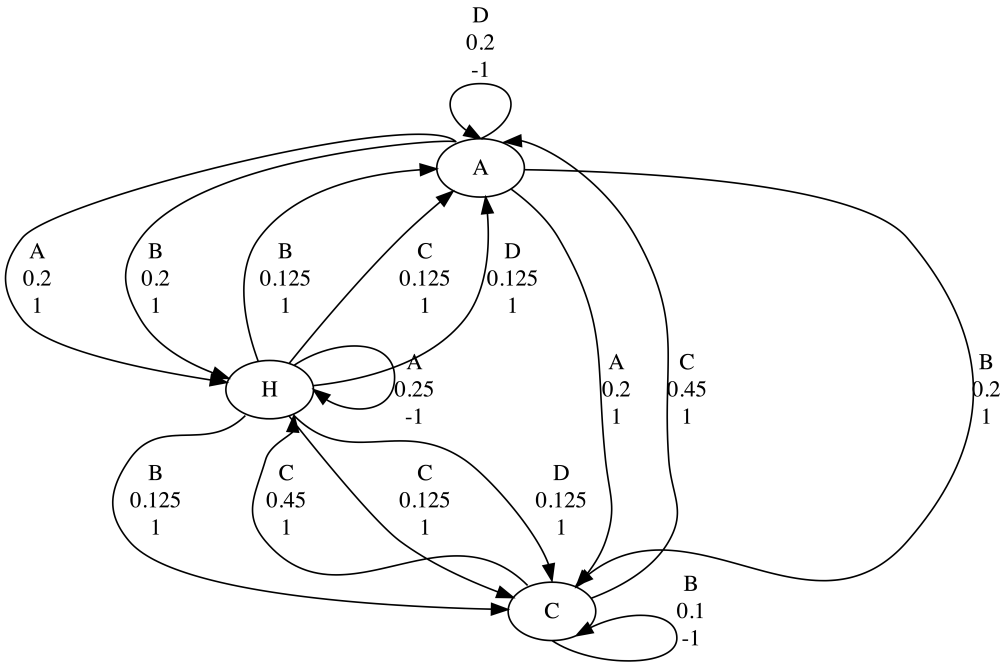


Figure 2: Model Transition Graph

Probability Calculations

Each probability on the edges was calculated by taking the action and its corresponding probability given by the policy, determining whether the combination of the action and the state would match the user's preference or not, and if it did match their preference, splitting the probability into two equal parts and assigning it to the two other genres, while if it did match the user's preference, assigning the full probability to the same genre as given in the current state. For example, if the user asks for Action movies and there is a 0.4 probability that the system recommends movie A, which is an Action movie, then if the system does recommend A, the user would have a 50% chance of selecting either Horror or Comedy as the next state, so we take the 0.4 probability, multiply it by 0.5, and assign that probability to each of the transitions from $A \rightarrow H$ and $A \rightarrow C$, marking which action was taken to get there.

The given policy is stationary as it only depends on the user's current genre preference rather than their past or future ones. It is also a stochastic policy rather than a deterministic one as the action taken is chosen with some degree of randomness as indicated by the probability values on the edges/transitions in the graph.

6.3

With the policy in 6.2 prescribed, a Markov chain (with reward) is induced over the MDP from 6.1. It is defined by the tuple $\mu = (\mu_0, \mathcal{S}, \mathcal{P}, \mathcal{R}, \gamma)$.

We can simplify the transition-based graph in Figure 2 by removing the specification on which action is taken and summing the remaining edges which make the same transition. This shows how inducing the policy removes the need to specify $\mathcal{A}$ in the model definition: now that a specific set of actions is prescribed for each state, even if we don't know which exact action will be taken, we can describe the transition from state-to-state via a probability that does not depend on the exact action taken, but rather the total probability of moving from one state to another across all actions that lead to the given transition.

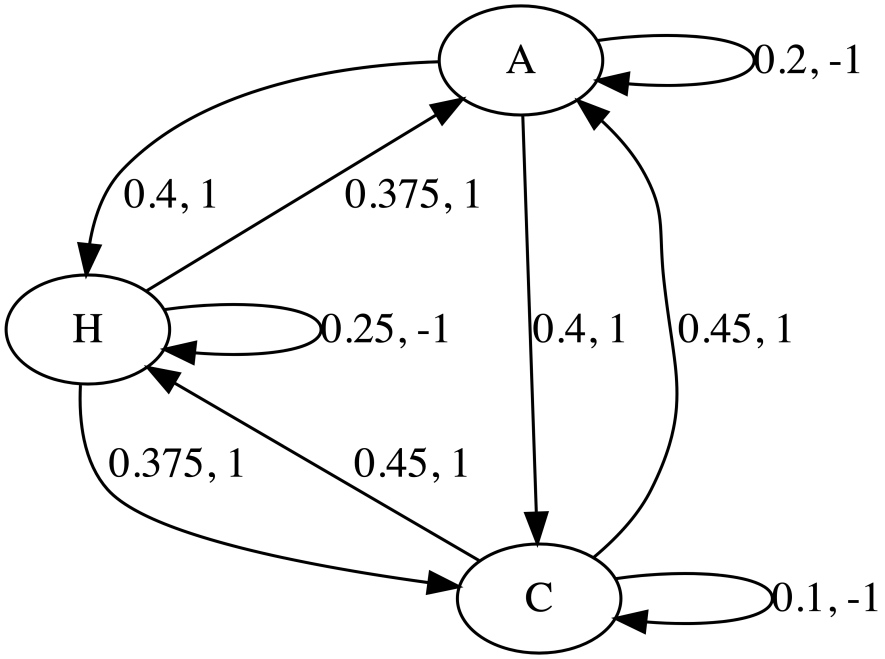


Figure 3: Simplified Model Transition Graph with Actions Removed

Using the same helper sets and function from 6.1, we can define each element of the model:

$\mathcal{S} = \{s_A, s_H, s_C\}$ where each state corresponds to the user's desired genre, which is either Action, Horror, or Comedy, respectively.

$\mu_0 = [\frac{1}{3}, \frac{1}{3}, \frac{1}{3}]$ where the order of the indices is Action, Horror, Comedy.

$\mathcal{P} = \mathcal{P}(s) : \mathcal{S} \to \Delta(\mathcal{S})$ where the function can be described by the transition matrix as:

$$\mathcal{P}(s) = \begin{bmatrix} 0.2 & 0.4 & 0.4 \\ 0.375 & 0.25 & 0.375 \\ 0.45 & 0.45 & 0.1 \end{bmatrix}$$

Where the rows and columns follow the same Action, Horror, Comedy order as usual. The values of this matrix are found in the graph in Figure 3 above, which were calculated as described in Section .

$\mathcal{R} = \mathcal{R}(s) : \mathcal{S} \to [-1, 1]$

Where the reward function can be defined piecewise as:

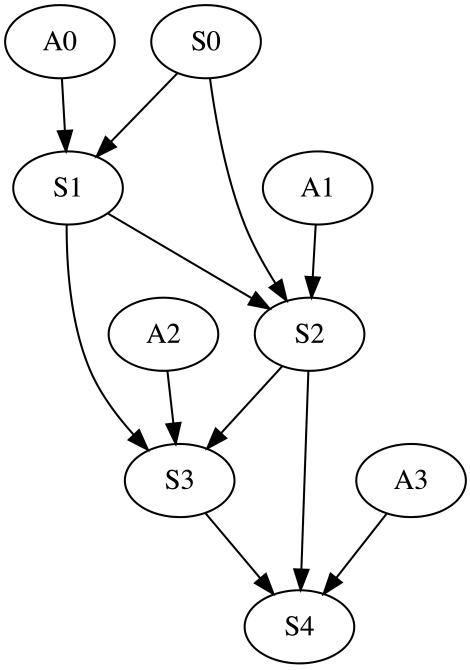
$$\mathcal{R}(s_t) = \begin{cases} 1, & s_t \\ 0, & i \notin \mathcal{G}(j), i \in G, j \in M \end{cases}$$

Which is to say that if the genre preference defined by the input state s_i (where i is the genre in the set of genres G) is in the set of genres of the movie chosen in action a_j (which is the output of the function $\mathcal{G}(j)$ where j is the movie in the set of movies M), then the reward will be 1, which corresponds to the user submitting a Like, and if the desired genre is *not* in the set of genres for the movie chosen, the reward is -1, corresponding to the user submitting a Dislike.

$\gamma = 0.95$

6.4

The first five timesteps are represented below:


A0
S0
S1
A1
A2
S2
S3
A3
S4