תעודת זהות:

מספר מחברת:

1	2	3	4	5	6	7	סהב

<u>מבחן מועד א' – למידה מחיזוקים</u>
<u>סמסטר ב' תשע"ח (2018/9)</u>

בית הספר למדעי המחשב, אוניברסיטת תל-אביב

מרצה: פרופ' ישי מנצור

מתרגלת: גב' לי כהן

9.7.2019

<u>הוראות</u>

1. מומלץ לקרוא את כל ההנחיות והשאלות בתחילת המבחן, לפני תחילת כתיבת התשובות.
2. משך הבחינה – **שלוש שעות**. לא תינתן כל הארכה נוספת.
3. חומר עזר מותר: דף עזר אחד בגודל A4 דו-צדדי , **<u>ומחשבון</u>**.
4. **יש לענות על השאלות במקום המיועד לכך בטופס השאלון (טופס זה).** מחברות הבחינה לא ייקראו, וישמשו כטיוטה בלבד.
5. יש למלא בכל דף של השאלון מספר ת.ז. ומספר מחברת.
6. במבחן 4 שאלות:
   - הניקוד לכל שאלה מופיע לידה מספר השאלה.
   - יש לענות תשובות ברורות ענייניות ותמציתיות.
7. מותר להשתמש בכל טענה שהוכחה בכיתה (בהרצאה, בתרגול, או בתרגיל בית) בתנאי שמצטטים אותה במדויק. טענות אחרות (כאלה שהוכחו בספרים, בהרצאות מהסמסטר הקודם, וכו') יש להוכיח.

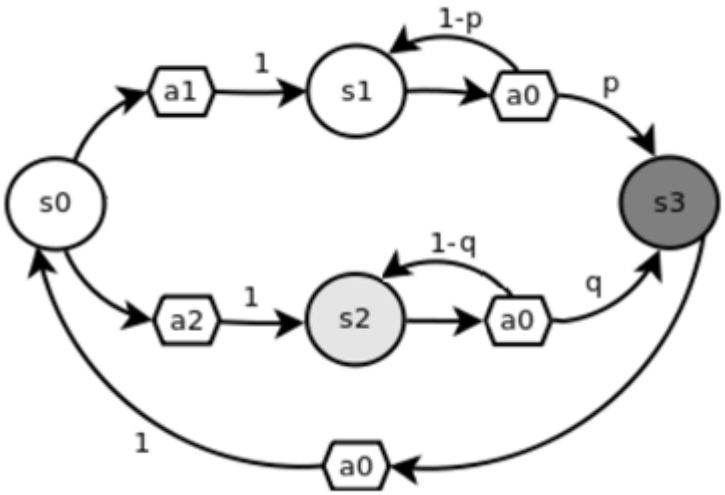
בהצלחה!

תעודת זהות:

מספר מחברת:

שאלה 1 (30 נקודות)

נתון MDP שמוגדר ע"י הגרף שבציור, כאשר המצבים מסומנים במעגלים והמצב ההתחלתי הוא s_0.
ישנן שלוש פעולות המתוארות באמצעות משושים. המספרים מעל החצים המחברים את הפעולות למצבים מתארים את ההסתברות למעבר מהמצב הנוכחי באמצעות הפעולה הזו למצב הבא.
לדוגמא, ההסתברות לעבור למצב s_3 ממצב s_2 באמצעות a_0 הינה q. במידה ואין חץ בין פעולה למצב, ההסתברות לעבור למצב הזה הינה 0. הרווח המיידי ממצב s_3 הוא 10, ממצב s_2 הוא 1 ו0 מהשאר.
ההחזר (return) הוא discounted עם פרמטר γ.



א. כתוב/כתבי את ה- MDP של הבעיה באופן פורמלי.

S =
A =
s_0 =
R:
P:

ב. כתוב/כתבי את משוואות בלמן לערך האופטימלי לכל מצב ($V^*(s)$) במונחים של γ, p, q ו- $V^*(s')$.

תעודת זהות:

מספר מחברת:

ג. מעתה נניח ש $q > p$. מתי מתקיים $V^*(s_2) > V^*(s_1)$?
רמז: התחילו מלהביע את $V^*(s_1), V^*(s_2)$ באמצעות γ, p, q ו-$V^*(s_3)$ בלבד.

תשובה:
נמק:

ד. האם קיימים q, γ שעבורם $\pi^*(s_0) = a_1$? (זכרו ש$q > p$).

תשובה: כן/ לא

נמק:

תעודת זהות:

מספר מחברת:

ה. חשב/י לכל מצב את $V^*(s)$ (כפונקציה של $p, q, \gamma, V^*(s_3)$)

פתרו את $V^*(s_3)$ (כלומר הבע אותו באמצעות p, q, γ בלבד)

תעודת זהות:

מספר מחברת:

שאלה 2 (30 נקודות)

בשאלה זאת נדון במודל ה-MAB. במודל ישנם $n = 2k + 1$ פעולות.
הרווח של פעולה a_i נדגם ממשתנה ברנולי עם פרמטר q_i (בהסתברות q_i זה 1, ואחרת 0)
כל הפרמטרים q_i לא ידועים שונים, וכן $|q_i - q_j| \geq \Delta$ לכל שתי פעולות $a_i \neq a_j$.
פעולה a_i נקראת **<u>החציון</u>** אם ישנן k פעולות a_j עם $q_j < q_i$ וכן k פעולות a_j עם $q_j > q_i$

א. בהינתן דגימה של n_i רווחים מפעולה a_i, מתוכם s_i הם 1 (והשאר 0),
הגדר/י UCB ו- LCB כפונקציה של n_i, s_i, δ עבור פעולה a_i כך
שבהסתברות $1 - \delta$ לפחות יתקיים $q_i \in [LCB(n_i, s_i), UCB(n_i, s_i)]$.

$\boldsymbol{UCB(n_i, s_i)}$ =

$\boldsymbol{LCB(n_i, s_i)}$ =

מנקודה זאת והלאה, הניחו שה-UCB וה-LCB תמיד מתקיימים, וענה על הסעיפים הבאים

ב. כתוב/י תנאי מספיק לכך שפעולה a_i היא החציון (כתלות ב-$UCB(n_i, s_i)$ ו-$LCB(n_i, s_i)$ של כל הפעולות). הנח שתמיד מתקיים $q_i \in [LCB(n_i, s_i), UCB(n_i, s_i)]$.

תעודת זהות:

מספר מחברת:

ג. כתוב/י תנאי מספיק (טוב ביותר) לכך שפעולה a_i אינה החציון (כתלות ב-$UCB(n_i, s_i)$ ו-$LCB(n_i, s_i)$ של כל הפעולות). הנח שתמיד מתקיים $q_i \in [LCB(n_i, s_i), UCB(n_i, s_i)]$.
אם יש מספר תנאים מספיקים, העדף זה שפוסל הכי הרבה.

ד. שנה את אלגוריתם ה- Successive Elimination כך שיחזיר את הפעולה שהיא חציון (האלגוריתם לא יודע את הפרמטר Δ). הדגש איזה שורה/ות משתנים, ואיך!

Algorithm Successive Elimination

1: Initially $S = A$

2: While $|S| \geq 2$:

3: Try each action $a_i \in S$ once, and update n_i and s_i

4: For each $a_i \in S$, such that $\exists a_j \in S: UCB(n_i, s_i) < LCB(n_j, s_j)$ THEN $S \leftarrow S - \{a_i\}$

5: END WHILE

6: Return S

תעודת זהות:

מספר מחברת:

שאלה 3 (20 נקודות)

התפלגות Weibull מוגדרת עם פרמטרים λ, $0 < k$, כאשר פונקציית הצפיפות היא $\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1}\exp(-\left(\frac{x}{\lambda}\right)^{k})$ לכל $x \geq 0$.

משתמשים בהתפלגות הנ"ל על מנת לדגום פעולות באופן הבא:

לכל מצב s יש וקטור $\phi(s) \in \mathbb{R}^d$ שמאפיין אותו.

מדיניות π מאופיינת על ידי שני וקטורים $\theta_\lambda, \theta_k \in \mathbb{R}^d$.

במצב s דוגמים פעולה $a \geq 0$ לפי התפלגות המוגדרת לעיל עם פרמטרים:

$\lambda = \exp(\theta_\lambda^\top \phi(s)), k = \theta_k^\top \phi(s)$.

א. מה ההסתברות שנדגום במצב s פעולה a כך ש$a \in [0,1]$ כפונקציה של θ_λ, θ_k ו- $\phi(s)$?

ב. חשב/י את $\nabla_\theta \log \pi(a|s;\theta)$.

תעודת זהות:

מספר מחברת:

ג. כתוב/י את העדכון של REINFORCE עבור ההתפלגות Weibull:

תעודת זהות:

מספר מחברת:

שאלה 4 (20 נקודות)

נתון MDP M המוגדר ע"י (S, A, p, s_0, R), ומדיניות π

לכל מצב s נגדיר וקטור יחידה e_s כך ש- $e_s(s) = 1$ ו- $\forall s' \neq s: e_s[s'] = 0$

לכל פעולה a נגדיר מטריצה $P^a[s_i, s_j] = p(s_i|s_j, a)$

למדיניות π נגדיר וקטור רווחים $r[s] = r(s, \pi(s))$ ומטריצה סטוכסטית $P^\pi[s_i, s_j] = p(s_i|s_i, \pi(s_i))$

מגדירים את האופרטורים הבאים עבור $V \in \mathbb{R}^{|S|}$ ו- $Q \in \mathbb{R}^{|S|\cdot|A|}$ בהתאמה:

(1) $\left(R^{(3)}V\right)(s) = r^\top e_s + \gamma r^\top P^\pi e_s + \gamma^2 r^\top (P^\pi)^2 e_s + \gamma^3 V^\top (P^\pi)^3 e_s$

(2) $\left(G^{(3)}Q\right)(s,a) = r(s,a) + \gamma r^\top P^a e_s + \gamma^2 r^\top P^\pi P^a e_s + \max_a \gamma^3 Q(\cdot, a)^\top (P^\pi)^2 P^a e_s$

א. כתוב/י את ההגדרה מתי אופרטור H כלשהו הוא $\gamma - contracting$, עבור נורמה $\|\cdot\|_\infty$

הגדרה:

ב. לכל אחד מהאופרטורים, ציין למה הוא מתכנס

תשובה:

$R^{(3)}$:

$G^{(3)}$:

הסבר (למה הפתרון המוצע מקיים את הזהות של האופרטור):

תעודת זהות:

מספר מחברת:

ג. הוכח/י שהאופרטורים $R^{(3)}$ R ו- $G^{(3)}$ הינם $\gamma^3 - contracting$.

הוכחה: