سوال ۱:

الف)

$$\text{utility\_up} = -۵۰ + \gamma + \gamma^{۲} + \gamma^{۳} + \dots + \gamma^{۱۰۰}$$

$$\text{utility\_down} = ۵۰ - \gamma - \gamma^{۲} - \dots - \gamma^{۱۰۰}$$

ب)

$$\text{utility\_up} > \text{utility\_down} \longrightarrow \underbrace{\gamma + \gamma^{۲} + \gamma^{۳} + \dots + \gamma^{۱۰۰}}_{\Downarrow} > ۵۰$$

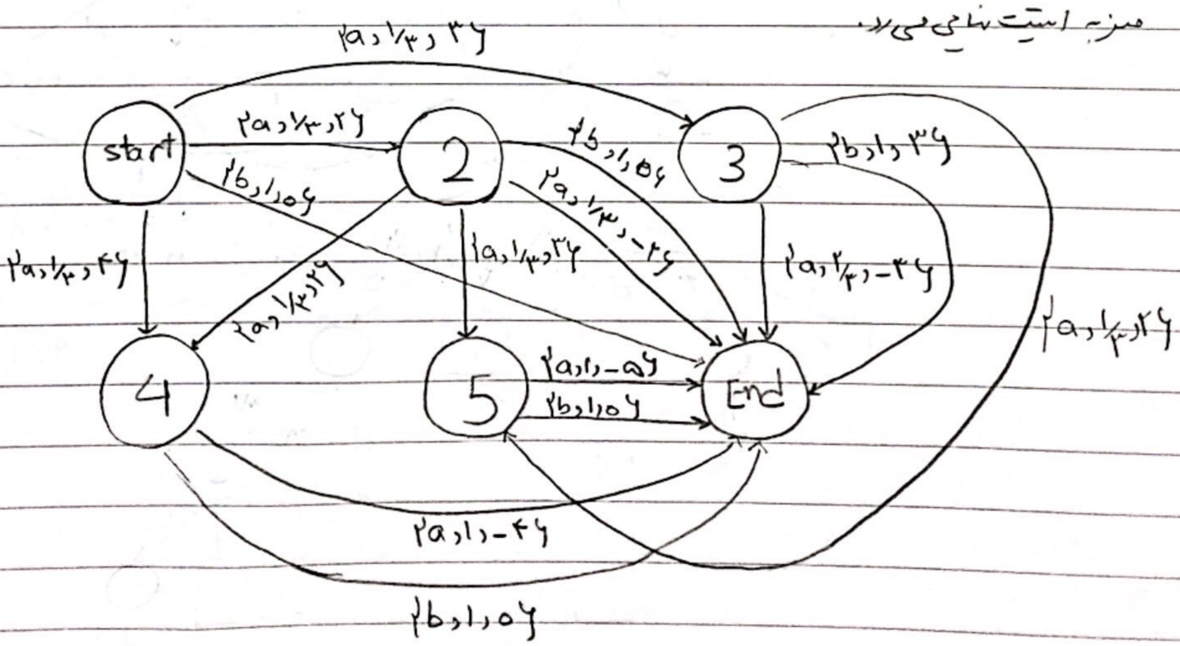
$$\gamma \frac{(۱-\gamma^{۱۰۰})}{۱-\gamma} > ۵۰ \longrightarrow \gamma - \gamma^{۱۰۱} > ۵۰ - ۵۰\gamma$$

$$\longrightarrow ۵۱\gamma - \gamma^{۱۰۱} - ۵۰ > ۰ \quad \boxed{\longrightarrow ۰٫۹۸۴ < \gamma < ۱}$$

به ازای تمام مقادیری که در این معادله صدق کند ایجنت ما باید بالا برود.

سوال ۲:

الف و ب) اکشن های ما a و b است که a برداشتن کارت و b اتمام بازی است (الی در لیبل ۳ چیز آمده که اولی اکشن دومی احتمال رفتن به یک استیت و سومی reward رفتن از یک استیت به استیت بعدی است برای اینکه اگر از ۵ بیشتر شد صفر شود با reward منفی مقدار استیت به استیت نهایی می رود در غیر این صورت اگر انصراف دهد با reward مجموع به استیت نهایی می رود.



ج) برای حساب کردن optimal policy یک policy دلخواه می‌گیریم آن را بسط می‌دهیم تا converge کند
پالیسی اول را به صورت زیر می‌گیریم:

	S	2	3	4	5	E
$\Pi_0 =$	a	a	b	b	b	_
	۰	۰	۰	۰	۰	۰
	۳	۱	۰	۰	۰	۰
	۱۰/۳	۱	۰	۰	۰	۰

با این پالیسی ها جدول را پر می‌کنیم

$$\Pi_{i+1} = \arg\max E(R(S,a,S') + V^{\pi_i}(S_i))]$$

$$\Pi_1(S) = \arg\max_a \begin{cases} b \to ۰ \\ a \to \frac{۱}{۳}(۲+۱) + \frac{۱}{۳}(۳+۰) + \frac{۱}{۳}(۴+۰) = ۱۰/۳ \end{cases} \to a$$

$$\Pi_1(2) = \arg\max_a \begin{cases} b \to ۰ \\ a \to \frac{۱}{۳}(۲+۰) + \frac{۱}{۳}(۳+۰) + \frac{۱}{۳}(-۲+۰) = ۱ \end{cases} \to a$$

$$\Pi_1(3) = \arg\max_a \begin{cases} b \to ۰ \\ a \to \frac{۱}{۳}(۲+۰) + \frac{۱}{۳}(-۳+۰) + \frac{۱}{۳}(-۳+۰) = -۴/۳ \end{cases} \to b$$

$$\Pi_1(۴) = \arg\max_a \begin{cases} b \to ۰ \\ a \to \frac{۳}{۳}(-۴+۰) = -۴ \end{cases} \to b$$

$$\Pi_1(۵) = \arg\max_a \begin{cases} b \to ۰ \\ a \to ۳/۳(-۵) = -۵ \end{cases} \to b$$

$\Pi_1 =$ a a b b b _

در نتیجه $\Pi_0 = \Pi_1$ شده یعنی converge کرده است و $\Pi^* = \Pi_1$ است

(۲

S ۲ ۳ ۴ ۵ E

$V_0 =$ ۰ ۰ ۰ ۰ ۰ ۰

$V = \max_a \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma V_k($

k=۱

$\max_a (S) \begin{cases} b \Rightarrow ۰ \\ a \Rightarrow \frac{1}{۳}(۲) + \frac{1}{۳}(۳) + \frac{1}{۳}(٤) = ۳ \end{cases} \Rightarrow ۳$

$\max_a (۱) = \begin{cases} b = ۰ \\ a = \frac{1}{۳}(۲) + \frac{1}{۳}(۳) + \frac{1}{۳}(-۲) = ۱ \end{cases} \Rightarrow ۱$

$\max_a (٤) \begin{cases} b = ۰ \\ a = -٤ \end{cases}$

$\max_a (۲) = \begin{cases} b = ۰ \\ a = \frac{1}{۳}(۲) + \frac{۲}{۳}(-۳) = -٤/۳ \end{cases} \Rightarrow ۰$

$\max_a (۵) \begin{cases} b = ۰ \\ a = -۵ \end{cases}$

$V_1 =$ ۳ ۱ ۰ ۰ ۰ ۰

k=۲

$\max_a (S) \begin{cases} b = ۰ \\ a = \frac{1}{۳}(۲+۱) + \frac{1}{۳}(۳+۰) + \frac{1}{۳}(٤+۰) = ۱۰/۳ \end{cases} \rightarrow ۱۰/۳$

$\max_a (۲) = \begin{cases} b = ۰ \\ a = \frac{1}{۳}(۲+۰) + \frac{1}{۳}(۳+۰) + \frac{1}{۳}(-۲+۰) \end{cases} \rightarrow ۱$

$\max_a (٤) \begin{cases} b = ۰ \\ a = -٤ \end{cases} \rightarrow ۰$

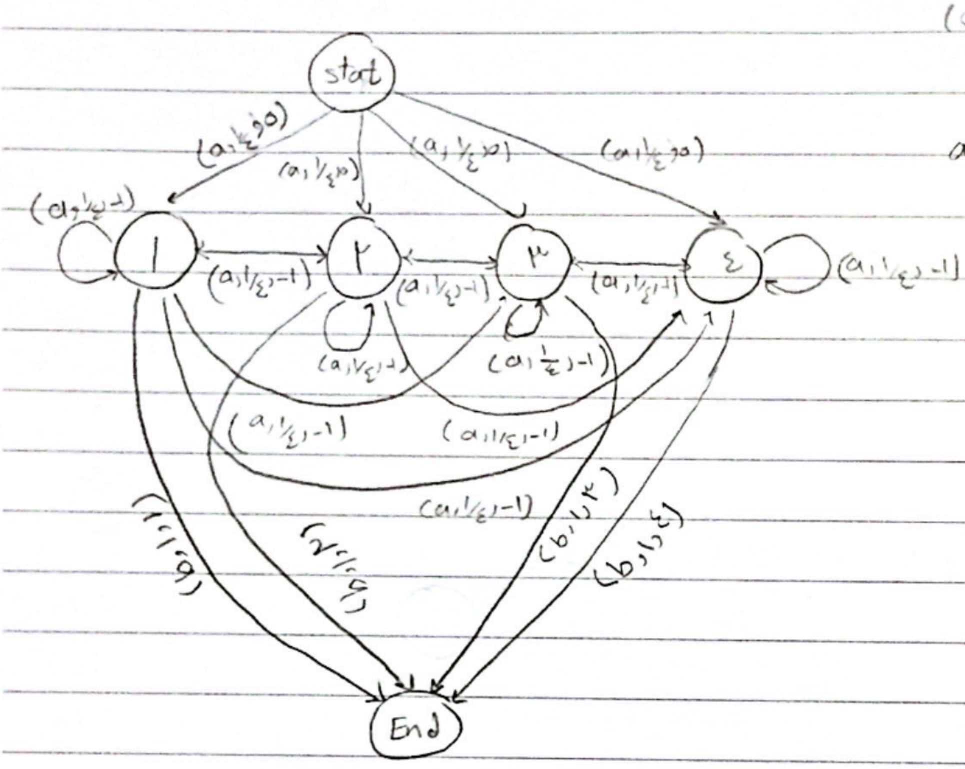
$\max_a (۳) = \begin{cases} b = ۰ \\ a = ۱/۳(۲+۰) + ۲/۳(-۳) = -٤/۳ \end{cases} = ۰$

$\max_a (۵) \begin{cases} b = ۰ \\ a = -۵ \end{cases} \rightarrow ۰$

$V_۲ =$ ۱۰/۳ ۱ ۰ ۰ ۰ ۰

چون مقادیر استیت های ۲ تا ۵ تغییر نکرد converge می‌کند در **k=۲**

سوال ۳: بازی یک استیت شروع، یک پایان و ۴ استیت میانی هم دارد (از استیت start - ۱ تا ۴ را مقدار R را صفر می‌گیریم (چون هنوز بازی شروع نشده است)



روی هر یال ۳ چیز مشخص شده که اولی action دومی احتمال آن و سومی reward آن یال هست

ب)

V₀ = ۱ ۲ ۳ ٤ e

۰ ۰ ۰ ۰ ۰

argmax(۱) over a: a → (-۱×۱/٤) + (۱/٤×-۱) = -۱ ; b → ۱ → ۱

argmax(۲) over a: a → (۱/٤×-۱) = -۱ ; b → ۲ → ۲

V₁ = ۱ ۲ ۳ ۴ ۰

argmax(۱) over a: a → ۱/٤(-۱+۹/۱۰) + ۱/٤(-۱+۲×۹/۱۰) + ۱/٤(-۱+۳×۹/۱۰) + ۱/٤(-۱+٤×۹/۱۰) = ۵/٤ ; b → ۱ → ۵/٤

argmax(۲) over a: a = ۳/۲ ; b → ۲ → ۲

argmax(۳) over a: a = ۳/۲ ; b → ۳ → ۳

V₂ = ۵/٤ ۲ ۳ ۴ ۰

argmax(۱) over a: a → ۱/٤(-۱+۵×۹/٤۰) + ۱/٤(-۱+۲×۹/۱۰) + ۱/٤(-۱+۳×۹/۱۰) + ۱/٤(-۱+٤×۹/۱۰) = ۱٫۳۰۶ ; b → ۱ → ۱۳/۸

V₃ = ۱٫۳۰۶ ۲ ۳ ۴ ۰ → π = a b b b ۰

ج)

۱ ۲ ۳ ۴ E

$\pi_0 =$ b b a a -

$V_0 =$ ۰ ۰ ۰ ۰ ۰

$V_1 =$ ۱ ۲ -۱ -۱ ۰ $\quad + ۱/۴ \left(-۴ + \frac{۱ \times ۹}{۱۰} + \frac{۲ \times ۹}{۱۰} - \frac{۱ \times ۹}{۱۰} - \frac{۱ \times ۹}{۱۰} \right)$

$V_2 =$ ۱ ۲ -۰٫۷۷ -۰٫۷۷

$V_3 =$ ۱ ۲ -۰٫۴۷ -۰٫۴۷ $\quad ۱/۴ (-۴ + ۹/۱۰ (۱ + ۲ - \quad))$

$\pi_1(۱) = \underset{a}{\arg\max} \begin{cases} b \to ۱ \\ a \to \frac{۱}{۴} \left(-۴ + \frac{۹}{۱۰} \left(۱ + ۲ - \frac{۷۰۱}{۴۰۰} \right) \right) = -۰٫۷۲ \end{cases} \longrightarrow b$

$\pi_1(۲) = \underset{a}{\arg\max} \begin{cases} b = ۲ \\ a \to -۰٫۷۲ \end{cases} \to b$

$\to \pi_1 =$ b b b b -

$V_0^{\pi_1}$ ۰ ۰ ۰ ۰ ۰

$V_1^{\pi_1}$ ۱ ۲ ۳ ۴ ۰

$\to \pi_2(۱) = \begin{cases} b = ۱ \\ a = ۱/۴ (-۴ + \frac{۹}{۱۰} \times ۱۰) = ۵/۴ \end{cases} = ۵/۴ \to a$

$\pi_2(۲) = b \qquad \pi_2(۳) = b \qquad \pi_2(۴) = b$

$\pi_2 =$ a b b b -

state	۱	۲	۳	۴	E
π_3	a	b	b	b	-

↑

V_2 (روش ب) = ۱٫۳۱ ۲ ۳ ۴ ۰

$\pi_3(۱) \to \begin{cases} b = ۱ \\ a = ۱/۴ (-۴ + ۹/۱۰ (۱۰٫۳۱)) = ۱٫۳۲ \end{cases} \to a \qquad \pi_3(۲) , \pi_3(۳) , \pi_3(۴) = b$