CS 234 H.W. #3

1. a.)

From the "CartPole_1000" and "CartPole_~~std~~_1000" plots, it's clear that a score of 200 is achieved well before iteration 100, but after iteration 105 the returns stabilize. Increasing the batch_size hyper-parameter to 2000 caused the agent to achieve the 200 score in 60 iterations and less variance (since the gradient estimates were less noisy), ~~however~~ and the returns ~~fail to~~ stabilize and ~~continue to~~ tend to intermittently oscillate ~~significantly~~ slightly ~~from~~ 200. Decreasing the batch_size to 500 caused the agent to achieve the 200 score in about 20 iterations, much faster than the default size of 1000 and on-par with setting size to 2000. Of course, this came at the expense of higher variance as indicated by the "CartPole_std_500" plot. (noisier gradient estimates). Finally, removing the baseline after re-setting to original configuration settings yielded a significantly less stable agent that only achieves 200 at iteration 120 and variance on par with batch_size = 500 with a baseline.

Scanned by CamScanner

b.)

We see from plot "Inverted Pendulum_1000" that for the default configuration settings the agent doesn't reach the required score of 1000 within 100 iterations. However, after doubling batch-size to 2000, the desired target is met ("Inverted Pendulum_2000").

c.)

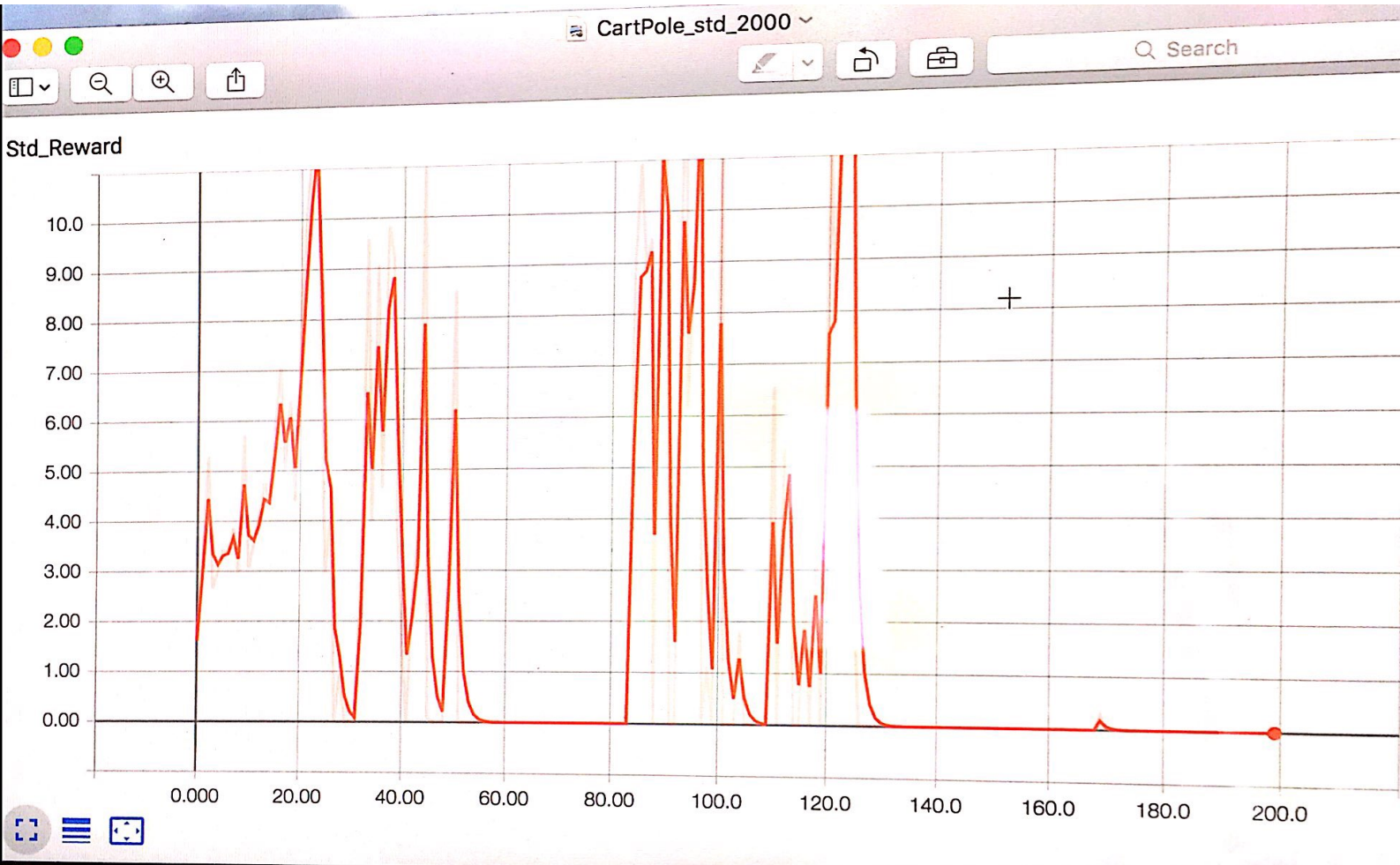
After adjusting gamma to 0.9, batch size to 50,000, and the policy network to 2-hidden layers of size 32 nodes each, we obtain the required score of 200 within 100 iterations as shown in "Half Cheetah_adj

Scanned by CamScanner

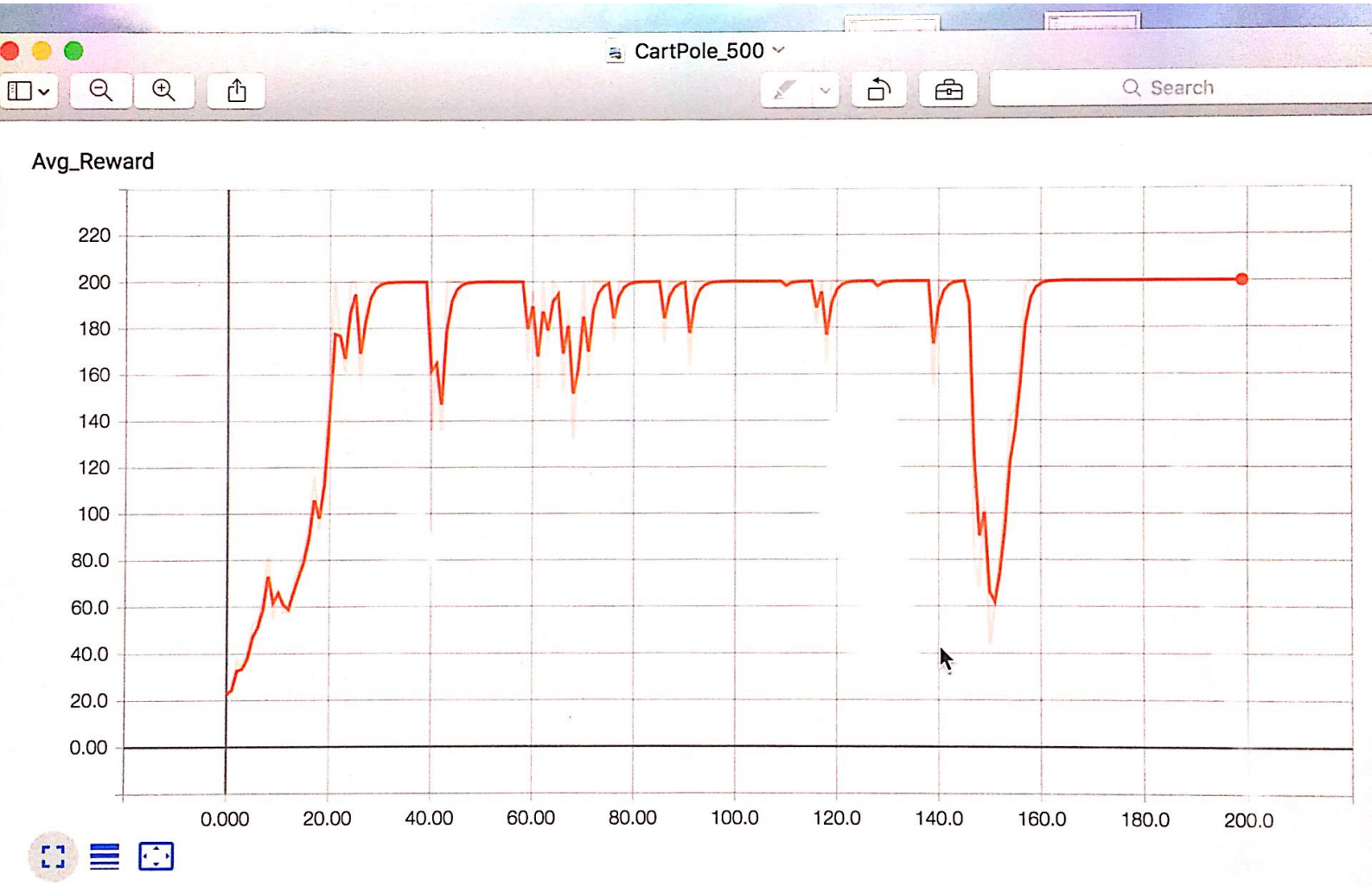
Scanned by CamScanner

Scanned by CamScanner

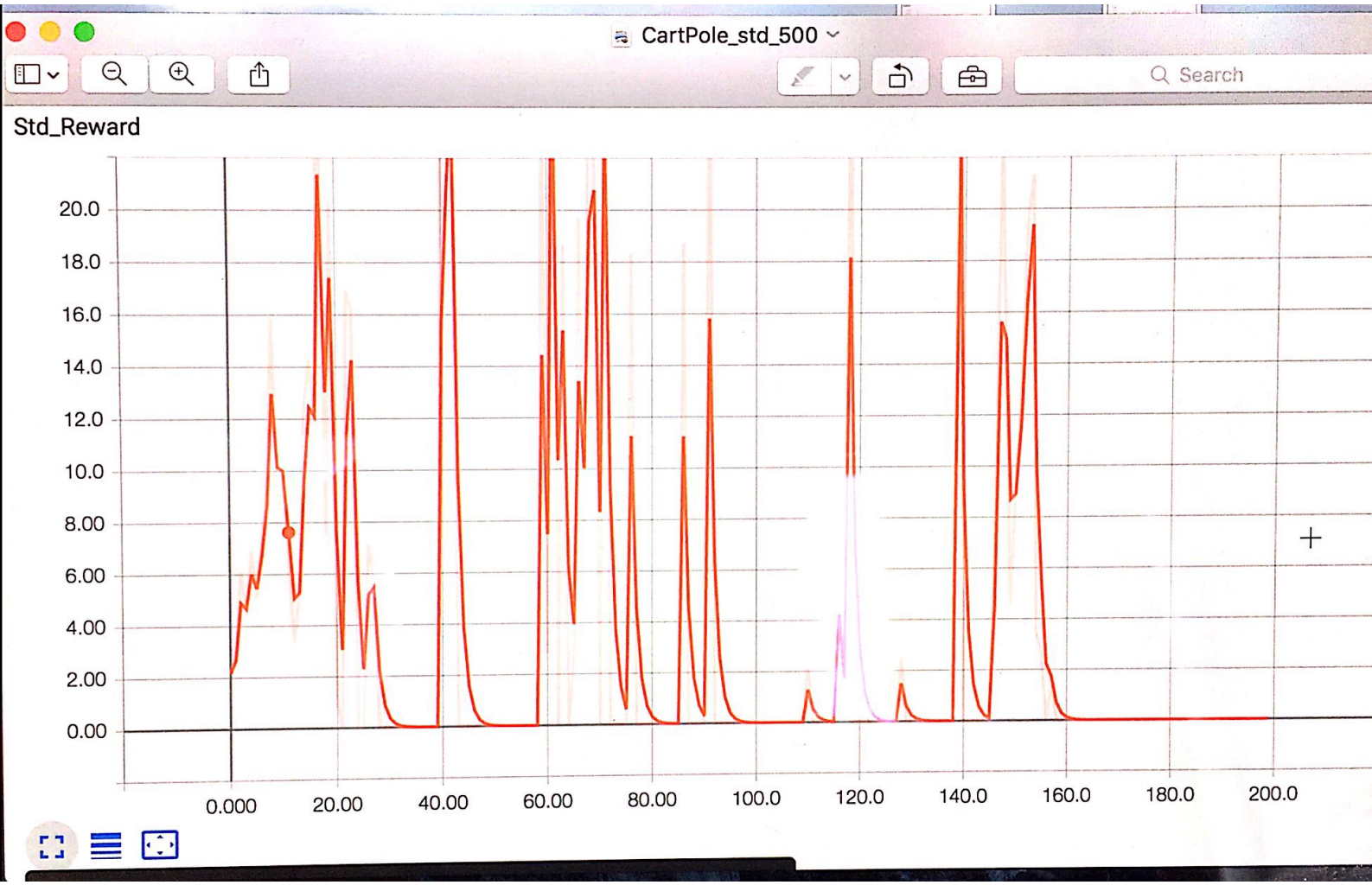
Scanned by CamScanner



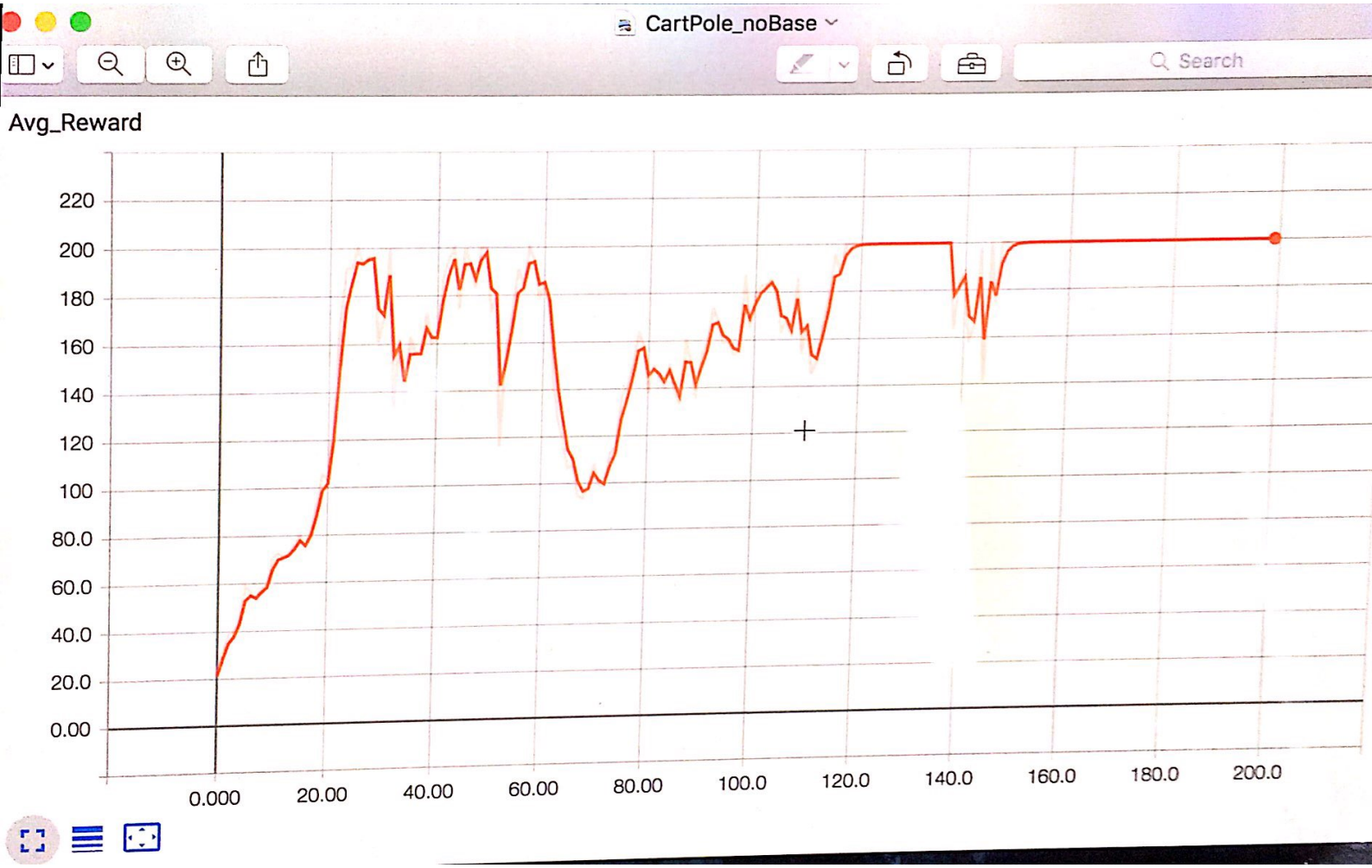
Scanned by CamScanner



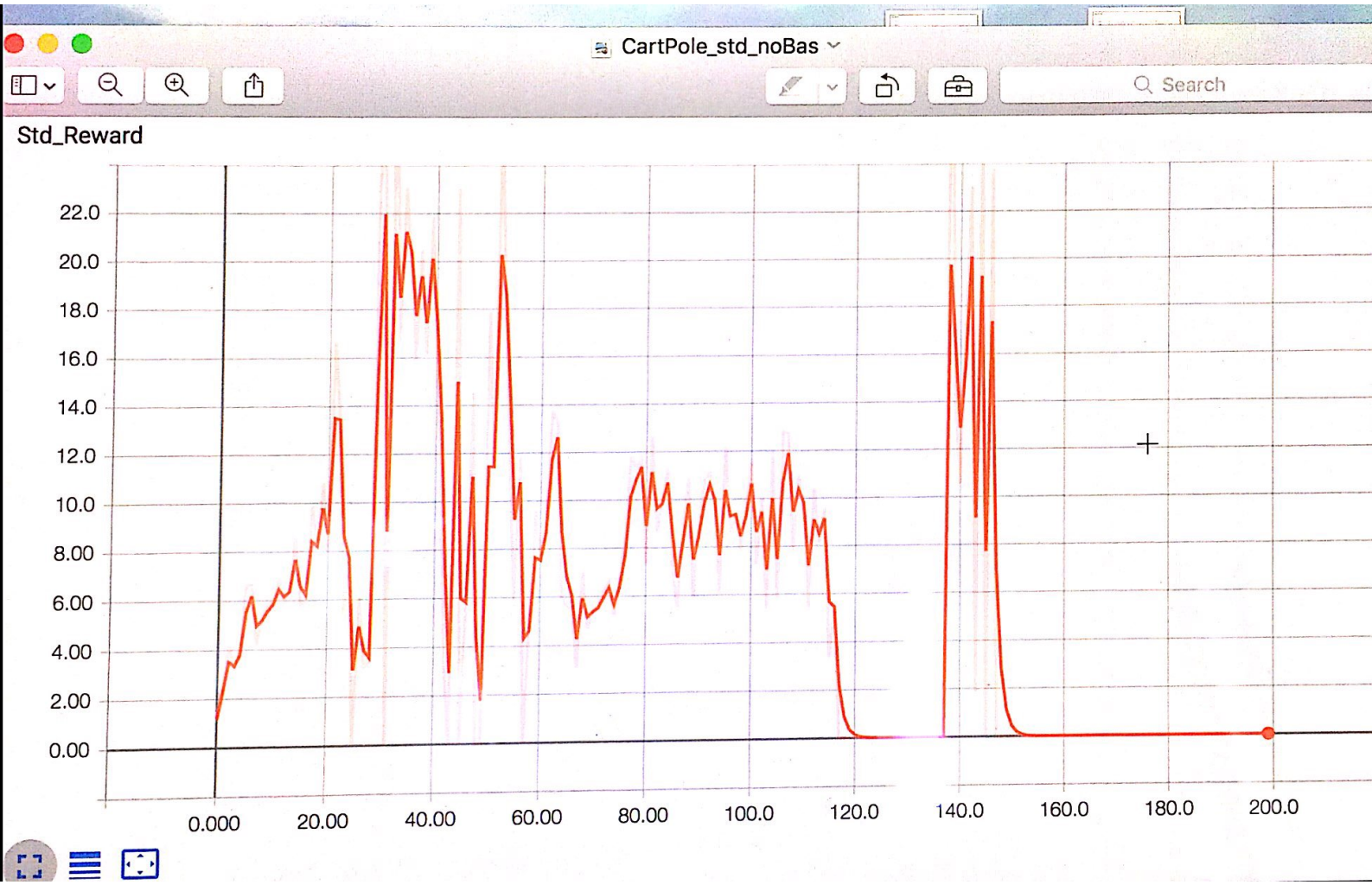
Scanned by CamScanner



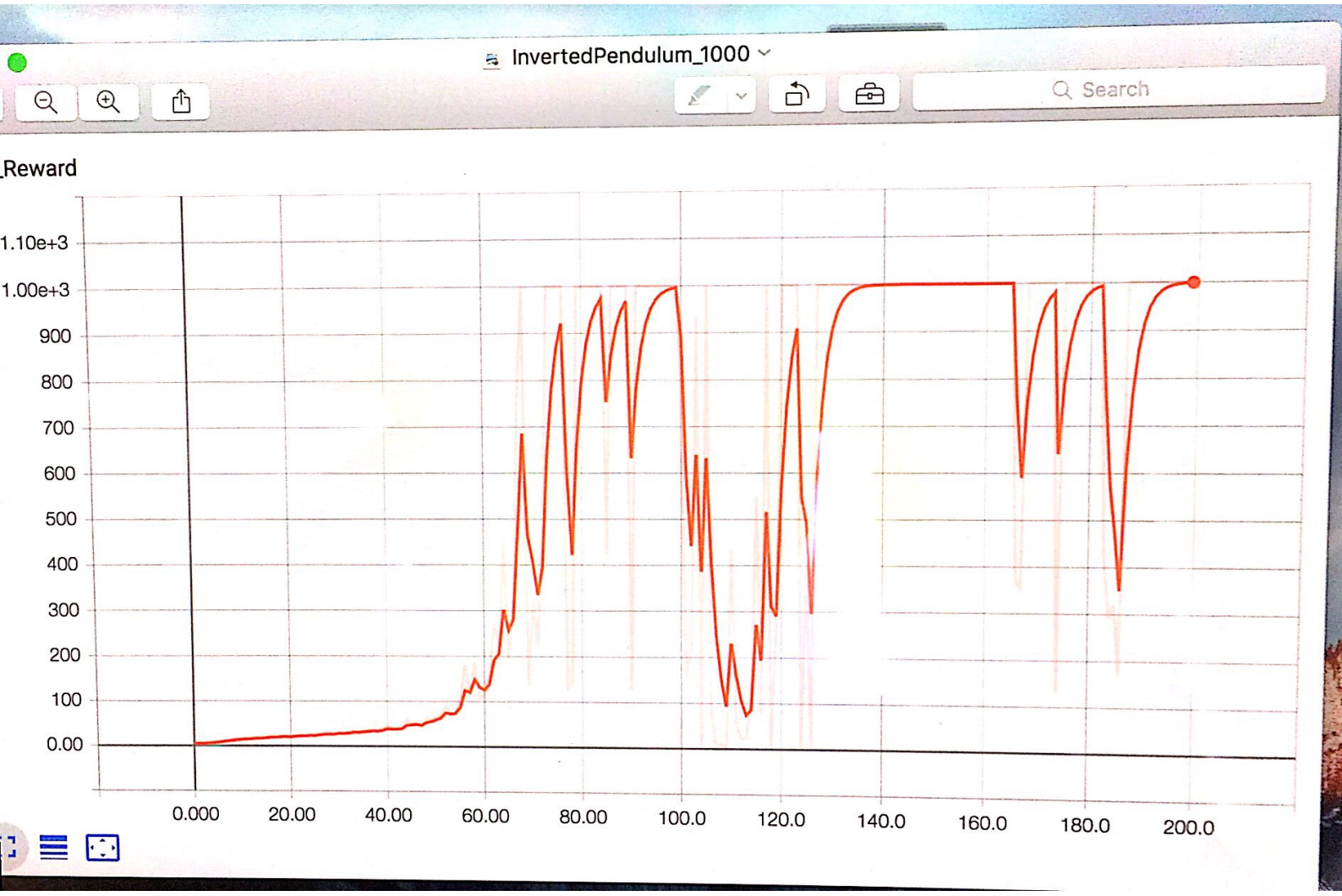
Scanned by CamScanner



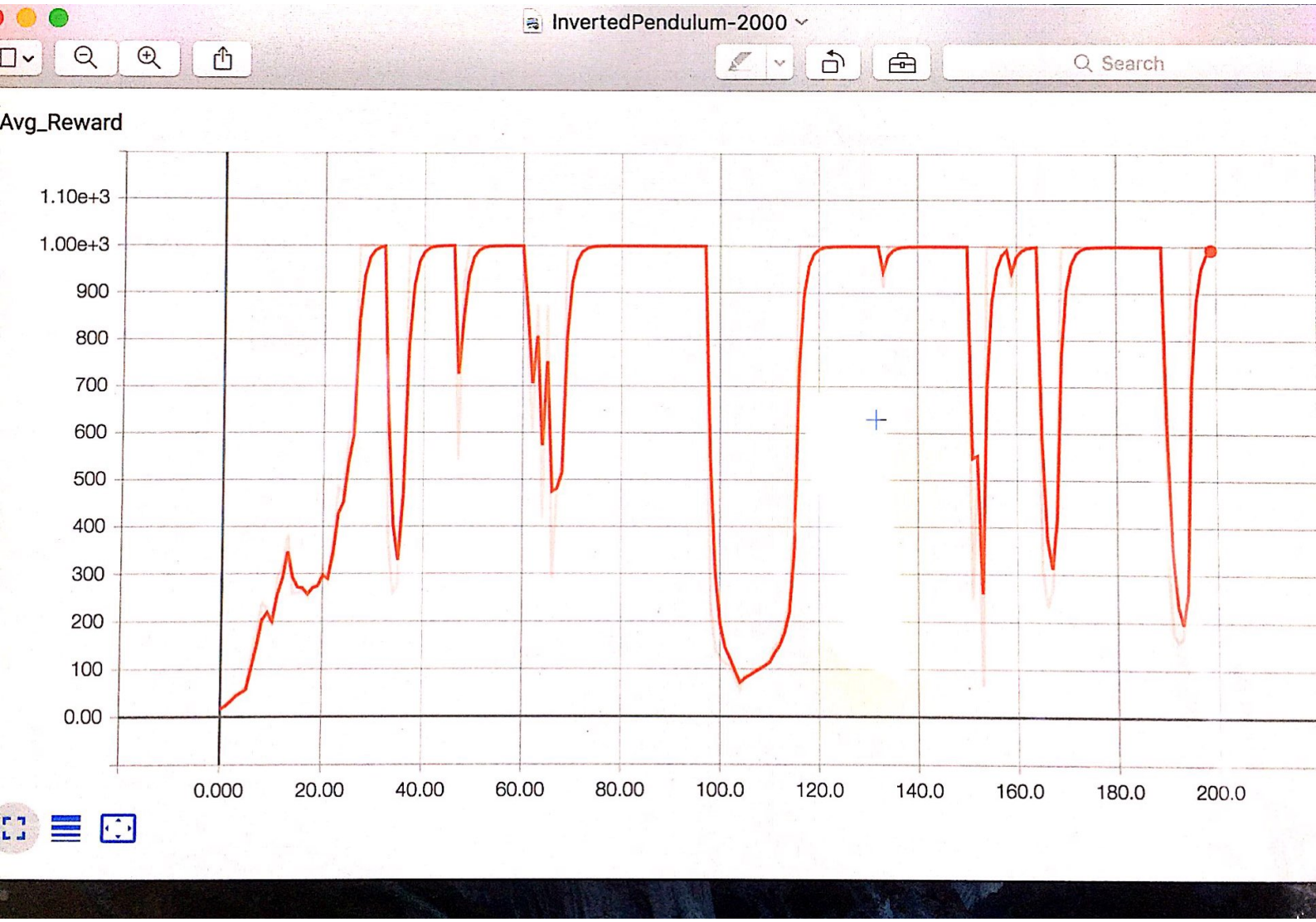
Scanned by CamScanner



Scanned by CamScanner



Scanned by CamScanner



Scanned by CamScanner

2. a.) By definition,

$$V_\pi(s) = E_\pi\left[\sum_{t=1}^{\infty} \gamma^{t-1} r_t \mid s_1 = s\right]$$

$$= \sum_{t=1}^{\infty} \gamma^{t-1} E_\pi[r_t \mid s_1 = s]$$

$$\therefore \left| V_\pi(s) - \sum_{t=1}^{T'} \gamma^{t-1} E_\pi[r_t \mid s_1 = s] \right|$$

$$= \left| \sum_{t=1}^{\infty} \gamma^{t-1} E_\pi[r_t \mid s_1 = s] - \sum_{t=1}^{T'} \gamma^{t-1} E_\pi[r_t \mid s_1 = s] \right|$$

$$= \left| \sum_{t=T'+1}^{\infty} \gamma^{t-1} E_\pi[r_t \mid s_1 = s] \right|$$

as $r_t \in [0, 1], \Rightarrow E_\pi[r_t \mid s_1 = s] \leq 1$

Scanned by CamScanner

This expression becomes

$$\leq \sum_{t=T'+1}^{\infty} \gamma^{t-1} \quad \left(\text{Infinite geo. series}\right)$$

$$= \frac{\gamma^{T'}}{1-\gamma}$$ Now, As $T' \geq \log_{\frac{1}{\gamma}} \frac{H}{\varepsilon}$,

$$\Rightarrow \left(\frac{1}{\gamma}\right)^{T'} \geq \frac{H}{\varepsilon} \Rightarrow \gamma^{T'} \leq \frac{\varepsilon}{H}$$

$$\therefore \left| V_\pi(s) - \sum_{t=1}^{T'} \gamma^{t-1} E_\pi(r_t \mid s_1 = s) \right| \leq \frac{\varepsilon}{H(1-\gamma)}$$

$\therefore$ as $H = \frac{1}{1-\gamma}$,

$$\boxed{\left| V_\pi(s) - \sum_{t=1}^{T'} \gamma^{t-1} E_\pi(r_t \mid s_1 = s) \right| \leq \varepsilon}$$

Scanned by CamScanner

b.)

First things first: we define

$$d(t) = f(T+t) - f(t)$$
$$= \sqrt{T+t} - \sqrt{t}$$

Now, $d'(t) = \dfrac{1}{2\sqrt{T+t}} - \dfrac{1}{2\sqrt{t}} = \dfrac{\sqrt{t} - \sqrt{T+t}}{2\sqrt{t^2+Tt}}$

$< 0 \quad \forall\, t, T > 0$

$\therefore$ $d(t)$ is a decreasing function.

So

$$\left| E_* [V_*(S_t)] - E_A [V_A(S_t)] \right| \leq \varepsilon$$

$$\Rightarrow f(T'+t-1) - f(T'+t-1) + 2\varepsilon' \leq \varepsilon$$

$\Rightarrow$ For any $t^* \geq t$, say $\left(\frac{1+\varepsilon}{\varepsilon}\right)(t-1) = t^* - 1$,

$$f(T'+t^*-1) - f(T'+t^*-1) + 2\varepsilon' \leq \varepsilon$$

Upon substitution for T' and t^*,

Scanned by CamScanner

$$\leq \sqrt{\log_{\frac{1}{4}} \frac{4}{\varepsilon'} + \left(\frac{1+\varepsilon}{\varepsilon}\right)^2 \left(\log_{\frac{1}{4}} \frac{4}{\varepsilon}\right)^2 - \left(\frac{1+\varepsilon}{\varepsilon}\right) \log_{\frac{1}{4}} \frac{4}{\varepsilon} + 2\varepsilon'}$$

choosing $\varepsilon' = \frac{\varepsilon}{4}$, we obtain

(let $x = \log_{\frac{1}{4}} \frac{4}{\varepsilon}$)

$$\leq \sqrt{x + \left(\frac{1+\varepsilon}{\varepsilon}\right)^2 x^2 - \left(\frac{1+\varepsilon}{\varepsilon}\right) x + \frac{\varepsilon}{2}}$$

$$\leq \sqrt{\left(\frac{1+\varepsilon}{\varepsilon} x + \frac{\varepsilon}{2(1+\varepsilon)}\right)^2 - \frac{\varepsilon^2}{4(1+\varepsilon^2)} - \left(\frac{1+\varepsilon}{\varepsilon}\right) x + \frac{\varepsilon}{2}}$$

(∵ completing the square)

$$\leq \sqrt{\left(\left(\frac{1+\varepsilon}{\varepsilon}\right) x + \frac{\varepsilon}{2(1+\varepsilon)}\right)^2} - \left(\frac{1+\varepsilon}{\varepsilon}\right) x + \frac{\varepsilon}{2}$$

$$\leq \frac{\varepsilon}{2(1+\varepsilon)} + \frac{\varepsilon}{2}$$

$$\leq \frac{\varepsilon}{2}\left(\frac{2+\varepsilon}{1+\varepsilon}\right) \quad \varepsilon > 0 \Rightarrow \frac{2+\varepsilon}{1+\varepsilon} < 2$$

$$\leq \varepsilon$$

∴ Result is proven

Scanned by CamScanner