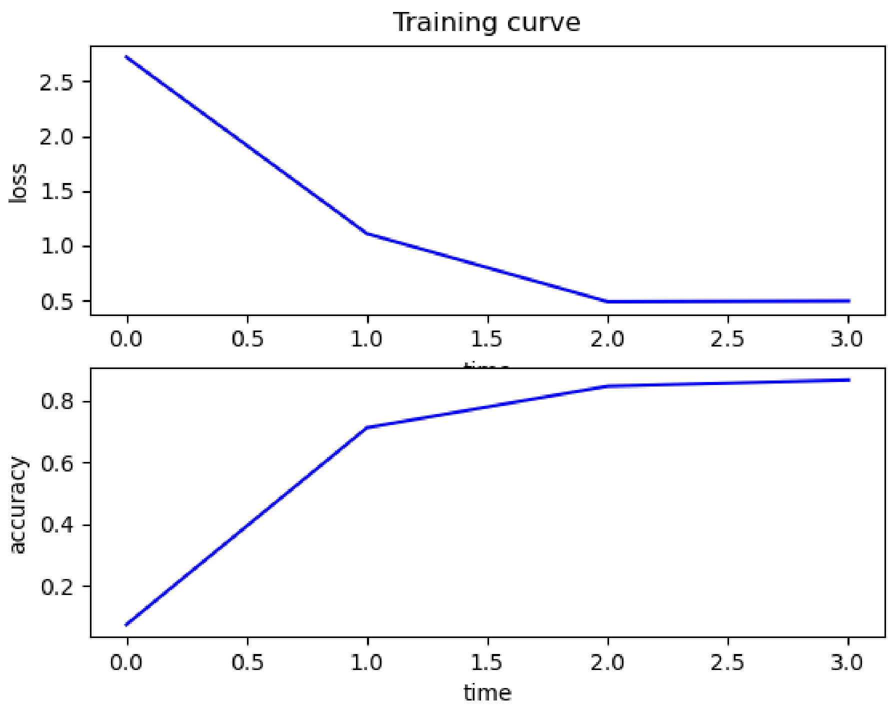
CS 229, Fall 2018

Problem Set #4 Solutions: EM, DL, & RL

1.



2.

(a)

If $\hat{\pi}_0 = \pi_0$, then

$$\begin{aligned}
\mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}} \frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)} R(s,a) &= \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}} \frac{\pi_1(s,a)}{\pi_0(s,a)} R(s,a) \\
&= \sum_{(s,a)} \frac{\pi_1(s,a)}{\pi_0(s,a)} R(s,a) p(s) \pi_0(s,a) \\
&= \sum_{(s,a)} R(s,a) p(s) \pi_1(s,a) \\
&= \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} R(s,a)
\end{aligned}$$

(b)

If $\hat{\pi}_0 = \pi_0$, then

$$\begin{aligned}
\mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}} \frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)} &= \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}} \frac{\pi_1(s,a)}{\pi_0(s,a)} \\
&= \sum_{(s,a)} p(s,a)\frac{\pi_1(s,a)}{\pi_0(s,a)} \\
&= \sum_{(s,a)} p(s)\pi_0(s,a)\frac{\pi_1(s,a)}{\pi_0(s,a)} \\
&= \sum_{(s,a)} p(s)\pi_1(s,a) \\
&= 1
\end{aligned}$$

$$\frac{\mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}} \frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)} R(s,a)}{\mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}} \frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}} = \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} R(s,a)$$

(c)

$$\frac{\mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}} \frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)} R(s,a)}{\mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}} \frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}} = \frac{\sum_{(s,a)} p(s,a)\frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)} R(s,a)}{\sum_{(s,a)} p(s,a)\frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}}$$

If there is only a single data element in the dataset, then

$$\frac{\sum_{(s,a)} p(s,a)\frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)} R(s,a)}{\sum_{(s,a)} p(s,a)\frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}} = \frac{p(s,a)\frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)} R(s,a)}{p(s,a)\frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}} = R(s,a)$$

So if $\pi_0 \neq \pi_1$, then

$$\mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}} R(s,a) \neq \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} R(s,a)$$

(d)

i.

If $\hat{\pi}_0 = \pi_0$, then

$$\mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}} ((\mathbb{E}_{a\sim\pi_1(s,a)} \hat{R}(s,a)) = \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} \hat{R}(s,a)$$

$$\mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}} (\frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}(R(s,a) - \hat{R}(s,a))) = \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} (R(s,a) - \hat{R}(s,a))$$

$$\mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}} ((\mathbb{E}_{a\sim\pi_1(s,a)} \hat{R}(s,a)) + \frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}(R(s,a) - \hat{R}(s,a))) = \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} R(s,a)$$

ii.

If $\hat{R}(s,a) = R(s,a)$, then

$$\mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}} ((\mathbb{E}_{a\sim\pi_1(s,a)} \hat{R}(s,a)) = \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} R(s,a)$$

$$\mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}} (\frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}(R(s,a) - \hat{R}(s,a))) = 0$$

$$\mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}}((\mathbb{E}_{a\sim\pi_1(s,a)}\hat{R}(s,a)) + \frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}(R(s,a)-\hat{R}(s,a))) = \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} R(s,a)$$

(e)

i.

Importance sampling estimator. Estimating $\hat{\pi}_0$ is easier in this situation.

ii.

Regression estimator. Estimating $\hat{R}(s,a)$ is easier in this situation.

3.

$$f_u(x) = \arg\min_{v\in\mathcal{V}} ||x-v||^2 = \frac{uu^Tx}{u^Tu} = uu^Tx$$

$$\begin{aligned}\arg\min_{u:u^Tu=1}\sum_{i=1}^m \|x^{(i)} - f_u(x^{(i)})\|_2^2 &= \arg\min_{u:u^Tu=1}\sum_{i=1}^m \|x^{(i)} - uu^Tx^{(i)}\|_2^2\\ &= \arg\min_{u:u^Tu=1}\sum_{i=1}^m (x^{(i)} - uu^Tx^{(i)})^T(x^{(i)} - uu^Tx^{(i)})\\ &= \arg\min_{u:u^Tu=1}\sum_{i=1}^m {x^{(i)}}^Tx^{(i)} - {x^{(i)}}^Tuu^Tx^{(i)}\\ &= \arg\max_{u:u^Tu=1}\sum_{i=1}^m {x^{(i)}}^Tuu^Tx^{(i)}\\ &= \arg\max_{u:u^Tu=1}\sum_{i=1}^m u^Tx^{(i)}{x^{(i)}}^Tu\\ &= \arg\max_{u:u^Tu=1} u^T\Big(\sum_{i=1}^m x^{(i)}{x^{(i)}}^T\Big)u\end{aligned}$$

4.

(a)

$$\begin{aligned}\nabla_W\ell(W) &= \nabla_W\sum_{i=1}^n\Bigg(\log|W| + \sum_{j=1}^d\log\frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{1}{2}(w_j^Tx^{(i)})^2\Big)\Bigg)\\ &= m(W^{-1})^T - \sum_{i=1}^n\nabla_W\sum_{j=1}^d\frac{1}{2}(w_j^Tx^{(i)})^2\\ &= m(W^{-1})^T - \sum_{i=1}^n Wx^{(i)}{x^{(i)}}^T\\ &= m(W^{-1})^T - WX^TX\\ &= 0\end{aligned}$$

$$W^TW = m(X^TX)^{-1}$$

Let R be an arbitrary orthogonal matrix, and let $W' = RW$. Then

$$W'^TW' = (RW)^TRW = W^TR^TRW = W^TW$$

So if W is a solution, then any W' is also a solution.

(b)

$$\begin{aligned}
\nabla_W \ell(W) &= \nabla_W \Big(\log|W| + \sum_{j=1}^{d} \log \frac{1}{2}\exp\big(-|w_j^T x^{(i)}|\big)\Big) \\
&= (W^{-1})^T - \nabla_W \sum_{j=1}^{d} |w_j^T x^{(i)}| \\
&= (W^T)^{-1} - \text{sign}(Wx^{(i)}){x^{(i)}}^T \\
W &:= W + \alpha\Big((W^T)^{-1} - \text{sign}(Wx^{(i)}){x^{(i)}}^T\Big)
\end{aligned}$$

5.

(a)

$$\begin{aligned}
\|B(V_1) - B(V_2)\|_\infty &= \gamma \left\| \max_{a\in A} \sum_{s'\in S} P_{sa}(s')\big[V_1(s') - V_2(s')\big] \right\|_\infty \\
&= \gamma \max_{s'\in S} \left| \max_{a\in A} \sum_{s'\in S} P_{sa}(s')\big[V_1(s') - V_2(s')\big] \right| \\
&\le \gamma \|V_1 - V_2\|_\infty
\end{aligned}$$

The inequality holds because for any $\alpha, x \in \mathbb{R}^n$, if $\sum_i \alpha_i = 1$ and $\alpha_i \ge 0$, then $\sum_i \alpha_i x_i \le \max_i x_i$

(b)

Assume that V_1 and V_2 are both fixed points, i.e., $B(V_1) = V_1, B(V_2) = V_2$.

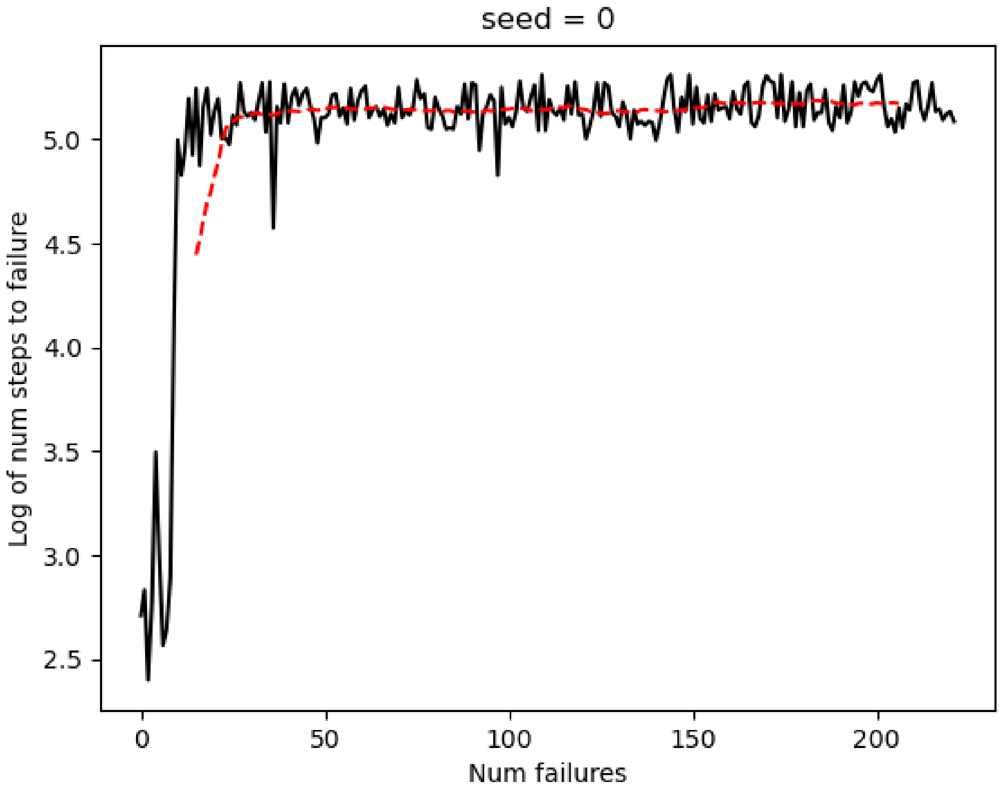
$$\|V_1 - V_2\|_\infty = \|B(V_1) - B(V_2)\|_\infty \le \gamma \|V_1 - V_2\|_\infty$$

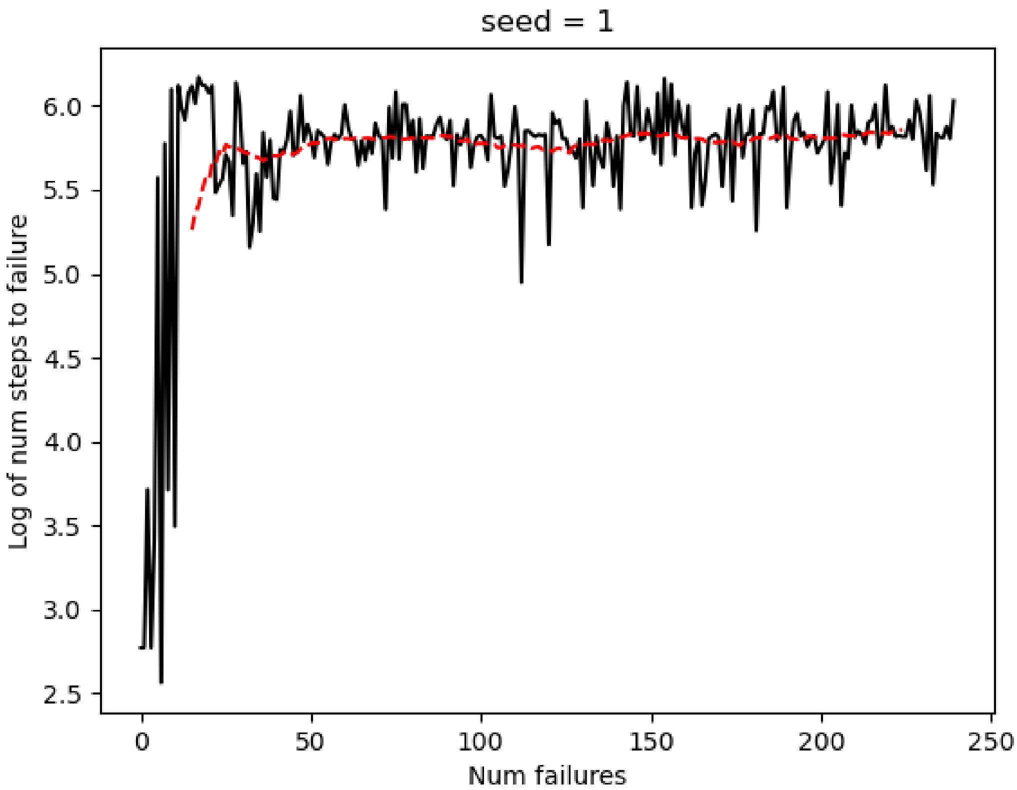
$$\|V_1 - V_2\|_\infty = 0$$

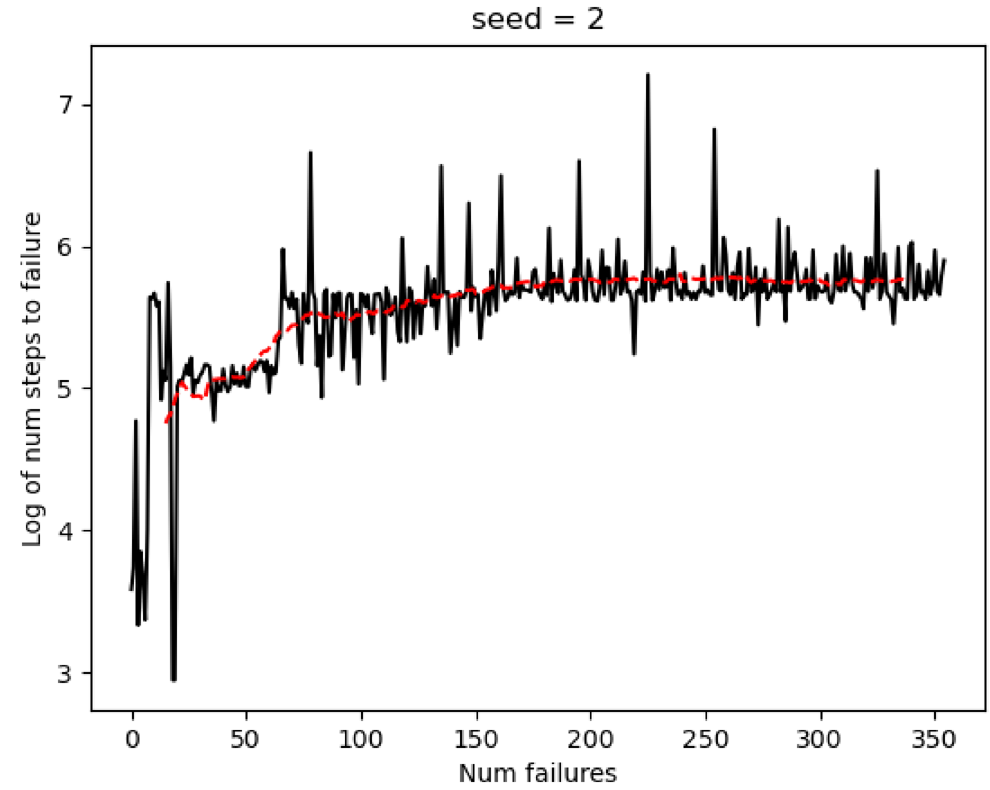
$$V_1 = V_2$$

So B has at most one fixed point.

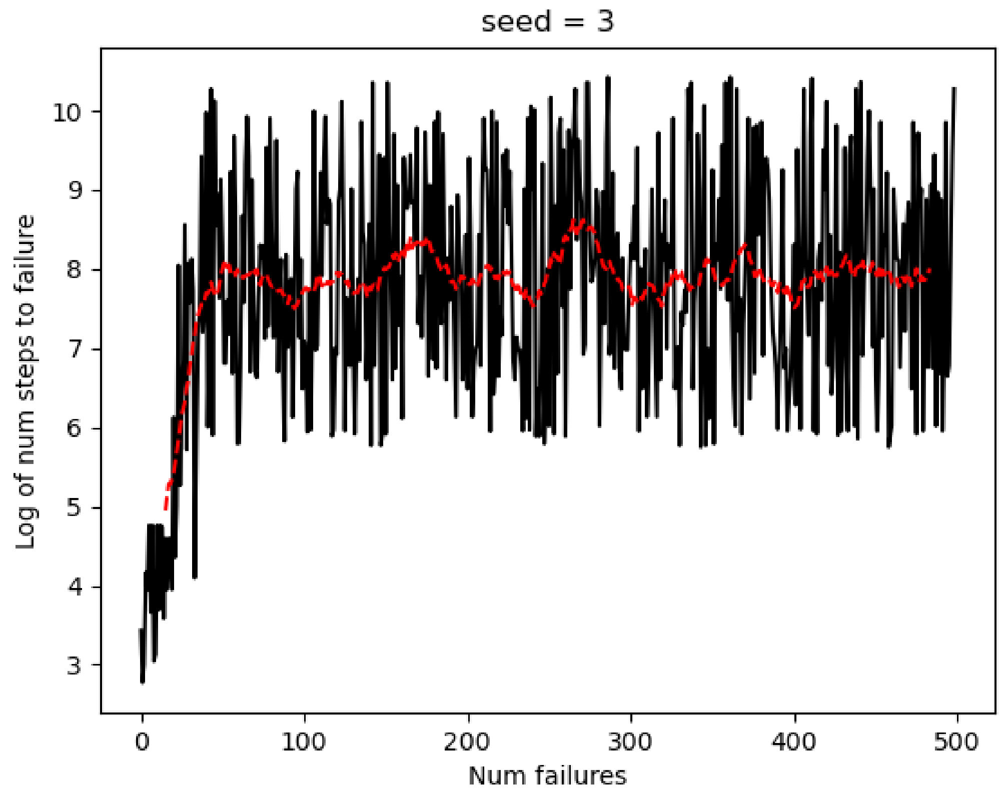
6.


seed = 0
Log of num steps to failure
2.5
3.0
3.5
4.0
4.5
5.0
0
50
100
150
200
Num failures


seed = 1
Log of num steps to failure
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
0
50
100
150
200
250
Num failures



seed = 2
Log of num steps to failure
Num failures



seed = 3
Log of num steps to failure
Num failures