PS4.

2. Off policy Evaluation & Causal Inference.

(a). Importance Sampling.

PROOF: If $\hat{\pi}_0 = \pi_0$, we can see that:

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

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

$$= \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} R(s,a).$$ done!

(b). Weighted Importance Sampling.

PROOF.

Starting at the ~~weighted~~ $\overbrace{\text{weighted}}^{\text{importance}}$ sampling:

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

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

$$\cancel{= \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} R(s,a) \Big/ \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} \mathbb{1} = \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} R(s,a), \text{ done!}}$$

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

$$= \mathbb{E}_{s\sim p(s)} \sum_a \pi_1(s,a) R(s,a) \Big/ \mathbb{E}_{s\sim p(s)} \sum_a \pi_1(s,a) \qquad (1)$$

$$= \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} R(s,a) \Big/ \mathbb{E}_{\substack{s\sim p(s)\\ a}} \mathbb{1}$$

$$= \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} R(s,a).$$ done!

Note, equation (1) is just by replacing $\hat{\pi}_0$ with π_0. ~~and cancel~~

2.(c):

- Consider there is only one single data element.
The weighted importance sampling estimator is:

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

$$= \mathbb{E}_{a\sim\pi_0(s,a)} \frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)} R(s,a) \Big/ \mathbb{E}_{a\sim\pi_0(s,a)} \frac{\pi_1(s,a)}{\pi_0(s,a)} \qquad ①$$

$$= \sum_a \frac{\pi_0(s,a)}{\hat{\pi}_0(s,a)} \pi_1 R(s,a) \Big/ \sum_a \frac{\pi_0(s,a)}{\hat{\pi}_0(s,a)} \pi_1(s,a)$$

$$= \mathbb{E}_{\pi_1}\left[\frac{\pi_0(s,a)}{\hat{\pi}_0(s,a)} R(s,a)\right] \Big/ \mathbb{E}_{\pi_1}\left[\frac{\pi_0(s,a)}{\hat{\pi}_0(s,a)}\right] \qquad ②$$

$$\neq \mathbb{E}_{\pi_1}\left[R(s,a)\right] \qquad ③$$

From ② to ③, Only if $\pi_0(s,a) = \hat{\pi}_0$ is it possible to have the result $\mathbb{E}_{\pi_1}(R(s,a))$, which means whenever $\pi_0 \neq \pi_1$, the weighted importance sampling is biased.

① is because we only consider one data sample.

(d). Doubly Robust. Denote Doubly Robust Estimator as: DRE.

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

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

The left part of above equation, $\mathbb{E}_{a\sim\pi_1(s,a)}\hat{R}(s,a)$ is irrelevant to "a" any more, So the ~~left~~ part is just:

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

Now consider the second part of the equation:

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

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

$= \mathbb{E}_{s\sim p(s)}\left(\sum \pi_1(s,a)(R(s,a)-\hat{R}(s,a))\right)$ (1)

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

The (1) holds from the fact that $\hat{\pi}_0 = \pi_0$.

Sum them up we have the Doubly Robust Estimator:

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

$= \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} R(s,a)$ when $\pi_0 = \hat{\pi}_0$. (i) done!

(ii). Show that ~~$\mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}}$~~ Doubly Robust estimator is equal to $\mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} R(s,a)$ when $R(s,a) = \hat{R}(s,a)$.

PROOF: Since $R=\hat{R}$, Doubly Robust Estimator can be written:

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

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

$= \mathbb{E}_{s\sim p(s)}\left(\mathbb{E}_{a\sim\pi_1(s,a)}\hat{R}(s,a)\right)$, ($\mathbb{E}_{a\sim\pi_1(s,a)}\hat{R}$ is independent of a)

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

(ii) Done!

2,(e).

(i) Importance Sampling.

Since the actions are easy to know, It's natural to use $\hat{\pi}_0$ as an estimate of true π_0 and use importance sampling to estimate π_1.

(ii):

Regression is the choice.

In this case the action is hard to know, but the lifespan can be easily obtained, which means we can use it as $\hat{R}$ to estimate the true $R(s,a)$.

3. PCA. Given any point $x^{(i)}$, the projection to u is given by:
$f_u(x^{(i)}) = (u^T x^{(i)})u$. $u^T x^{(i)}$ is the magnitude of the projected vector, and the magnitude times the unit vector u is the projected vector.

$$\arg\min_{u:u^Tu=1} \sum_{i=1}^{m} \|x^{(i)} - f_u(x^{(i)})\|_2^2 = \arg\min_{u:u^Tu=1} \sum_{i=1}^{m} (x^{(i)} - u^T x^{(i)} u)^T (x^{(i)} - u^T x^{(i)} u)$$

$$= \arg\min_{u:u^Tu=1} \sum^{m} \left(x^{(i)T} x^{(i)} - x^{(i)T}(u^T x^{(i)} u) - u^T x^{(i)} u^T x^{(i)} + (u^T x^{(i)})^2 u^T u \right)$$

$$= \arg\min_{u:u^Tu=1} \sum_{i=1}^{m} \left(x^{(i)T} x^{(i)} - 2(u^T x^{(i)})^2 + u^T u (u^T x^{(i)})^2 \right)$$

$$= \arg\min_{u:u^Tu=1} \sum_{i=1}^{m} \left(x^{(i)T} x^{(i)} - (u^T x^{(i)})^2 \right)$$

$$= \arg\min_{u:u^Tu=1} \sum_{i=1}^{m} \left(-(u^T x^{(i)})^2 \right)$$

$$= \arg\max_{u:u^Tu=1} \sum_{i=1}^{m} (u^T x^{(i)})^2$$

$$= \arg\max_{u:u^Tu=1} \frac{1}{m} \sum_{i=1}^{m} (u^T x^{(i)})^2$$

Since $u^T x^{(i)}$ is just the magnitude of the projection, the last equation is just the variance of the projections, which is to be maximized by ~~se~~ choosing a u.
From lecture notes we know that this ~~as~~ can be written as: $\arg\max_{u:u^Tu=1} u^T \left(\frac{1}{m} \sum_{i=1}^{m} x^{(i)} x^{(i)T} \right) u$.
Maximizing it s.t. $||u||\_2=1$ gives the first principal eigenvector of $\Sigma = \frac{1}{m} \sum_{i=1}^{m} x^{(i)} x^{(i)T}$
Since $V = \{\alpha u : \alpha \in \mathbb{R}\}$, we choose to find a 1-dimensional subspace, the u we choose is the first principal eigenvector of Σ.
Proved!

4. ICA :

(a): $\ell(W) = \sum_{i=1}^{m}\left(\log|W| + \sum_{j=1}^{n}\log g'(w_j^T x^{(i)})\right)$

$= m\log|W| + \sum_{i=1}^{m}\log g'(W^T x^{(i)})$, use Vectorized Gaussian

$= m\log|W| + \sum_{i=1}^{m}\log\frac{1}{(2\pi)^{n/2}}\exp\left(-\frac{1}{2}(W^T x^{(i)})^T(W^T x^{(i)})\right)$

$= m\log|W| - \sum_{i=1}^{m}\frac{1}{2}(W^T x^{(i)})^T(W^T x^{(i)}) - C$. the last term is a constant,

The gradient is :

$\nabla_W \ell(W) = m(W^T)^{-1} - \sum_{i=1}^{m}\frac{1}{2}\nabla_W\left(x^{(i)T}W W^T x^{(i)}\right)$

$= m(W^T)^{-1} - \sum_{i=1}^{m}\frac{1}{2}\nabla_W\left(\mathrm{tr}\, x^{(i)T}W W^T x^{(i)}\right)$

$= m(W^T)^{-1} - \sum_{i=1}^{m}\frac{1}{2}\nabla_W\left(\mathrm{tr}\, x^{(i)}x^{(i)T}W W^T\right)$

$= m(W^T)^{-1} - \sum_{i=1}^{m}x^{(i)}x^{(i)T}\cdot\frac{1}{2}\nabla_W\left(\mathrm{tr}\, W W^T\right)$

$= m(W^T)^{-1} - \sum_{i=1}^{m}x^{(i)}x^{(i)T}\cdot W = m(W^T)^{-1} - XX^T W$

Set it to 0, we have :

$(W^T X)\cdot(W^T X)^T = mI$

$W^T X X^T W = mI$

~~To describe in linear algebra language, the n principal eigenvectors are all with the same eigenvalue m,~~

which means ~~the~~ ~~so~~ we can change the transformation matrix ~~matrix~~ W to be anything in the form of WR, where R is an orthogonal matrix and $RR^T = I$, which means there is no way to tell if the unmix is from W or $W' = WR$.

Here orthogonal matrix R is less formally called a rotation/reflection matrix, the result above is what we call 'rotational invariance'.

4. (b). Laplace source, the PDF: $p_x(x) = p_s(W^T x) \cdot |W| = \frac{1}{2}\exp(-|W^T x|)|W|$

Similar to lecture notes, the log-likelihood is: $\ell(W) = \log(p_s(W^T x))$,

$$\Rightarrow \ell(W) = \sum_{i=1}^{m}\left(\log|W| + \log\left(\frac{1}{2}\exp\left(-|W^T x^{(i)}|\right)\right)\right)$$

$$= m\log|W| + m\log\left(\tfrac{1}{2}\right) - \sum_{i=1}^{m}|W^T x^{(i)}|$$

For a single $x^{(i)}$: $\ell_i = \log|W| + \cdot\log\left(\frac{1}{2}\right) - |W^T x^{(i)}|$

The gradient of the first term w.r.t W is $(W^T)^{-1}$;

The second term is just a constant;

Now consider the gradient ~~of~~ w.r.t W for the last term:

For term $W^T x^{(i)}$, it's an n dimensional vector; It's L1 norm is: $|W^T x^{(i)}| = \sum_{j=1}^{n}|W_j^T x^{(i)}|$, ~~to~~ use chain rule of derivatives:

$$\nabla_W |W^T x^{(i)}| = \begin{bmatrix} \frac{|W^T x^{(i)}|}{W_1^T x^{(i)}} \\ \frac{|W^T x^{(i)}|}{W_2^T x^{(i)}} \\ \vdots \\ \frac{|W^T x^{(i)}|}{W_n^T x^{(i)}} \end{bmatrix} x^{(i)T};$$

Sum ~~the~~ three up: the gradient for a given sample $x^{(i)}$ w.r.t W:

$$\nabla_W = (W^T)^{-1} - \nabla_W |W^T x^{(i)}|$$

$$= (W^T)^{-1} - \begin{bmatrix} \frac{1}{W_1^T x^{(i)}} \\ \frac{1}{W_2^T x^{(i)}} \\ \vdots \\ \frac{1}{W_n^T x^{(i)}} \end{bmatrix} \cdot |W^T x^{(i)}| \, x^{(i)T}.$$

The update rule is thus:

$$W := W + \alpha \left((W^T)^{-1} - \begin{bmatrix} \frac{1}{W_1^T x^{(i)}} \\ \frac{1}{W_2^T x^{(i)}} \\ \vdots \\ \frac{1}{W_n^T x^{(i)}} \end{bmatrix} |W^T x^{(i)}| \, x^{(i)T} \right).$$

5. MDP

(a). From definition in problem description,

$$\|B(V_1) - B(V_2)\|_\infty = \|\gamma \max_{a\in A} \sum_{s'\in S} P_{sa}(s') V_1(s') - \gamma \max_{a\in A} \sum_{s'\in S} P_{sa}(s') V_2(s')\|_\infty$$

For the right part, we can assume that a_1, a_2 are taken to maximize $\sum_{s'\in S} P_{sa_1}(s') V_1(s')$ and $\sum_{s'\in S} P_{sa_2}(s') V_2(s')$, hence:

$$\|B(V_1) - B(V_2)\|_\infty = \|\gamma \sum_{s'\in S} P_{sa_1}(s') V_1(s') - \gamma \sum_{s'\in S} P_{sa_2}(s') V_2(s')\|_\infty$$

$$= \gamma \max_{s\in S} \left|\sum_{s'\in S} P_{sa_1} V_1(s') - \sum_{s'\in S} P_{sa_2}(s') V_2(s')\right|$$

$$\leq \gamma \max_{s\in S} \left|\sum_{s'\in S} P_{sa_1} V_1(s') - \sum_{s'\in S} P_{sa_1}(s') V_2(s')\right|$$, change from a_2 to a_1 for the 2nd term makes it smaller,

$$= \gamma \max_{s\in S} \left|\sum_{s'\in S} P_{sa_1}(s')\left(V_1(s') - V_2(s')\right)\right|$$

$$\leq \gamma \max_{s\in S} \max_{s'\in S} \left|V_1(s') - V_2(s')\right|.$$ [expected value is smaller than max value]

$$= \gamma \max_{s'\in S} \left|V_1(s') - V_2(s')\right|$$ $\max_{s\in S}$ has no effect on the result any more.

$$= \|V_1 - V_2\|_\infty.$$ done.

(b): Suppose B has more than 1 fixed point, they are V_1, V_2. According to γ-contraction, it's obvious that:

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

$$\leq \gamma^2\|V_1 - V_2\|_\infty = \gamma^2\|B(V_1) - B(V_2)\|_\infty$$

$$\leq \gamma^3\|V_1 - V_2\|_\infty$$

$$\cdots$$

$$\leq \gamma^n\|B(V_1) - B(V_2)\|_\infty \to 0,$$

Which is contradictory to the assumption that V_1, V_2 are different.

So it's not possible that B has multiple fixed point.

done with the proof that B has at most one fixed point.