2.

(a) [2 points] **Importance Sampling:** One commonly used estimator is known as the importance sampling estimator. Let $\hat{\pi}_0$ be an estimate of the true π_0. The importance sampling estimator uses that $\hat{\pi}_0$ and has the form:

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

Please show that if $\hat{\pi}_0 = \pi_0$, then the importance sampling estimator is equal to:

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

Note that this estimator only requires us to model π_0 as we have the $R(s,a)$ values for the items in the observational data.

$$E_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}} \frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)} R(s,a)$$

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

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

$$= \sum_{(s,a)} \frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)} \pi_0(s,a) p(s) R(s,a)$$

if $\hat{\pi}_0 = \pi_0$

$$\sum_{(s,a)} \pi_1(s,a) p(s) R(s,a)$$

$$= \sum_{(s,a)} p(a|s) p(s) R(s,a) = E_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} R(s,a)$$

(b) [2 points] **Weighted Importance Sampling:** One variant of the importance sampling estimator is known as the weighted importance sampling estimator. The weighted importance sampling estimator has the form:

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

Please show that if $\hat{\pi}_0 = \pi_0$, then the weighted importance sampling estimator is equal to:

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

$\Leftarrow$ if $\hat{\pi}_0 = \pi_0$, $\mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}} \frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)} = 1$

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

$$= \sum_{(s,a)} p(s)\pi_0(s,a) \frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}$$

$\hat{\pi}_0 = \pi_0$

$$\sum_{(s,a)} p(s,a) = 1$$

(c) [2 points] One issue with the weighted importance sampling estimator is that it can be biased in many finite sample situations. In finite samples, we replace the expected value with a sum over the seen values in our observational dataset. Please show that the weighted importance sampling estimator is biased in these situations.

Hint: Consider the case where there is only a single data element in your observational dataset.

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

$$= \frac{\sum_{(s,a)} p(s,a) \frac{\pi_1(s,a)}{\pi_0(s,a)} R(s,a)}{\sum_{(s,a)} p(s,a) \frac{\pi_1(s,a)}{\pi_0(s,a)}}$$

only one example $\quad \frac{p(s,a) \frac{\pi_1(s,a)}{\pi_0(s,a)} R(s,a)}{p(s,a) \frac{\pi_1(s,a)}{\pi_0(s,a)}} = R(s,a)$

$R(s,a)$ does not have to be equal to $\mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} R(s,a)$

(d) [7 points] **Doubly Robust**: One final commonly used estimator is the doubly robust estimator. The doubly robust estimator has the form:

$$\mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}}\left((\mathbb{E}_{a\sim\pi_1(s,a)}\hat{R}(s,a)) + \frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}(R(s,a)-\hat{R}(s,a))\right) \quad (1)$$

One advantage of the doubly robust estimator is that it works if either $\hat{\pi}_0 = \pi_0$ or $\hat{R}(s,a) = R(s,a)$

i. [4 points] Please show that the doubly robust estimator is equal to $\mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} R(s,a)$ when $\hat{\pi}_0 = \pi_0$

ii. [3 points] Please show that the doubly robust estimator is equal to $\mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} R(s,a)$ when $\hat{R}(s,a) = R(s,a)$

(i)

$$E_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}}\left(\left(E_{a\sim\pi_1(s,a)}\hat{R}(s,a)\right) + \frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}\left(R(s,a)-\hat{R}(s,a)\right)\right)$$

if $\pi_0 = \hat{\pi}_0$:

$$E_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}}\frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}\left(R(s,a)-\hat{R}(s,a)\right) = E_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}}\left(R(s,a)-\hat{R}(s,a)\right)$$

$$E_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}}\left(E_{a\sim\pi_1(s,a)}\hat{R}(s,a)\right)$$

$$= \sum_{(s,a)} E_{a\sim\pi_1(s,a)}\hat{R}(s,a)\,P(a|s,\pi_0)\,P(s)$$

$$= \sum_{(s,a)}\left(\sum_a P(a|s,\pi_1)\hat{R}(s,a)\right)P(s)\,P(a|s,\pi_0)$$

$$= \sum_a\left(\sum_{(s,a)} P(s,a|\pi_1)\hat{R}(s,a)\right)P(a|s,\pi_0)$$

$$= E_{a\sim\pi_0(s,a)}\left(\underbrace{E_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}}\hat{R}(s,a)}_{\text{constant}}\right) = E_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}}\hat{R}(s,a)$$

$$(1) = E_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} R(s,a) \quad \text{if } \pi_0 = \hat{\pi}_0$$

(ii) if $\hat{R}(s,a) = R(s,a)$

$$(1) = \mathbb{E}_{\substack{s\sim p(s) \\ a\sim\pi_0(s,a)}} \left(\mathbb{E}_{a\sim\pi_1(s,a)} R(s,a) \right)$$

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

(e) [2 points] We will now consider several situations where you might have a choice between the importance sampling estimator and the regression estimator. Please state whether the importance sampling estimator or the regression estimator would probably work best in each situation and explain why it would work better. In all of these situations, your states s consist of patients, your actions a represent the drugs to give to certain patients and your $R(s, a)$ is the lifespan of the patient after receiving the drug.

i. [1 points] Drugs are randomly assigned to patients, but the interaction between the drug, patient and lifespan is very complicated.

ii. [1 points] Drugs are assigned to patients in a very complicated manner, but the interaction between the drug, patient and lifespan is very simple.

i. Importance sampling estimator. As policy function $\hat{\pi}_0$ is easier to estimate

ii. Regression estimator. Interaction between drug & patient meaning $\hat{R}(s,a)$ would be easy to get.

3. **[10 points] PCA**

In class, we showed that PCA finds the "variance maximizing" directions onto which to project the data. In this problem, we find another interpretation of PCA.

Suppose we are given a set of points $\{x^{(1)}, \ldots, x^{(m)}\}$. Let us assume that we have as usual preprocessed the data to have zero mean and unit variance in each coordinate. For a given unit-length vector u, let $f_u(x)$ be the projection of point x onto the direction given by u. I.e., if $\mathcal{V} = \{\alpha u : \alpha \in \mathbb{R}\}$, then

$$f_u(x) = \arg\min_{v \in \mathcal{V}} ||x - v||^2.$$

Show that the unit-length vector u that minimizes the mean squared error between projected points and original points corresponds to the first principal component for the data. I.e., show that

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

gives the first principal component.

Remark. If we are asked to find a k-dimensional subspace onto which to project the data so as to minimize the sum of squares distance between the original data and their projections, then we should choose the k-dimensional subspace spanned by the first k principal components of the data. This problem shows that this result holds for the case of $k = 1$.

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

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

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

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

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

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

4.

(a) [5 points] **Gaussian source**

For this sub-question, we assume sources are distributed according to a standard normal distribution, i.e $s_j \sim \mathcal{N}(0,1), j = \{1,\dots,d\}$. The likelihood of our unmixing matrix, as described in the notes, is

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

where g is the cumulative distribution function, and g' is the probability density function of the source distribution (in this sub-question it is a standard normal distribution). Whereas in the notes we derive an update rule to train W iteratively, for the cause of Gaussian distributed sources, we can analytically reason about the resulting W.

Try to derive a closed form expression for W in terms of X when g is the standard normal CDF. Deduce the relation between W and X in the simplest terms, and highlight the ambiguity (in terms of rotational invariance) in computing W.

$$\nabla_W(\ell(W)) = \nabla_W \sum_{i=1}^{n} \left(\log|W| + \sum_{j=1}^{d} \log\left[\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(w_j^T x^{(i)})\right)\right]\right)$$

$$= n(W^{-1})^T - \sum_{i=1}^{n} \nabla_W \sum_{j=1}^{d} \frac{1}{2}(w_j^T x^{(i)})^2 + 0$$

$$= n(W^{-1})^T - \sum_{i=1}^{n} \nabla_W \left(\frac{1}{2}(Wx^{(i)})^T (Wx^{(i)}) \right)$$

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

$$= n(W^{-1})^T - \sum_{i=1}^{n} W x^{(i)} x^{(i)T}$$

$$= n(W^T)^T - WX^TX \quad \overset{\text{set}=0}{=\!=} \quad W^TW = n(X^TX)^{-1} \quad (1)$$

Let R be an arbitrary orthogonal matrix, $R^TR = RR^T = I$

let $W' = RW$

$$(W')^TW' = W^TR^TRW = W^TW = n(X^TX)^{-1}$$

meaning RW is also a solution of (1)

(b) [10 points] **Laplace source.**

For this sub-question, we assume sources are distributed according to a standard Laplace distribution, i.e $s_i \sim \mathcal{L}(0,1)$. The Laplace distribution $\mathcal{L}(0,1)$ has PDF $f_{\mathcal{L}}(s) = \frac{1}{2}\exp(-|s|)$. With this assumption, derive the update rule for a single example in the form

$$W := W + \alpha(\dots).$$

$$\ell(W) = \left(\log|W| + \sum_{j=1}^{d} \log\left(f_{\mathcal{L}}(w_j^T x^{(i)})\right)\right)$$

$$= \left(\log|W| + \sum_{j=1}^{d}\left(\log\frac{1}{2} - |w_j^T x^{(i)}|\right)\right)$$

$$\nabla_W \ell(W) = (W^{-1})^T - \nabla_W \sum_{j=1}^{d} |w_j^T x^{(i)}|$$

$$= (W^{-1})^T - \underbrace{\underbrace{\operatorname{sign}(Wx^{(i)})}_{d\times 1}\, x^{(i)T}}_{d\times d} \longrightarrow 1\times d$$

$$\therefore\ W := W + \alpha(\nabla_W \ell(W))$$

$$= W + \alpha\left[(W^{-1})^T - \operatorname{sign}(Wx^{(i)})\, x^{(i)T}\right]$$

5. **[15 points] Markov decision processes**

Consider an MDP with finite state and action spaces, and discount factor $\gamma < 1$. Let B be the Bellman update operator with V a vector of values for each state. I.e., if $V' = B(V)$, then

$$V'(s) = R(s) + \gamma \max_{a \in A} \sum_{s' \in S} P_{sa}(s') V(s').$$

(a) **[10 points]** Prove that, for any two finite-valued vectors V_1, V_2, it holds true that

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

where

$$||V||_\infty = \max_{s \in S} |V(s)|.$$

(This shows that the Bellman update operator is a "γ-contraction in the max-norm.")

$$\|B(V_1) - B(V_2)\|_\infty$$

$$= \max_{s' \in S} |B(V_1) - B(V_2)|$$

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

$$\leq \gamma \max_{s' \in S} \cdot \max_{a \in A} \left| \sum_{s' \in S} P_{sa}(s') \left(V_1(s') - V_2(s') \right) \right|$$

$$= \gamma \max_{s' \in S} \left| E_{s' \sim P_{sa^*}} \left(V_1(s') - V_2(s') \right) \right|$$, given $g(x) = |x|$ is convex with Jensen's Inequality:

$$\leq \gamma \max_{s' \in S} E_{s' \sim P_{sa^*}} \left| V_1(s') - V_2(s') \right|$$

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

(b) **[5 points]** We say that V is a **fixed point** of B if $B(V) = V$. Using the fact that the Bellman update operator is a γ-contraction in the max-norm, prove that B has at most one fixed point—i.e., that there is at most one solution to the Bellman equations. You may assume that B has at least one fixed point.

If V_1, V_2 are both B' fixed points

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

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

so $\|V_1 - V_2\|_\infty \leq \gamma \|V_1 - V_2\|_\infty$

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

so B has at most one fixed point.