Problem Set 4

1. (a). (i).

$$\text{softmax}(x_i) = \frac{e^{x_i - \max(x)}}{\sum_j e^{x_j - \max(x)}} \qquad \frac{\partial\, \text{softmax}(x)}{\partial x_i} = \left[\left(e^{x_i - \max(x)} \cdot (1 - 0/1) \cdot \sum_j e^{x_j - \max(x)}\right)\right.$$

$$\left. - \left(e^{x_i - \max(x)} \cdot e^{x_i - \max(x)} \cdot (1 - 0/1)\right)\right] \div \left(\sum_j e^{x_j - \max(x)}\right)^2$$

(ii) $\text{Relu}.(x_i) = \begin{cases} x_i & \text{if } x_i > 0 \\ 0 & \text{otherwise.} \end{cases} \qquad \frac{\partial \text{Relu}(x_i)}{\partial x_i} = \begin{cases} 1 & \text{if } x_i > 0 \\ 0 & \text{otherwise.} \end{cases}$

(iii). $CE(y, \hat{y}) = -\sum_{k=1}^{K} y_k \log \hat{y}_k \qquad \frac{\partial\, CE(y,\hat{y})}{\partial \hat{y}_k} = -\sum_{k=1}^{K} \frac{y_k}{\hat{y}_k}$

(iv). $\frac{\partial}{\partial x} = \vec{w} \quad \frac{\partial}{\partial w} = \vec{x} \quad \frac{\partial}{\partial b} = \vec{1}$

(v). $\frac{\partial_{1,1}}{\partial b} = 1 \qquad \frac{\partial\, \text{channel}}{\partial b, c} = \begin{bmatrix} 1 & 1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \cdots & \cdots & 1 \end{bmatrix} \qquad \frac{\partial\, \text{loss}}{\partial b} = \underbrace{\left\langle \text{output\_grad}, \frac{\partial\, \text{channel}}{\partial b} \right\rangle}_{\text{inner product.}}$

$$\frac{\partial_{1,1}}{\partial w} = \sum_{c,x,y} \text{data}[:, x:(x+\text{width}), y:(y+\text{height})] * \text{output\_grad}[c, x, y]$$

broadcast.

$$\frac{\partial_{11}}{\partial \text{input}} = \sum_{c,x,y} \text{out\_grad}[c, x, y] * \text{con\_W}[c, 0, :, :]$$

(vi). for c, x, y in output_grad:

index := argmax (data [c, x · pool_width, y · pool_height]). size pool_width, pool_height.

grad_data [range] [index] = out_grad [c, x, y].

2. (a). if $\hat{\pi}_0 = \pi_0$, then

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

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

$$= \sum_{(s,a)} \frac{\pi_1(s,a)}{\cancel{\pi_0(s,a)}} \cdot R(s,a) \cdot \cancel{\pi_0(s,a)} \cdot P(s).$$

$$= \sum_{(s,a)} P(a|s;\pi_1) \cdot P(s) \cdot R(s,a)$$

$$= \sum_{(s,a)} P(s,a;\pi_1) \cdot R(s,a)$$

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

(b).

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

$$= \frac{\mathbb{E}_{\substack{s\sim P(s)\\ a\sim\pi_1(s,a)}} R(s,a).}{\sum_{(s,a)} \frac{\pi_1(s,a)}{\cancel{\pi_0(s,a)}} \cdot P(s) \cdot \cancel{P(a|s;\pi_0)}}$$

$$= \frac{\mathbb{E}_{\substack{s\sim P(s)\\ a\sim\pi_1(s,a)}} R(s,a)}{\sum_{(s,a)} P(s,a;\pi_1)}$$

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

(c). if there is only a single element in the observational dataset, then.

the weighted Importance Sampling $= \dfrac{\sum_{(s,a)} \frac{\pi_1(s,a)}{\pi_0(s,a)} R(s,a)}{\sum_{(s,a)} \frac{\pi_1(s,a)}{\pi_0(s,a)}} = R(s,a),$

which can be biased.

(add:).

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

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

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

hence $A = B - C + \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_0(s,a)}} \frac{\pi_1(s,a)}{\pi_0(s,a)} R(s,a)$

$$= 0 + \mathbb{E}_{\substack{s\sim p(s)\\ a\sim\pi_1(s,a)}} R(s,a)$$

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

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

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

$$= \sum_s p(s) \sum_a \pi_0(s,a) \left(\sum_a \pi_1(s,a) R(s,a) \right)$$

$$= \sum_s p(s) \left(\sum_a \pi_1(s,a) R(s,a) \right)$$

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

(e). (i). since the interaction between the three components is complicated, it would be ~~differ~~ difficalt to estimate $\hat{R}(s,a) = R(s,a)$.

On the other hand, because drugs are randomly assign, $\pi_0(s,a)$ can be easily esitimated by $\hat{\pi}_0(s,a)$. Let m be the total actions can be token in state s', Then

$$\hat{\pi}_0(s,a) = \frac{1}{m} = \pi_0(s,a).$$

(ii). since $\pi_0(s,a)$ is very complicated, importance sampling is not a good way because it can't give a good estimate on $\hat{\pi}_0(s,a)$.

On the other hand, the interaction is very simple. so $\hat{R}(s,a)$ can be easily estimated so that $\hat{R}(s,a)$ is closed to $R(s,a)$.

3. $f_u(x) = \arg\min_{a \in \mathbb{R}} \|x - au\|_2^2$

$$\|x-au\|_2^2 = (x-au)^T(x-au)$$

$$= x^Tx - ax^Tu - au^Tx + a^2u^Tu$$

$$= x^Tx - 2ax^Tu + a^2 \qquad u^Tu=1$$

$$\frac{\partial}{\partial a} = -2x^Tu + 2a = 0$$

$$a = x^Tu$$

Then $au = (x^Tu)u$

$$\sum_{i=1}^m \|x^{(i)} - (x^{(i)T}u)u\|_2^2 = \sum_{i=1}^m (x^{(i)} - (x^{(i)T}u)u)^T(x^{(i)} - (x^{(i)T}u)u)$$

$$= \sum_{i=1}^m x^{(i)T}x^{(i)} - 2(x^{(i)T}u)^2 + (x^{(i)T}u)^2u^Tu$$

$$= \sum_{i=1}^m x^{(i)T}x^{(i)} - (x^{(i)T}u)^2$$

$$= \sum_{i=1}^m \|x^{(i)}\|_2^2 - \sum_{i=1}^m u^Tx^{(i)}x^{(i)T}u$$

$$= \sum_{i=1}^m \|x^{(i)}\|_2^2 - u^T\Sigma u$$

$$\mathcal{L} = \sum_{i=1}^m \|x^{(i)}\|_2^2 - u^T\Sigma u + \lambda(u^Tu - 1)$$

$$\nabla_u \mathcal{L} = -2\Sigma u + 2\lambda u = 0$$

$$\Sigma u = \lambda u$$

Therefore u is eigenvectors of the covariance matrix.

Thus. it is same as "variance maximizing".

4. (a). we know $g(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$ $\quad g''(z) = g'(z) \cdot (-z)$.

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

$$\nabla \ell(W) = \sum_{i=1}^{n} (W^{-1})^T + \sum_{j=1}^{d} \frac{1}{g'(z)} \cdot g'(z)(-z) \cdot x^{(i)T} = 0$$

$$n(W^{-1})^T = \sum_{i=1}^{n} \begin{bmatrix} w_1^T x^{(i)} \\ \vdots \\ w_d^T x^{(i)} \end{bmatrix} \cdot x^{(i)T}$$

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

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

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

Let R be an orthogonal matrix. then $RR^T = I$.

~~Let Suppose the true W is W_{true} s.t. $W_{true} R$.~~

$\cancel{(W_{true}R)^T = R^T W_{true}, \quad (W_{true}R)^T (W_{true}R) =}$

Then $W_* = WR$, $\quad W_*^T = R^T W^T$

Then $W_* W_*^T = WRR^TW^T = WW^T$.

Since R can be arbitrary orthogonal matrix, then there are multiple solutions for W.

(b). for Laplace distribution, $g'(z) = \frac{1}{2}\exp(-|z|)$, $g''(z) = \frac{1}{2}\exp(-|z|) \cdot (+1)$ when $z<0$

(-1) when $z>0$.

$$\nabla \ell(W) = (W^{-1})^T + \sum_{j=1}^{d} \frac{1}{g'(z)} \cdot g''(z) \cdot x^{(i)T}$$

$$= (W^{-1})^T + \begin{bmatrix} \pm x^{(i)T} \\ \vdots \\ \pm x^{(i)T} \end{bmatrix} \cdot$$

(1) or (-1)

~~when~~ $w_j^T x^{(i)} > 0$, for (-1)

$w_j^T x^{(i)} < 0$ for $(+1)$.

$$\text{Thus. } W := W + \alpha \left(\begin{bmatrix} \pm 1 \\ \vdots \\ \pm 1 \end{bmatrix} x^{(i)T} + (W^T)^{-1} \right).$$

$$W := W + \alpha \left(\begin{bmatrix} 2 \times 1(w_1^T x^{(i)}) - 1 \\ \vdots \\ 2 \times 1(w_d^T x^{(i)}) - 1 \end{bmatrix} x^{(i)T} + (W^T)^{-1} \right)$$

5, (a). since $|B(V_1)_s - B(V_2)_s| \le \|B(V_1) - B(V_2)\|_\infty$ for $\forall s \in S$.

wts. $|B(V_1)_s - B(V_2)_s| \le \gamma \|V_1 - V_2\|_\infty$

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

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

suppose Taking a_1 in V_1 and a_2 in V_2

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

since a_2 is the action that maximize the total reward, any other actions would have less or equal reward.

$$\sum_{s'\in S} P_{sa_1}(s')V_2(s') \le \sum_{s'\in S} P_{sa_2}(s')V_2(s')$$

$$\text{Then } X \le \gamma \left| \sum_{s'\in S} P_{sa_1}(s')V_1(s') - \sum_{s'\in S} P_{sa_1}(s')V_2(s') \right| = Y$$

except the first term is negative and second term is positive.

5 (a) continued.

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

for the case where first term is negative and the second term is positive.

$$X \leq \gamma \left| \sum_{s' \in S} P_{sa_2}(s') V_1(s') - \sum_{s' \in S} P_{sa_2}(s') V_2(s') \right| = Y.$$

Then $= \gamma \left| \sum_{s' \in S} P_{sa_2}(s')(V_1(s') - V_2(s')) \right| \leq \gamma \|V_1 - V_2\|_\infty$

Q.E.D.

(b). Assume there are two fixed points of B, V_1, V_2 s.t. $V_1 \neq V_2$.

Then $B(V_1) = V_1$ and $B(V_2) = V_2$.

from part (a), we know $\|B(V_1) - B(V_2)\|_\infty \leq \gamma \|V_1 - V_2\|_\infty$

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

$$(1-\gamma)\|V_1 - V_2\|_\infty \leq 0$$

since $1-\gamma > 0$ ($\gamma < 1$ is given).

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

Thus $V_1 = V_2$ which contridicts the assumption.

---

6. • around 60 trials.

• plot

• a lot of jitting in some random seed.

They all converge to $\approx$ 4.0 (log of # failures).