1.

$$\Pi_C(y) = \underset{x\in C}{\operatorname{argmin}} \|x-y\|^2$$

$$= \underset{(a,x_2)\in C}{\operatorname{argmin}} \cancel{(a-y_1)^2} + (x_2-y_2)^2$$

$$= \begin{cases} (a,0) & \text{if} \quad y_2<0 \\ (a,y_2) & \text{if} \quad 0\le y_2<1 \\ (a,1) & \text{if} \quad y_2\ge 1 \end{cases}$$

$$= \begin{bmatrix} a \\ \min\{\max\{y_2,0\},\ 1\} \end{bmatrix}$$

$\square$

3.

(a)
$$\log(p_\theta(x)) = \log \mathbb{E}_{Z\sim p_Z}\left[p_\theta(x|Z)\right]$$

$$= \log \mathbb{E}_{Z\sim q_\phi(z|x)}\left[p_\theta(x|Z)\frac{p_Z(Z)}{q_\phi(z|x)}\right]$$

$$= \log \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}_{Z\sim q_\phi(z|x)}\left[p_\theta(x|Z)\frac{p_Z(Z)}{q_\phi(z|x)}\right]$$

$$= \log \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}_{Z_1,\cdots,Z_K\sim q_\phi(z|x)}\left[p_\theta(x|Z_k)\frac{p_Z(Z_k)}{q_\phi(z_k|x)}\right]$$

$$= \log \mathbb{E}_{Z_1,\cdots,Z_K\sim q_\phi(z|x)}\left[\frac{1}{K}\sum_{k=1}^{K} p_\theta(x|Z_k)\frac{p_Z(Z_k)}{q_\phi(z_k|x)}\right]$$

$$\underset{\text{Jensen}}{\ge} \mathbb{E}_{Z_1,\cdots,Z_K\sim q_\phi(z|x)}\left[\log\frac{1}{K}\sum_{k=1}^{K} p_\theta(x|Z_k)\frac{p_Z(Z_k)}{q_\phi(z_k|x)}\right] = \mathrm{VLB}^{(K)}_{\theta,\phi}(x)$$

(b)
$$\mathrm{VLB}^{(K)}_{\theta,\phi}(x) = \mathbb{E}_{Z_1,\cdots,Z_K\sim q_\phi(z|x)}\left[\log\frac{1}{K}\sum_{k=1}^{K} p_\theta(x|Z_k)\frac{p_Z(Z_k)}{q_\phi(z_k|x)}\right]$$

$$= \mathbb{E}_{Z_1,\cdots,Z_K\sim q_\phi(z|x)}\left[\log \mathbb{E}_{I=\{i_1,\cdots i_M\}}\frac{1}{M}\sum_{k=1}^{M} p_\theta(x|Z_{i_k})\frac{p_Z(Z_{i_k})}{q_\phi(z_{i_k}|x)}\right]$$

$$\underset{\text{Jensen}}{\ge} \mathbb{E}_{Z_1,\cdots,Z_K\sim q_\phi(z|x)}\ \mathbb{E}_{I=\{i_1,\cdots i_M\}}\left[\log\frac{1}{M}\sum_{k=1}^{M} p_\theta(x|Z_{i_k})\frac{p_Z(Z_{i_k})}{q_\phi(z_{i_k}|x)}\right]$$

$$= \mathbb{E}_{Z_1,\cdots,Z_M\sim q_\phi(z|x)}\left[\log\frac{1}{M}\sum_{k=1}^{M} p_\theta(x|Z_k)\frac{p_Z(Z_k)}{q_\phi(z_k|x)}\right] = \mathrm{VLB}^{(M)}_{\theta,\phi}$$

(c) q_ϕ is powerful enough.

$$\Leftrightarrow\ D_{KL}\left[q_\phi(\cdot|X_i)\,\|\,p_\theta(\cdot|X_i)\right] = 0$$

Then, $0 \le \log p_\theta(X_i) - \mathrm{VLB}^{(K)}_{\theta,\phi} \le \log p_\theta(X_i) - \mathrm{VLB}^{(1)}_{\theta,\phi} = D_{KL}\left[q_\phi(\cdot|X_i)\,\|\,p_\theta(\cdot|X_i)\right] = 0$

$$\therefore\ \underset{\theta\in\Theta}{\text{maximize}} \sum_{i=1}^{N}\log p_\theta(X_i) = \underset{\theta\in\Theta,\phi\in\Phi}{\text{maximize}} \sum_{i=1}^{N}\mathrm{VLB}^{(K)}_{\theta,\phi}(X_i)$$

$\square$

5.

(a) Define W : event that B win

L : event that B lose

D : event that draw

$$\mathbb{E}_{P_A,P_B}[\text{Points for } B] = P_{P_A,P_B}(W) - P_{P_A,P_B}(L)$$

$$= P(B \text{ plays rocks})P(A \text{ plays Scissors}) + P(B \text{ plays paper})P(A \text{ plays rock}) + P(B \text{ plays scissors})P(A \text{ plays paper})$$
$$- P(B \text{ plays rocks})P(A \text{ plays paper}) - P(B \text{ plays paper})P(A \text{ plays scissors}) - P(B \text{ plays scissors})P(A \text{ plays rock})$$

$$= P(B \text{ plays rocks})\big(P(A \text{ plays Scissors}) - P(A \text{ plays paper})\big)$$
$$+ P(B \text{ plays paper})\big(P(A \text{ plays rock}) - P(A \text{ plays scissors})\big)$$
$$+ P(B \text{ plays scissors})\big(P(A \text{ plays paper}) - P(A \text{ plays rock})\big)$$

$$E_{P_A^*,P_B^*}[\text{Points for } B] = 0.$$

$$E_{P_A^*,P_B}[\text{Points for } B] = 0.$$

$$E_{P_A,P_B^*}[\text{Points for } B] = 0.$$

$\therefore$ $E_{P_A^*,P_B}[\text{Points for } B] \le E_{P_A^*,P_B^*}[\text{Points for } B] \le E_{P_A,P_B^*}[\text{Points for } B]$

$\therefore$ P_A^*, P_B^* is a Solution.

Suppose that one element of Solution $P_B^{*\prime}$ is bigger than other.

WLOG, $P^{*\prime}(B \text{ plays rock})$ is the biggest. Clearly, $P_A^{*\prime} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$.

Then, $E_{P_A^{*\prime},P_B^{*\prime}}[\text{Points for } B] \le E_{P_A^{*\prime},\tilde{P}_B}[\text{Points for } B]$

where $\tilde{P}_B := \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

$\therefore$ $P_A^{*\prime}, P_B^{*\prime}$ is not a solution in this case.

$\therefore$ $P_A^* = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}$, $P_B^* = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}$ is the unique solution.

(b) If $P_B = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$,

$P(A \text{ win} \mid A \text{ plays rock}) = P(B \text{ plays scissors} \mid A \text{ plays rock}) = P(B \text{ plays scissors}) = \frac{1}{3}$

Likewise, $P(A \text{ win} \mid A \text{ plays paper}) = P(A \text{ win} \mid A \text{ plays Scissors}) = \frac{1}{3}$

$$P(A \text{ win}) = P(A \text{ win} \mid A \text{ plays rock})P(A \text{ plays rock}) + P(A \text{ win} \mid A \text{ plays paper})P(A \text{ plays paper}) + P(A \text{ win} \mid A \text{ plays scissors})P(A \text{ plays scissors}) = \frac{1}{3}.$$

It means any strategy is optimal for player A.

$\square$