(1)

(a) $D_{KL}(p\|q) = D_{KL}(p(x,y)\|q(x,y))$

$$= \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{q_1(x)q_2(y)}$$

$$= \sum_x \sum_y p(x,y) \log p(x,y) - \sum_x \sum_y p(x,y) \log q_1(x) q_2(y)$$

To minimize $D_{KL}(p\|q)$, we just need to minimize the right part.

$$-\sum_x \sum_y p(x,y) \log q_1(x) = -\sum_x \left[\left(\sum_y p(x,y)\right) \log q_1(x)\right] = -\sum_x p(x) \log q_1(x) = H(p,q)$$

As $H(p,q) \geq H(p)$, thus $\arg\min_{q_1(x)} -\sum_x \sum_y p(x,y) \log q_1(x) = p(x)$

Similarly, $\arg\min_{q_2(y)} -\sum_x \sum_y p(x,y) \log q_2(y) = p(y)$

Above all, the optimal approximation is a product of marginals.

(b) $D_{KL}(q\|p) = \sum_x \sum_y q(x,y) \log \frac{q(x,y)}{p(x,y)} = \sum_x \sum_y q_1(x) q_2(y) \log \frac{q_1(x) q_2(y)}{p(x,y)}$

For $p(x_i, y_i) = 0$, we need to set $q_1(x_i) q_2(y_i) = 0$ to avoid ∞ in $D_{KL}(q\|p)$, since $\lim_{x \to 0} x \log x = 0$. Therefore,

For $q_1(x_i) q_2(y_i) \log \frac{q_1(x_i) q_2(y_j)}{p(x_i, y_j)} \neq 0$, $\forall i, j$. Then there are 3 scenarios.

1. only $q_1(x_3), q_2(x_3) \neq 0$ 2. $q_1(x_4) q_2(x_4) \neq 0$ only.
3. $q_1(x_3), q_1(x_4), q_2(x_3), q_2(x_4) = 0$

Considering the general case: to minimize D_{KL},

s.t. $\sum_i q_1(x_i) = 1$, $\sum_j q_2(y_j) = 1$

Thus $L = \sum_{ij} q_1(x_i) q_2(y_j) \log \frac{q_1(x_i) q_2(y_j)}{p(x_i, y_j)} + \lambda\left(\sum_i q_1(x_i) - 1\right) + \beta\left(\sum_j q_2(y_j) - 1\right)$

$$\forall i, \quad \frac{\partial L}{\partial q_1(x_i)} = \log q_1(x_i) + \sum_j q_2(y_j) \log q_2(y_j) - \sum_j q_2(y_j) \log p(x_i, y_j) + 1 + \lambda = 0$$

$$\forall j, \quad \frac{\partial L}{\partial q_2(y_j)} = \log q_2(y_j) + \sum_i q_1(x_i) \log q_1(x_i) - \sum_i q_1(x_i) \log p(x_i, y_j) + 1 + \beta = 0$$

$$\frac{\partial L}{\partial \lambda} = \sum_i q_1(x_i) = 0, \qquad \frac{\partial L}{\partial \beta} = \sum_j q_2(y_j) - 1 = 0$$

For case (1), $q_1(x_3) = q_2(y_3) = 1$,

$$D_{KL} = 1 \cdot \log \frac{1}{\frac{1}{4}} = \log 4,$$

For case (2), $q_1(x_4) = q_2(y_4) = 1$,

$$D_{KL} = 1 \cdot \log \frac{1}{\frac{1}{4}} = \log 4$$

For case 3,

$$\begin{cases} \log q_1(x_1) + H(q_{12}) - \log \frac{1}{8} + 1 + \lambda = 0 \\ \log q_1(x_2) + H(q_{12}) - \log \frac{1}{8} + 1 + \lambda = 0 \\ \log q_2(y_1) + H(q_{11}) - \log \frac{1}{8} + 1 + \beta = 0 \\ \log q_2(y_2) + H(q_{11}) - \log \frac{1}{8} + 1 + \beta = 0 \\ q_2(y_1) + q_2(y_2) = 1 \\ q_1(x_1) + q_1(x_2) = 1 \end{cases} \Rightarrow q_1(x_1) = q_1(x_2) = q_2(y_1) = q_2(y_2) = 0.5$$

Thus $\min D_{KL}(q \| p) = \log 2$

(c) If we set $q(x,y) = p(x)p(y)$

$$\Rightarrow D_{KL} = \sum p(x_i) p(y_j) \log \frac{p(x_i)p(y_j)}{p(x_i, y_j)}$$

when $p(x_i, y_j) = 0$, $p(x_i) p(y_j) \neq 0$

$$\Rightarrow D_{KL}(q \| p) \to \infty$$

2. (a) $\Sigma = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}$,

Thus $\Sigma^{-1} = \begin{bmatrix} \frac{4}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{4}{3} \end{bmatrix}$, $|\Sigma| = \frac{3}{4}$

$\therefore f(x_1, x_2) = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu))$

$(x-\mu)^T \Sigma^{-1} (x-\mu) = (x_1-1, x_2-1) \begin{pmatrix} \frac{4}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{4}{3} \end{pmatrix} \begin{pmatrix} x_1-1 \\ x_2-1 \end{pmatrix}$

$= \frac{4}{3}[(x_1-1)^2 - (x_1-1)(x_2-1) + (x_2-1)^2]$

$\therefore f(x_1, x_2) = \frac{1}{2\pi} \cdot \frac{1}{\sqrt{|\Sigma|}} \exp(-\frac{2}{3}[(x_1-1)^2 - (x_1-1)(x_2-1) + (x_2-1)^2])$

$= \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}(x_1-1)^2) \cdot \frac{1}{\sqrt{2\pi}\sqrt{\frac{3}{4}}} \exp(-\frac{1}{2}(x_2-1-\frac{1}{2}(x_1-1)) \cdot \frac{4}{3}(x_2-1-\frac{1}{2}(x_1-1)))$

Thus $f_1(x_1) = \int f(x_1, x_2)\, dx_2 = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}(x_1-1)^2)$

$f(x_2|x_1) = \frac{f(x_1, x_2)}{f(x_1)} = \frac{1}{\sqrt{2\pi}\sqrt{\frac{3}{4}}} \exp(-\frac{1}{2}(x_2-1-\frac{1}{2}(x_1-1))\frac{4}{3}(x_2-1-\frac{1}{2}(x_1-1)))$

In the similar way,

$f(x_1|x_2) = \frac{f(x_1, x_2)}{f(x_2)} = \frac{1}{\sqrt{2\pi}\sqrt{\frac{3}{4}}} \exp(-\frac{1}{2}(x_1-\frac{1}{2}x_2-\frac{1}{2}) \cdot \frac{4}{3} \cdot (x_1-\frac{1}{2}x_2-\frac{1}{2}))$

5. (a) $\because \check{X}$ is white

$\therefore \frac{1}{N} \check{X} \check{X}^T = I$

As $\check{X} = DX \Rightarrow \frac{1}{N} DXX^T D^T = I$

Since XX^T is PSD, thus we can use eigenvalue decomposition for XX^T, i.e. we can write

$XX^T = EVE^T$, E is the orthogonal matrix of eigenvectors of XX^T

V is the diagonal matrix of its eigenvalue $V = diag(V_1, \cdots V_n)$

Let $V^{\frac{1}{2}} = diag(V_1^{\frac{1}{2}}, V_2^{\frac{1}{2}}, \cdots V_n^{\frac{1}{2}})$

$$\therefore \frac{1}{N} DXX^T D^T = \frac{1}{N} DEV^{\frac{1}{2}} V^{\frac{1}{2}} E^T D^T$$
$$= \frac{1}{N} DEV^{\frac{1}{2}} E^{-1} E V^{\frac{1}{2}} E^T D^T$$
$$= \frac{1}{N} D (EV^{\frac{1}{2}} E^T)(EV^{\frac{1}{2}} E^T)^T D^T$$

$$\therefore D = \sqrt{N} E V^{\frac{1}{2}} E^T$$

$$\frac{1}{N} DXX^T D^T = \frac{1}{N} N \cdot (EV^{-\frac{1}{2}} E^T)(EV^{\frac{1}{2}} E^T)(EV^{\frac{1}{2}} E^T)^T (EV^{-\frac{1}{2}} E^T)^T$$
$$= I. \qquad \text{white.}$$