Problem 3.

(1) Let

$$q_\phi(z|x,y) \sim \mathcal{N}(\mu_1, \Sigma_1) \qquad p_\theta(z|y) \sim \mathcal{N}(\mu_2, \Sigma_2).$$

We use part of the results in (2)

$$p(x) = \frac{1}{(2\pi)^{\frac{n}{2}} \det(\Sigma)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right).$$

$$D_{KL}(q_\phi \| p_\theta) = \frac{1}{2}\left[\log\left(\frac{\det \Sigma_2}{\det \Sigma_1}\right) - n + \mathrm{tr}(\Sigma_2^{-1}\Sigma_1) + (\mu_2-\mu_1)^T \Sigma_2^{-1}(\mu_2-\mu_1)\right]$$

$$\mathbb{E}_{q_\phi}[\log p_\theta] = \mathop{\mathbb{E}}_{q_\phi}\left[-\frac{n}{2}\cdot(2\pi) - \frac{1}{2}\log(\det(\Sigma_2)) - \frac{1}{2}(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)\right].$$

$$= \mathop{\mathbb{E}}_{q_\phi}\left[-\frac{n}{2}(2\pi) - \frac{1}{2}\log(\det(\Sigma_2)) - \frac{1}{2}\mathrm{tr}\left(\Sigma_2^{-1}(x-\mu_2)(x-\mu_2)^T\right)\right].$$

$$= \mathbb{E}_{q_\phi}\left[-\frac{n}{2}(2\pi) - \frac{1}{2}\log(\det(\Sigma_2)) - \frac{1}{2}\mathrm{tr}\left(\Sigma_2^{-1}(xx^T - 2x\mu_2^T + \mu_2\mu_2^T)\right)\right]$$

$$= -\frac{n}{2}(2\pi) - \frac{1}{2}\log(\det(\Sigma_2)) - \frac{1}{2}\mathrm{tr}\left(\Sigma_2^{-1}(\Sigma_1 + \mu_1\mu_1^T - 2\mu_1\mu_2^T + \mu_2\mu_2)\right)$$

$$= -\frac{n}{2}(2\pi) - \frac{1}{2}\log(\det(\Sigma_2)) - \frac{1}{2}\mathrm{tr}(\Sigma_2^{-1}\Sigma_1) + \frac{1}{2}(\mu_1-\mu_2)^T\Sigma_2^{-1}(\mu_1-\mu_2).$$

$$\mathbb{E}_{q_\phi} - D_{KL} = -\frac{n}{2}(2\pi) + \frac{n}{2} - \log(\det(\Sigma_2)) + \frac{1}{2}\log(\det(\Sigma_1)) - \mathrm{tr}(\Sigma_2^{-1}\Sigma_1)$$

$$\log p_\theta = -\frac{n}{2}(2\pi) - \frac{1}{2}\log(\det(\Sigma_2)) - \frac{1}{2}(x-\mu_2)\Sigma_2^{-1}(x-\mu_2).$$

$$\Rightarrow \quad \log p_\theta \geq \mathbb{E}_{q_\phi}[\log p_\theta] - D_{KL}(q_\phi \| p_\theta).$$

2. We directly cite John Duchi's lecture notes.

Assume P_1, P_2 Gaussian vectors.

$$P_1 \sim N(\mu_1, \Sigma_1) \qquad P_2 \sim N(\mu_2, \Sigma_2).$$

$$D_{KL}(P_1 \| P_2) = \mathbb{E}_{P_1}[\log P_1 - \log P_2]$$

$$= \frac{1}{2} \mathbb{E}_{P_1}[-\log(\det \Sigma_1) - (x-\mu_1)^T \Sigma_1^{-1}(x-\mu_1) + \log(\det \Sigma_2) + (x-\mu_2)^T \Sigma_2^{-1}(x-\mu_2)]$$

This is from the density

$$p(x) = \frac{1}{(2\pi)^{\frac{n}{2}} \det(\Sigma)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right).$$

Then

$$D_{KL}(P_1 \| P_2) = \frac{1}{2}\log\left(\frac{\det \Sigma_2}{\det \Sigma_1}\right) + \frac{1}{2}\mathbb{E}_{P_1}\left[-\operatorname{tr}\left(\Sigma_1^{-1}(x-\mu_1)(x-\mu_1)^T\right) + \operatorname{tr}\left(\Sigma_2^{-1}(x-\mu_2)(x-\mu_2)^T\right)\right]$$

$$= \frac{1}{2}\log\left(\frac{\det \Sigma_2}{\det \Sigma_1}\right) + \frac{1}{2}\mathbb{E}_{P_1}\left[-\operatorname{tr}(\Sigma_1^{-1}\Sigma_1) + \operatorname{tr}\left(\Sigma_2^{-1}(xx^T - 2x\mu_2^T + \mu_2\mu_2^T)\right)\right]$$

$$= \frac{1}{2}\log\left(\frac{\det \Sigma_2}{\det \Sigma_1}\right) - \frac{1}{2}n + \frac{1}{2}\operatorname{tr}\left[\Sigma_2^{-1}\left(\Sigma_1 + \mu_1\mu_1^T - 2\mu_2\mu_1^T + \mu_2\mu_2^T\right)\right]$$

$$= \frac{1}{2}\left\{\log\left(\frac{\det \Sigma_2}{\det \Sigma_1}\right) - n + \operatorname{tr}(\Sigma_2^{-1}\Sigma_1) + \operatorname{tr}\left(\mu_1^T \Sigma_2^{-1}\mu_1 - 2\mu_1^T \Sigma_2^{-1}\mu_2 + \mu_2^T \Sigma_2^{-1}\mu_2\right)\right\}$$

$$= \frac{1}{2}\left[\log\left(\frac{\det \Sigma_2}{\det \Sigma_1}\right) - n + \operatorname{tr}(\Sigma_2^{-1}\Sigma_1) + (\mu_2 - \mu_1)^T \Sigma_2^{-1}(\mu_2 - \mu_1)\right].$$

Now take $\Sigma_1 = \operatorname{diag}(\sigma_1^2, \cdots, \sigma_J^2)$ $\mu_1 = (\mu_1 \cdots \mu_J)$

$\Sigma_2 = I$ $\mu_2 = (0, \cdots, 0)$.

$$D_{KL}(P_1 \| P_2)$$

$$= \frac{1}{2}\left(\log \frac{1}{\prod_{i=1}^{J} \sigma_i^2} - J + \operatorname{tr}(\Sigma_1) + \mu_1^T \mu_1 \right).$$

$$= \frac{1}{2}\left[-\sum_{i=1}^{J} \log \sigma_i^2 - \sum_{i=1}^{J} 1 + \sum_{i=1}^{J} \sigma_i^2 + \sum_{i=1}^{J} \mu_i^2 \right].$$

$$= -\frac{1}{2} \sum_{j=1}^{J} \left(1 + \log(\sigma_j^2) - \mu_j^2 - \sigma_j^2 \right).$$