1. (a) $\log x \le x-1 \Rightarrow -\log x \ge 1-x \Rightarrow \log y \ge 1-\frac{1}{y}$. 取等当且仅当 $y=1$.

用 $\frac{P(x)}{Q(x)}$ 替换 y, $\log\frac{P(x)}{Q(x)} \ge 1-\frac{Q(x)}{P(x)}$. 由 $P(x)\ge 0$, 有 $P(x)\log\frac{P(x)}{Q(x)} \ge P(x)-Q(x)$

对 x 求和, $D_{KL}(P\|Q)=\sum_x P(x)\log\frac{P(x)}{Q(x)} \ge \sum_x (P(x)-Q(x)) = 1-1=0$

取等当且仅当 $\frac{P(x)}{Q(x)}=1$, 即 $P(X)\equiv Q(X)$

(b)
$$
\begin{aligned}
D_{KL}(P(X,Y)\|Q(X,Y)) &= \sum_{x,y} P(x,y)\log\frac{P(x,y)}{Q(x,y)} = \sum_{x,y}P(x,y)\log\frac{P(x)P(y|x)}{Q(x)Q(y|x)} \\
&= \sum_{x,y}P(x,y)\left(\log\frac{P(x)}{Q(x)}+\log\frac{P(y|x)}{P(y|x)}\right) = \sum_{x,y}P(x,y)\log\frac{P(x)}{Q(x)}+\sum_{x,y}P(x,y)\log\frac{P(y|x)}{Q(y|x)} \\
&= \sum_x\left(\sum_y P(x,y)\right)\log\frac{P(x)}{Q(x)}+\sum_x P(x)\left(\sum_y P(y|x)\right)\log\frac{P(y|x)}{Q(y|x)} \\
&= \sum_x P(x)\log\frac{P(x)}{Q(x)}+\left(\sum_x P(x)\right)\left(\sum_y P(y|x)\log\frac{P(y|x)}{Q(y|x)}\right) \\
&= D_{KL}(P(X)\|Q(X))+D_{KL}(P(Y|X)\|Q(Y|X)).
\end{aligned}
$$

(c)
$$
\begin{aligned}
\arg\min_\theta D_{KL}(\hat{P}\|P_\theta) &= \arg\min_\theta \sum_x \hat{P}(x)\log\frac{\hat{P}(x)}{P_\theta(x)} = \arg\min_\theta \sum_{i=1}^n \frac{1}{n}\log\frac{\frac{1}{n}}{P_\theta(x^{(i)})} \\
&= \arg\min_\theta \frac{1}{n}\sum_{i=1}^n\left(-\log n-\log P_\theta(x^{(i)})\right) \\
&= \arg\min_\theta\left(-\log n-\frac{1}{n}\sum_{i=1}^n\log P_\theta(x^{(i)})\right) \\
&= \arg\min_\theta\left(-\frac{1}{n}\sum_{i=1}^n\log P_\theta(x^{(i)})\right) \\
&= \arg\max_\theta \sum_{i=1}^n \log P_\theta(x^{(i)})
\end{aligned}
$$

2. (a) 记 $L(\theta,Q)=\sum_{i=1}^n\sum_{z^{(i)}}Q_i(z^{(i)})\log\frac{p(x^{(i)},z^{(i)};\theta)}{Q_i(z^{(i)})}+\alpha\sum_{i=1}^{\hat{n}}\log p(\hat{x}^{(i)},\hat{z}^{(i)};\theta)$

由 Jensen Ineq, $\log\sum_{z^{(i)}}p(x^{(i)},z^{(i)};\theta)=\log\sum_{z^{(i)}}\log Q_i(z^{(i)})\frac{p(x^{(i)},z^{(i)},\theta)}{Q_i(z^{(i)})}\ge\sum_{z^{(i)}}Q_i(z^{(i)})\log\frac{p(x^{(i)},z^{(i)};\theta)}{Q_i(z^{(i)})}$

$\Rightarrow \ell_{unsup}(\theta)=\sum_{i=1}^n\log\sum_{z^{(i)}}p(x^{(i)},z^{(i)};\theta)\ge\sum_{i=1}^n\sum_{z^{(i)}}Q_i(z^{(i)})\log\frac{p(x^{(i)},z^{(i)};\theta)}{Q_i(z^{(i)})}$

$\Rightarrow \ell_{semi\text{-}sup}(\theta)=\ell_{unsup}(\theta)+\alpha\ell_{sup}(\theta)\ge L(\theta,Q)$ (*)

E-step 中 $Q_i^{(t)}(z^{(i)})=p(z^{(i)}|x^{(i)};\theta^{(t)})$, 此时

$$
\begin{aligned}
\mathcal{L}(\theta^{(t)},Q_i^{(t)}) &= \sum_{i=1}^n\sum_{z^{(i)}}p(z^{(i)}|x^{(i)};\theta^{(t)})\log\frac{p(x^{(i)},z^{(i)};\theta^{(t)})}{p(z^{(i)}|x^{(i)};\theta^{(t)})}+\alpha\sum_{i=1}^{\hat{n}}\log p(\hat{x}^{(i)},\hat{z}^{(i)};\theta^{(t)}) \\
&= \sum_{i=1}^n\Bigg(\underbrace{\sum_{z^{(i)}}p(z^{(i)}|x^{(i)};\theta^{(t)})\log p(x^{(i)},z^{(i)};\theta^{(t)})}_{\mathbb{E}_{p(z|x;\theta^{(t)})}(\log p(x,z;\theta^{(t)}))}-\underbrace{\sum_{z^{(i)}}p(z^{(i)}|x^{(i)},\theta^{(t)})\log p(z^{(i)}|x^{(i)};\theta^{(t)})}_{-H(p(z|x;\theta^{(t)}))}\Bigg) \\
&\quad+\alpha\sum_{i=1}^{\hat{n}}\log p(\hat{x}^{(i)},\hat{z}^{(i)};\theta^{(t)}) \\
&= \sum_{i=1}^n\Bigg(\sum_{z^{(i)}}p(z^{(i)}|x^{(i)};\theta^{(t)})\log p(z^{(i)}|x^{(i)};\theta^{(t)})+\log p(x^{(i)};\theta^{(t)}) \\
&\quad-\sum_{z^{(i)}}p(z^{(i)}|x^{(i)};\theta^{(t)})\log p(z^{(i)}|x^{(i)};\theta^{(t)})\Bigg)+\alpha\sum_{i=1}^{\hat{n}}\log p(\hat{x}^{(i)},\hat{z}^{(i)};\theta^{(t)}) \\
&= \sum_{i=1}^n\log p(x^{(i)};\theta(t))+\alpha\sum_{i=1}^{\hat{n}}\log p(\hat{x}^{(i)},\hat{z}^{(i)},\theta^{(t)})=\ell_{semi\text{-}sup}(\theta^{(t)})
\end{aligned}
$$

即 $\mathcal{L}(\theta^{(t)},Q^{(t)})=\ell_{unsup}(\theta)+\alpha\ell_{sup}(\theta)=\ell_{semi\text{-}sup}(\theta^{(t)})$

M-step 中 $\theta^{(t+1)}=\arg\max\mathcal{L}(\theta,Q^{(t)}) \Rightarrow \mathcal{L}(\theta^{(t+1)},Q^{(t)})\ge\mathcal{L}(\theta^{(t)},Q^{(t)})$

在(*)中取 $Q=Q^{(t)}$, $\theta=\theta^{(t+1)}$. $\ell_{semi\text{-}sup}(\theta^{(t+1)})\ge\mathcal{L}(\theta^{(t+1)},Q^{(t)})\ge\mathcal{L}(\theta^{(t)},Q^{(t)})=\ell_{semi\text{-}sup}(\theta^{(t)})$

(b) 对未标记数据, $w_j^{(i)}=p(z_j|x^{(i)};\theta)$. $\theta=\{\mu_j,\Sigma_j,\phi_j\}_{j=1}^k$. 利用 Bayes Thm.

$$
w_j^{(i)}=p(z_j|x^{(i)};\theta)=\frac{p(x^{(i)}|z_j;\theta)p(z_j;\theta)}{\sum_{\ell=1}^k p(x^{(i)}|z_\ell;\theta)p(z_\ell;\theta)}=\frac{\phi_j N(x^{(i)};\mu_j,\Sigma_j)}{\sum_{\ell=1}^k\phi_\ell N(x^{(i)},\mu_\ell,\Sigma_\ell)}
$$

其中 $N(x;\mu_j,\Sigma_j)=\frac{1}{\sqrt{(2\pi)^d\det\Sigma_j}}e^{-\frac{1}{2}(x-\mu_j)^T\Sigma_j^{-1}(x-\mu_j)}$

对已标记数据, $w_j=\mathbb{1}_{j=z^{(i)}}$.

(c) 由于 $p(x^{(i)},z_j;\theta)=\phi_j N(x^{(i)};\mu_j,\Sigma_j)$, $p(\hat{x}^{(i)},\hat{z}^{(i)};\theta)=\phi_{\hat{z}^{(i)}}N(x^{(i)};\mu_{\hat{z}^{(i)}},\Sigma_{\hat{z}^{(i)}})$

$$
\begin{aligned}
\text{故 }\mathcal{L}(\theta,Q_t^{(i)}) &= \sum_{i=1}^n\sum_{j=1}^k w_j^{(i)}\left(\log\phi_j+\log N(x^{(i)},\mu_j,\Sigma_j)\right)+\alpha\sum_{i=1}^{\hat{n}}\left(\log\phi_{\hat{z}^{(i)}}+\log N(\hat{x}^{(i)};\mu_{\hat{z}^{(i)}},\Sigma_{\hat{z}^{(i)}})\right) \\
&= \sum_{i=1}^n w_j^{(i)}\log\phi_j+\alpha\sum_{i=1}^n\mathbb{1}_{j=z^{(i)}}\log\phi_j+\Bigg(\sum_{i=1}^n w_j^{(i)}\log N(x^{(i)};\mu_j,\Sigma_j)+\alpha\sum_{i=1}^{\hat{n}}\mathbb{1}_{j=z^{(i)}} \\
&\quad\log N(x^{(i)};\mu_j,\Sigma_j)\Bigg)\Bigg|_{\mu_j}+\left(\sum_{i=1}^n w_j^{(i)}\log N(x^{(i)},\mu_j,\Sigma_j)+\alpha\sum_{i=1}^N\mathbb{1}_{\hat{z}^{(i)}=j}\log N(\hat{x}^{(i)},\mu_j,\Sigma_j)\right)\Bigg|_{\Sigma_j} \\
&:= \mathcal{L}_{\phi_j}+\mathcal{L}_{\mu_j}+\mathcal{L}_{\Sigma_j}\quad(\text{由于可分离变量, 故可分别求解})
\end{aligned}
$$

令 $\hat{\mathcal{L}}_{\phi_j}=\mathcal{L}_{\phi_j}+\lambda(1-\sum_{j=1}^k\phi_j)=\sum_{i=1}^n w_j^{(i)}\log\phi_j+\alpha\sum_{i=1}^n\mathbb{1}_{j=z^{(i)}}\log\phi_j+\lambda(1-\sum_{j=1}^k\phi_j)$

$$
\frac{\partial\hat{L}_{\phi_j}}{\partial\phi_j}=\frac{1}{\phi_j}\left(\sum_{i=1}^n w_j^{(i)}+\alpha\sum_{i=1}^{\hat{n}}\mathbb{1}_{\hat{z}^{(i)}=j}\right)-\lambda=0 \Rightarrow \phi_j=\lambda\left(\sum_{i=1}^n w_j^{(i)}+\alpha\sum_{i=1}^{\hat{n}}\mathbb{1}_{z^{(i)}=j}\right)
$$

又 $1=\sum_{j=1}^k\phi_j=\lambda\sum_{j=1}^n\sum_{i=1}^n\left(\sum_{i=1}^n w_j^{(i)}+\alpha\sum_{i=1}^{\hat{n}}\mathbb{1}_{z^{(i)}=j}\right)=\lambda(n+\alpha\hat{n})$. 故 $\lambda=\frac{1}{n+\alpha\hat{n}}$

此时 $\phi_j=\frac{\sum_{i=1}^n w_j^{(i)}+\sum_{i=1}^{\hat{n}}\mathbb{1}_{z^{(i)}=j}}{n+\alpha\hat{n}}=\phi_j^{(t+1)}$.

$$
\begin{aligned}
\mathcal{L}_{\mu_j} &= \left(\sum_{i=1}^n w_j^{(i)}\log N(x^{(i)};\mu_j,\Sigma_j)+\alpha\sum_{i=1}^{\hat{n}}\mathbb{1}_{z^{(i)}=j}\log N(\hat{x}^{(i)};\mu_j,\Sigma_j)\right)\Bigg|_{\mu_j} \\
&= -\frac{1}{2}\sum_{i=1}^n w_j^{(i)}(x^{(i)}-\mu_j)^T\Sigma_j^{-1}(x^{(i)}-\mu_j)-\frac{\alpha}{2}\sum_{i=1}^{\hat{n}}\mathbb{1}_{z^{(i)}=j}(\hat{x}^{(i)}-\mu_j)^T\Sigma_j^{-1}(\hat{x}^{(i)}-\mu_j^T)
\end{aligned}
$$

$$
\frac{\partial\mathcal{L}_{\mu_j}}{\partial\mu_j}=\sum_{i=1}^n w_j^{(i)}\Sigma_j^{-1}(x^{(i)}-\mu_j)+\alpha\sum_{i=1}^{\hat{n}}\mathbb{1}_{z^{(i)}=j}\Sigma_j^{-1}(\hat{x}^{(i)}-\mu_j)=0
$$

$$
\Rightarrow\left(\sum_{i=1}^n w_j^{(i)}+\alpha\sum_{i=1}^{\hat{n}}\mathbb{1}_{z^{(i)}=j}\right)\mu_j=\sum_{i=1}^n w_j^{(i)}x^{(i)}+\alpha\sum_{i=1}^{\hat{n}}\mathbb{1}_{z^{(i)}=j}\hat{x}^{(i)}. \Rightarrow \mu_j=\frac{\sum_{i=1}^n w_j^{(i)}x^{(i)}+\alpha\sum_{i=1}^{\hat{n}}\mathbb{1}_{z^{(i)}=j}\hat{x}^{(i)}}{\sum_{i=1}^n w_j^{(i)}+\alpha\sum_{i=1}^{\hat{n}}\mathbb{1}_{z^{(i)}=j}}=\mu_j^{(t+1)}
$$

$$
\mathcal{L}_{\Sigma_j}=-\frac{1}{2}\sum_{i=1}^n w_j^{(i)}\left(\log\det\Sigma_j+(x^{(i)}-\mu_j)^T\Sigma_j^{-1}(x^{(i)}-\mu_j)\right)-\frac{\alpha}{2}\sum_{i=1}^{\hat{n}}\mathbb{1}_{z^{(i)}=j}\left(\log\det\Sigma_j+(\hat{x}^{(i)}-\mu_j)^T\Sigma_j^{-1}(\hat{x}^{(i)}-\mu_j)\right)
$$

$$
\frac{\partial L_{\Sigma_j}}{\partial\Sigma_j}=-\frac{1}{2}\sum_{i=1}^n w_j^{(i)}\left(\Sigma_j^{-1}-\Sigma_j^{-1}(x^{(i)}-\mu_j)(x^{(i)}-\mu_j)^T\Sigma_j^{-1}\right)-\frac{\alpha}{2}\sum_{i=1}^{\hat{n}}\mathbb{1}_{z^{(i)}=j}\left(\Sigma_j^{-1}-\Sigma_j^{-1}(\hat{x}^{(i)}-\mu_j)(\hat{x}^{(i)}-\mu_j)^T\Sigma_j^{-1}\right)=0
$$

左右同乘 Σ_j, $-\frac{1}{2}\sum_{i=1}^n w_j^{(i)}\left(\Sigma_j-(x^{(i)}-\mu_j)(x^{(i)}-\mu_j)^T\right)-\frac{\alpha}{2}\sum_{i=1}^{\hat{n}}\mathbb{1}_{z^{(i)}=j}\left(\Sigma_j-(\hat{x}^{(i)}-\mu_j)(\hat{x}^{(i)}-\mu_j)^T\right)=0$

$$
\Rightarrow\Sigma_j=\frac{\sum_{i=1}^n w_j^{(i)}(x^{(i)}-\mu_j)(x^{(i)}-\mu_j)^T+\alpha\sum_{i=1}^{\hat{n}}\mathbb{1}_{z^{(i)}=j}(\hat{x}^{(i)}-\mu_j)(\hat{x}^{(i)}-\mu_j)^T}{\sum_{i=1}^n w_j^{(i)}+\sum_{i=1}^{\hat{n}}\mathbb{1}_{z^{(i)}=j}}=\Sigma_j^{(t+1)}
$$

3. (a) $L_0(\mu)=4^2+2^2+(\sqrt{5})^2+(\sqrt{5})^2=30$.

(b) Store location: $(-4,3)$, $(4,0)$. $L(\mu)=4$.

(c) $L_C(\mu)=\sum_{i=1}^r C^{(i)}\min_{1\le j\le k}\|x^{(i)}-\mu^{(j)}\|^2$.

(d) $L_C(\mu)=\sum_{i=1}^2 C^{(i)}(x^{(i)}-\mu)^2=10(-1-\mu)^2+4(1-\mu)^2=14\mu^2+12\mu+14=14(\mu+\frac{3}{7})^2+\frac{560}{49}$

故应位于 $\mu=-\frac{3}{7}$ 处.

(e) $L_C(\mu)=\sum_{i=1}^r C^{(i)}\min_{1\le j\le k}\|x^{(i)}-\mu^{(j)}\|^2=\sum_{i=1}^r C^{(i)}\|x^{(i)}-\mu^{(y^{(i)})}\|^2$

$L_{DC}(\mu)=\sum_{j=1}^k\hat{C}^{(j)}\|\mu^{(j)}-x_{DC}\|^2=\sum_{j=1}^k\left(\sum_{i=1}^r C^{(i)}\mathbb{1}_{y^{(i)}=j}\right)\|\mu^{(j)}-x_{DC}\|^2$.

$L_S(\mu)=\sum_{i=1}^r C^{(i)}\|x^{(i)}-\mu^{(y^{(i)})}\|^2+\sum_{j=1}^k\left(\sum_{i=1}^r C^{(i)}\mathbb{1}_{y^{(i)}=j}\right)\|\mu^{(j)}-x_{DC}\|^2$

(f) $L_{DC}(\mu)=\left(\sum_{j=1}^2 C^{(i)}\right)(\mu-x_{DC})^2=14(\mu-10)^2=14\mu^2-280\mu+1400$

$L_S(\mu)=L_C(\mu)+L_{DC}(\mu)=28\mu^2-268\mu+1414=28(\mu-\frac{67}{14})^2+\frac{5409}{7}$

故应位于 $\mu=\frac{67}{14}$ 处

4. (a) 显然对 $t=1$, $q(x_1|x_0)=N(x_1;\sqrt{1-\beta_1}x_0;\beta_1 I)=N(x_1;\sqrt{\bar{\alpha}_1}x_0,(1-\bar{\alpha}_1)I)$

若对某一 $k\in\mathbb{N}_+$ 成立 $q(x_k|x_0)=N(x_k;\sqrt{\bar{\alpha}_k}x_0,(1-\bar{\alpha}_k)I)$

由 $x_{k+1}=\sqrt{1-\beta_{k+1}}x_k+\sqrt{\beta_{k+1}}\varepsilon_{k+1}=\sqrt{\alpha_{k+1}}x_k+\sqrt{1-\alpha_{k+1}}\varepsilon_{k+1}$. $\varepsilon_{k+1}\sim N(0,I)$

有 $x_{k+1}\sim N(\mu_{k+1},\Sigma_{k+1})$. $\mu_{k+1}=\mathbb{E}(x_{k+1}|x_0)$, $\Sigma_{k+1}=Var(x_{k+1}|x_0)$

$\mathbb{E}(x_{k+1}|x_0)=\mathbb{E}(\sqrt{\alpha_{k+1}}x_k+\sqrt{1-\alpha_{k+1}}\varepsilon_{k+1})=\mathbb{E}(\sqrt{\alpha_{k+1}}\sqrt{\bar{\alpha}_k}x_0)+0=\sqrt{\bar{\alpha}_{k+1}}x_0$

$$
\begin{aligned}
Var(x_{k+1}|x_0) &= Var(\sqrt{\alpha_{k+1}}x_k+\sqrt{1-\alpha_{k+1}}\varepsilon_{k+1})=\alpha_{k+1}Var(x_k)+(1-\alpha_{k+1})Var(\varepsilon_{k+1}) \\
&= \alpha_{k+1}(1-\bar{\alpha}_k I)+(1-\alpha_{k+1})I=(1-\alpha_{k+1}\bar{\alpha}_k)I=(1-\bar{\alpha}_{k+1})I
\end{aligned}
$$

故 $x_{k+1}\sim N(\sqrt{\bar{\alpha}_{k+1}}x_0,(1-\bar{\alpha}_{k+1})I)$, 即 $q(x_{k+1}|x_0)=N(x_{k+1};\sqrt{\bar{\alpha}_{k+1}}x_0,(1-\bar{\alpha}_{k+1})I)$

归纳即得 $\forall 1\le t\le T$, $q(x_t|x_0)=N(x_t;\sqrt{\bar{\alpha}_t}x_0,(1-\bar{\alpha}_t)I)$.

(b) $\log p(x_0)=\log\int p(x_{0:T})dx_{1:T}=\log\mathbb{E}_{q(x_{1:T}|x_0)}\frac{p(x_{0:T})}{q(x_{1:T}|x_0)}\ge\mathbb{E}_{q(x_{1:T}|x_0)}\log\frac{p(x_{0:T})}{q(x_{1:T}|x_0)}:=\text{ELBO}$

其中倒数第二个不等式使用了 Jensen Ineq.

$$
\begin{aligned}
\text{而 ELBO} &= \mathbb{E}_{q(x_{1:T}|x_0)}\left(\log p(x_{0:T})-\log q(x_{1:T}|x_0)\right) \\
&= \mathbb{E}_{q(x_{1:T}|x_0)}\left(\log\left(p(x_T)\prod_{t=1}^T p_\theta(x_{t-1}|x_t)\right)-\log\prod_{t=1}^T q(x_t|x_{t-1})\right) \\
&= \mathbb{E}_{q(x_{1:T}|x_0)}\left(\log p(x_T)+\sum_{t=1}^T\log p_\theta(x_{t-1}|x_t)-\sum_{t=1}^T\log q(x_t|x_{t-1})\right) \\
&= \mathbb{E}_{q(x_{1:T}|x_0)}\log p(x_T)+\sum_{t=1}^T\mathbb{E}_{q(x_{1:T}|x_0)}\left(\log p(x_{t-1}|x_t)-\log q(x_t|x_{t-1})\right) \\
&= \log p(x_0)-\sum_{k=1}^t\mathbb{E}_{q(x_{1:T}|x_0)}\log\frac{q(x_t|x_{t-1})}{p(x_{t-1}|x_t)} \\
&= \log p(x_0)-\sum_{k=1}^n\int_\Omega q(x_t|x_{t-1})\log\frac{q(x_t|x_{t-1})}{p(x_{t-1}|x_t)} \\
&= \log p(x_0)-\sum_{k=1}^n D_{KL}(q(x_t|x_{t-1})\|p(x_{t-1}|x_t)) \\
&= \log p(x_0)-D_{KL}(q(x_{1:T}|x_0)\|p(x_{1:T}|x_0))
\end{aligned}
$$

$\Rightarrow \log p(x_0)=\text{ELBO}+D_{KL}(q(x_{1:T}|x_0)\|p(x_{1:T}|x_0))$ 即为所求.

(c)
$$
\begin{aligned}
\arg\max_\theta\text{ ELBO} &= \arg\max_\theta\left(\mathbb{E}_{q(x_1|x_0)}\log p_\theta(x_0|x_1)-\sum_{t=2}^T\mathbb{E}_{q(x_t|x_0)}D_{KL}(q(x_{t-1}|x_t,x_0)\|p_\theta(x_{t-1}|x_t))\right) \\
&= \arg\min_\theta\sum_{t=1}^T\mathbb{E}_{q(x_t|x_0)}D_{KL}(q(x_{t-1}|x_t,x_0)\|p_\theta(x_{t-1}|x_t))
\end{aligned}
$$

由 $q(x_{t-1}|x_t,x_0)=N(x;\mu_q(x_t,x_0),\sigma_q^2 I)$, $p_\theta(x_{t-1}|x_t)=N(x;\mu_\theta(x_t,t),\sigma_q^2 I)$

$$
\begin{aligned}
\text{有 }D_{KL}(q(x_{t-1}|x_t,x_0)\|p_\theta(x_{t-1}|x_t)) &= \frac{d}{2}+\frac{1}{2}(\mu_\theta(x_t,t)-\mu_q(x_t,x_0))^T(\sigma_q^2 I)^{-1}(\mu_\theta(x_t,t)-\mu_q(x_t,x_0)) \\
&= \frac{d}{2}+\frac{1}{2\sigma_q^2}\|\mu_\theta(x_t,t)-\mu_q(x_t,x_0)\|_2^2
\end{aligned}
$$

又由 $\mu_\theta(x_t,t)=\frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})x_t+\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)\hat{x}_\theta(x_t,t)}{1-\bar{\alpha}_t}$, $\mu_q(x_t,x_0)=\frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})x_t+\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)x_0}{1-\bar{\alpha}_t}$

故 $\|\mu_\theta(x_t,t)-\mu_q(x_t,x_0)\|_2^2=\left\|\frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{1-\bar{\alpha}_t}(\hat{x}_\theta(x_t,t)-x_0)\right\|_2^2=\frac{\bar{\alpha}_{t-1}(1-\alpha_t)^2}{(1-\bar{\alpha}_t)^2}\|\hat{x}_\theta(x_t,t)-x_0\|_2^2$

$$
\begin{aligned}
\Rightarrow\arg\max_\theta\text{ ELBO} &= \arg\min_\theta\sum_{t=1}^T\mathbb{E}_{q(x_t|x_0)}D_{KL}(q(x_{t-1}|x_t,x_0)\|p_\theta(x_{t-1}|x_t)) \\
&= \arg\min_\theta\sum_{t=1}^T\mathbb{E}_{q(x_t|x_0)}\left(\frac{d}{2}+\frac{1}{2\sigma_q^2}\|\mu_\theta(x_t,t)-\mu_q(x_t,x_0)\|_2^2\right) \\
&= \arg\min_\theta\frac{1}{T}\sum_{t=1}^T\left(\frac{d}{2}+\mathbb{E}_{q(x_t|x_0)}\frac{\bar{\alpha}_{t-1}(1-\alpha_t)^2}{2\sigma_q^2(1-\bar{\alpha}_t)^2}\|\hat{x}_\theta(x_t,t)-x_0\|_2^2\right) \\
&= \arg\min_\theta\mathbb{E}_{t\sim U[1,T],q(x_t|x_0)}\frac{\bar{\alpha}_{t-1}(1-\alpha_t)^2}{2\sigma_q^2(1-\bar{\alpha}_t)^2}\|\hat{x}_\theta(x_t,t)-x_0\|_2^2.
\end{aligned}
$$