

1. Please derive the KL term as a close-form function of  $\theta'$  that can be used as a training objective for back propagation.

對於 VAE 中的 KL divergence, 假設  $q(z|x; \theta')$  的平均數為  $\mu(x)$ , 共變異數為  $\Sigma(x)$  的高斯分佈, 且當  $p(z)$  為標準常態分佈時, KL divergence 有 close-form:

$$KL(q(z|x; \theta') || p(z))$$

$$= \frac{1}{2} * [\text{tr}(\Sigma(x)) + \mu(x)^T \mu(x) - k - \log |\Sigma(x)|]$$

其中  $\text{tr}(\Sigma(x))$  為共變異數矩陣的 trace

$K$  為潛在空間的維度

$|\Sigma(x)|$  為共變異數矩陣的行列式

對於對角共變異數矩陣，可簡化為

$$KL = 0.5 * \sum_i [\sigma_i^2 + \mu_i^2 - 1 - \log(\sigma_i^2)]$$

這個 close-form 可直接用於 back propagation

因他對  $\mu(x)$  和  $\Sigma(x)$  都是可微的

2.

① GMM 做為 prior 是可行的

② 挑戰：

(i) KL divergence 不再有簡單的 close-form

，計算變更複雜

(ii) back propagation 變更困難，需要近似或採樣

(iii) 訓練穩定度降低，容易陷入局部最佳

解決方案：

(i) 使用對抗性方法，強制聚合 posterior 分佈匹配 GMM prior.

(ii) 採用 variational lower bound 的蒙地卡羅估計

(iii) 使用重要性採樣技術

(IV) 使用分層 VAE 的結構，頂層處理

# 混合成分的選擇