

A3-QA

Q1

- 在分配的步骤中，对于每个样本点 x 都将其分配到最近的质心所属的簇中，这样可以保证对于每个点 $\|x - c_x\|$ 都是最优的，即：

$$\|x - c_{old}\| \geq \|x - c_{new}\|$$

- 在更新步骤中，新的质心计算为

$$c_i = \frac{1}{|G_i|} \sum_{x \in G_i} x$$

由求导可以知道，这样的计算方式正是最小化 $\sum_{x \in G_i} \|x - c\|^2$ 的解析解。因此，这一步同样对于目标函数的值是不增的

- 因为目标函数有清晰的下界 $J \geq 0$ ，且函数单调递减，因此一定收敛
- 综上所述，目标函数是严格递减的且可以在有限步内收敛

Q2

1)

- 相同点：当自编码器采用题目结构时，学到的映射恰好会把数据投影到主成分子空间。从目标函数上看，自编码器要最小化

$$J = \sum_{n=1}^N \|x_n - D(E(n_n))\|^2$$

最优解满足 $W_d W_e$ 就是对数据做最优线性投影的矩阵，而这正是PCA投影到前 k 个主成分的算子

- 不同点：PCA中得到的是一组正交单位向量，而自编码器里的 W_e, W_d 可能并非正交，但它们张成的子空间与PCA是完全一致的

2)

- VAE采用的随即映射，但传统自编码器采用的确定性映射

VAE的损失函数采用重构误差和KL散度，但传统的采用普通误差 $\|x - \hat{x}\|^2$

VAE的潜变量分布近似服从先验 $p(z)$ ，而传统的没有约束

- 连续性：VAE 通过 KL 三度强制 $q(z|x)$ 与简单先验靠拢，保证了潜空间的连贯性

- 可解释性：因为直到先验分布的形式，可以在潜空间中有方向地移动以改变生成样本的属性

Q3

令 $q(z)$ 是我们任意指定的分布

$$\therefore \ln p(x) = \int q(z) \ln p(x) dz$$

由后验分布 $p(z|x) = \frac{p(x,z)}{p(x)}$

$$\therefore \ln p(x) = \int q(z) [\ln p(x,z) - \ln p(z|x)] dz$$

$$= \int q(z) [\ln p(x,z) - \ln q(z)] dz + \int q(z) [\ln q(z) - \ln p(z|x)] dz$$

$$= \mathcal{L}(q) + KL(q||p)$$

Q4

1)

- 优势：
 - 生成质量与稳定性：扩散模型通过逐步去噪的生成过程，能够生成更高质量的样本，且训练过程更加稳定，避免了GAN中常见的模式崩溃问题
 - 多样性：扩散模型在生成样本的多样性上表现更好，尤其是在复杂数据分布下，能够覆盖更多潜在的模式
 - 理论保障：扩散模型的训练目标具有更加明确的数学支持，优化过程更加直接
- 劣势
 - 计算成本：扩散模型需要多部迭代去噪，导致生成速度较慢，且训练和推理的计算资源消耗较高
 - 实时性不足：相比GAN的单步生成，扩散模型难以满足实时性要求较高的应用场景

2)

1. 文本编码器：使用与训练的CLIP文本编码器，将输入文本转换为高维语义向量，作为生成过程的引导条件
2. 潜在空间扩散：直接在像素空间进行扩散计算成本过高，因此 Stable Diffusion 将图像压缩到低维潜空间，在此空间中进行扩散和去噪，大幅降低计算量
3. 交叉注意力机制：引入条件交叉注意力层，将文本嵌入与图像嵌在特征对齐，注意力权重动态调整生成方向，确保输出与文本语义一致

4. 条件调度：在扩散过程的每一步，文本条件被动态注入UNet的各个层级中，通过缩放因子增强条件控制强度，平衡生成结果的逼真度与文本匹配度