

---

# 作业 5：推理

---

清华大学软件学院  
人工智能导论, 2025 年春季学期

## 介绍

本次作业需要提交说明文档（PDF 形式）和 Python 的源代码。注意事项如下：

- 本次作业满分为 100 分。
- 除简答题、编程题外的题目，请给出必要的解答过程，只有答案且过于简略的回答会酌情扣除一定分数。题目要求直接回答或只汇报结果的题目，则不需要给出过程或分析。
- **请不要使用他人的作业，也不要向他人公开自己的作业，复制网上内容须在报告中说明，否则将受到严厉处罚，作业分数扣至-100（即倒扣本次作业的全部分值）。**
- 完成作业过程中，如果使用了大模型辅助（如润色文笔、询问知识点等），请在作业末尾声明使用的方式和程度（不影响作业评分）。**禁止直接粘贴大模型输出的文本**，否则会扣除一定的作业分数。
- 统一文件的命名：{学号}\_{姓名}\_hw5.zip。所有解答和实验报告请写在一个 pdf 文件中，和代码一起压缩上传。

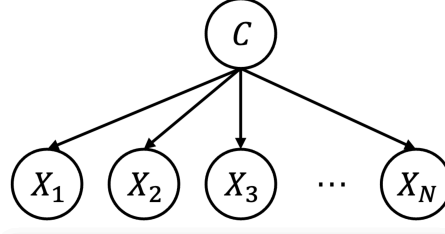
## 1 简答题（15 分）

1. 请简述最大似然估计（MLE）与最大后验估计（MAP）之间的主要区别，以及它们在数据量非常大时的关系。
2. 在 LDA 的生成模型中，为什么要分别为每个文档引入一个主题分布（Dirichlet 分布），而不是为整个语料库引入一个共享的主题分布？这样设计的统计意义与表达能力体现在哪里？
3. 变分自编码器（VAE）通过引入一个“推断网络”来近似潜在变量的后验分布。请解释为何在优化过程中不能直接最大化似然，而是要使用变分下界，并指出“重参数化技巧”解决了什么具体问题。

## 2 解答题

### 2.1 朴素贝叶斯网络 (20 分)

考虑如下朴素贝叶斯网络：



其中类别  $C$  是取值  $\{1, 2, \dots, K\}$  的离散随机变量，满足多项分布  $p(C = c) = \Theta_{C=c}$ ；特征  $\mathbf{X} = X_1 \dots X_N$  是  $N$  个取值  $\{0, 1\}$  的离散变量，满足伯努利分布

$$p(X_i = 1 \mid C = c) = \Phi_{X_i=1|c}.$$

- (a) **(最大似然估计)** 有  $M$  个独立同分布的样本  $D = \{(c_i, \mathbf{x}_i)\}_{i=1}^M$ ，请写出参数  $\Theta_C, \Phi_{X_i|C}$  的最大似然估计。

(提示：你可以假设样本足够多，答案中可能出现的计数均不为 0，下同)

- (b) **(最大后验估计)** Dirichlet 分布定义如下：

$$p(\Theta) = \frac{1}{Z} \prod_k \Theta_k^{\alpha_k - 1} \quad \text{where} \quad Z = \frac{\prod_{i=1}^k \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^k \alpha_i)}, \quad \Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$$

对  $\Theta_C, \Phi_{X_i|C=c}$  分别施加 Dirichlet 先验，分布参数分别为  $\alpha_C$  和  $\alpha_{x_i|c}$ ，请计算收集到题目定义的样本集  $D$  之后，参数  $\Theta_C, \Phi_{X_i|c}$  的最大后验估计。

(提示：答案中不应出现导数和积分项)

- (c) **(EM 算法)** 在现实应用中，类别变量  $C$  往往是隐藏的，即  $D = \{\mathbf{x}_i\}_{i=1}^M$ ，因此需要运用 EM 算法学习该模型的参数。请通过以下步骤证明：如果将参数初始化为均匀分布：

$$\Theta_{C=c}^{(0)} = \frac{1}{K}, \quad \Phi_{X_i=1|c}^{(0)} = \frac{1}{2},$$

那么 EM 算法将经过一次迭代之后就收敛。

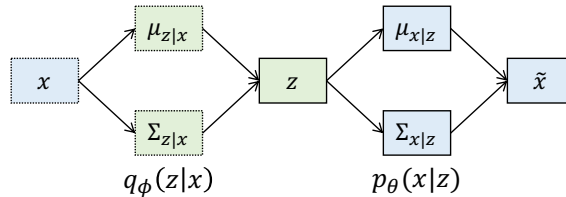
- (a) 请计算 E 步中得到的最优的证据下界 (ELBO)  $J(\Theta_C, \Phi_{X_i|c})$  的表达式

- (b) 请计算经过一次 EM 迭代之后的参数值  $\Theta_C^{(1)}, \Phi_{X_i|c}^{(1)}$

- (c) 请计算经过两次 EM 迭代之后的参数值  $\Theta_C^{(2)}, \Phi_{X_i|c}^{(2)}$

### 2.2 变分自编码器 (Variational Auto-Encoder, VAE, 15 分)

变分自编码器 (VAE) 是一种基于变分推断实现的深度生成模型，其模型框架如下图所示：

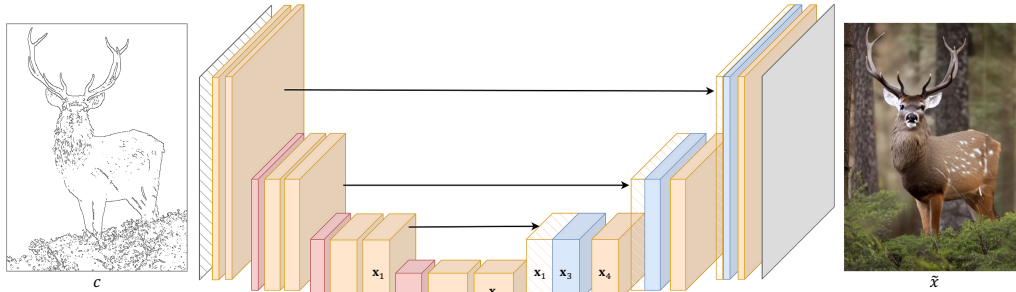


VAE 假设样本  $x$  由隐变量  $z$  通过解码器网络  $p_\theta(x|z)$  表示的分布决定，其中  $p(z) = \mathcal{N}(0, \mathbf{I})$ ，并设计编码器网络  $q_\phi(z|x)$  进行变分推断训练，优化如下的损失函数：

$$\mathcal{L}_{\text{ELBO}}(x; \phi, \theta) = \mathbb{E}_{z \sim q_\phi(z|x)} [-\log p_\theta(x|z)] + \text{KL}(q_\phi(z|x) \| p(z))$$

现将 VAE 扩展为条件生成模型，即 Conditional VAE (CVAE)，例如生成指定数字的手写数字图像。在 CVAE 中，样本  $x$  和隐变量  $z \sim \mathcal{N}(0, \mathbf{I})$  和给定的条件变量  $c$ （例如要生成的数字）共同决定，对应条件解码器网络  $p_\theta(x|z, c)$ ，并假设  $c$  与  $z$  独立。考虑模型的推理和训练过程：

- 画出  $x, z, c, x$  的概率图模型，并用  $p(z)$  和  $p_\theta(x|z, c)$  表示  $p(x|c)$ 。
- 为进行变分推断，需要设计编码器网络近似隐变量  $z$  的条件分布。注意到  $c$  与  $z$  独立，但在给定  $x$  的情况下，编码器建模的是  $q_\phi(z|x, c)$  而不是  $q_\phi(z|x)$ ，请从变量条件独立的角度解释原因。
- 根据 VAE 的损失函数，直接写出 CVAE 的损失函数  $\mathcal{L}_{\text{ELBO}}(x, c; \phi, \theta)$ 。提示： $c$  在概率分布中以条件的形式出现。
- 在  $p_\theta(x|z, c)$  和  $q_\phi(z|x, c)$  的建模中，若  $x \in \mathbb{R}^d$ ， $z \in \mathbb{R}^{d'}$  是一维向量， $c \in \{1, 2, \dots, k\}$  是类别标签，请写出合适的将  $c$  加到网络输入的方式。
- 若  $x$  是与图像对应的图像，则解码器通常使用 U-Net 作为网络架构，如下图所示（有关  $z$  的部分在此省略），其中每个立方体代表一个特征图（Feature Map），立方体的高、宽、厚分别代表特征图的高、宽和通道（Channel）数。图中所简略代表 U-Net 的特殊设计：它将浅层与深层的特征图在通道上进行拼接，并作为下一隐藏层的输入。例如，图中  $x_4$  的计算为  $\mathbf{x}_4 = F([\mathbf{x}_1, \mathbf{x}_3])$ ，其中中括号为对应隐藏层，方括号代表拼接操作。此种设计有什么好处？



### 3 图像生成: VAE 与 Diffusion (50 分)

**问题背景** 本编程作业旨在让同学们深入理解生成模型，特别是变分自编码器 (VAE) 和扩散模型 (Diffusion Models)，并应用于手写文字生成任务。在这次任务中，我们将依次完成多个步骤，包括架构设计、模型训练、推理生成、以及模型评估。

**任务目标** 本编程作业的目标是让学生掌握生成模型的核心概念与应用。学生需要依次实现条件变分自编码器 (CVAE) 架构，并通过定义训练目标进行模型训练；接着，通过 FID (Fréchet Inception Distance) 等指标评估生成图像的质量。作业还包括扩散模型 (Diffusion Model) 的理解与实现，特别是条件扩散模型 (Conditional Diffusion Model) 的前向加噪与标签注入机制。最后，学生将通过运行测评代码，对生成效果进行分析与评价，从而深入理解生成模型在手写文字生成中的应用和性能优化。

- 
1. **CVAE 的架构**: 相较于最简单的变分自编码器 (VAE), 条件变分自编码器 (CVAE) 最大的差异就在于其引入的条件变量。在我们的例子中, 这一条件即数据集中的对于手写的图片标注的从 0 到 9 的数字标签。请阅读 `unified_model/models/vae_model.py`, 实现条件变分自编码器 (CVAE) 架构, 该架构应包含:
    - 编码器 (Encoder): 输入图片和标签, 返回隐变量  $z$  的均质和方差;
    - 解码器 (Decoder): 输入隐变量  $z$  和标签, 返回重建的图像;请结合代码中的提示 (Hint), 补全 `vae_model.py` 中 CVAE 类的 `__init__` 函数。
  2. **CVAE 的训练目标**: 请参考课件中对 VAE 的讲解, 结合代码中的提示 (Hint), 补全 `unified_model/models/vae_model.py` 中的 `train_step` 函数。
  3. **CVAE 的重参数化**: 在课堂中, 我们说明了重参数化 (Reparameterization trick) 对于梯度反向传播的必要性, 请完成 `unified_model/models/vae_model.py` 中 `reparameterize` 函数的书写。
  4. **CVAE 的训练与推理**: 请阅读 `unified_model/train.py`, `unified_model/inference.py` 与 `cal_fid.py`, 参考 `README.md` 的脚本, 依次完成 CVAE 模型的训练、推理与指标计算。为达到更好的训练效果, 对于训练的超参数, 你可以根据需求调整。在你的报告中, 请包括 `out/samples/vae_samples.png` 与你每一类的 FID Score。`vae_samples.png` 的图像应当类似下面这个案例:



图 1: 例: `vae_samples.png`

- (注: FID 是衡量图像生成质量的常见指标, 越低代表生成的图像与真实图像越接近, 在本任务中, 所有类别的 FID Score 通常可以低于 30.0)
5. **走进扩散模型**: 扩散模型 (Diffusion Models) 是现代生成模型中的重要组成部分。在此, 我们提供了一份完整的基于 UNet 架构的扩散模型的训练、推理、可视化流程的代码。请你阅读 `unified_model/models/diffusion_model.py` 与 `samplers/diffusion_samplers.py`, 补全 `diffusion_samplers.py` 中的 `sample_forward` 函数。
  6. **扩散模型的训练与推理**: 请参考 `README.md` 的脚本, 依次完成扩散模型 (Diffusion Model) 的训练、推理与可视化。在你的报告中, 请包括 `out/samples/diffusion_ddpm_samples.png`

与 `out/visualization/diffusion_ddpm_visualization/denoising_grid.png`。图像示例如下：

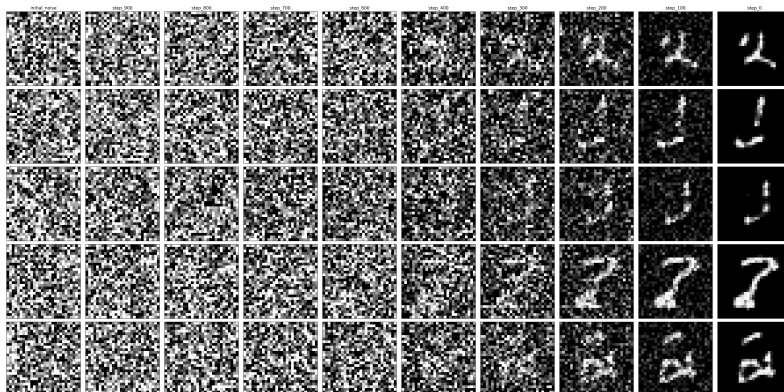


图 2: 例: denoising\_grid.png

7. **为扩散模型加入条件：**那么聪明的小朋友就会说了：“嗯？这个扩散模型并没有接收数字条件作为输入，那它在生成的东西我也不知道是什么数字呀”，这非常有道理。因此，我们参考条件变分自编码器 (CVAE)，对应的获得了条件扩散模型 (Conditional Diffusion Models)。在这里，一个核心的问题在于将数字条件加入模型的架构。请阅读 `unified_model/models/unet.py`，补全 `ConditionalUNet` 类 `forward` 函数中条件注入的部分。
8. **条件扩散模型的训练与推理：**请参考 `README.md` 的脚本，依次完成条件扩散模型 (Conditional Diffusion Model) 的训练、推理、可视化与评测。请在你的报告中包括推理得到的网格、去噪过程可视化的图像、以及评测得到的每一类的 FID Score。

**参数选择** 完成上述题目时，可以自由选择适合你的情况的参数进行训练，但过于不合理的参数设置可能会导致扣分（若报告未说明实验使用的参数，则会以提交的代码为准）。

## 4 提交格式

- 请先删除 `*.so *.pyd` 等文件和例如 `data/` 等数据与模型，再将你的代码目录内**所有代码文件**、三个模型评测使用的 **checkpoint**、**评测生成的图像目录** (以 `by_class` 结尾的文件夹) 以及你的**文字报告**打包提交。统一文件的命名：`{学号}_{姓名}_hw5.zip`。
- 请将本次作业所有问题回答写在同一份报告中，报告请导出为 **pdf 格式**。