

这份PDF是关于**生成模型 (Generative Models)** 的课程讲义，由西安电子科技大学人工智能学院制作。以下是逐页分析并整理的机器学习复习笔记，涵盖了核心概念、数学公式、模型对比及优缺点。

第一部分：生成模型基础 (P1 - P8)

1. 核心目标 (Motivation)

- **理查德·费曼名言**: “What I cannot create, I do not understand.” (我不能创造的，我也不理解)。
- **生成模型的目标**: 学习数据的真实底层分布 $p_{data}(x)$ ，从而能够生成新的、逼真的样本。
- **关键应用**: 图像合成、语音合成 (如WaveNet)、超分辨率、图像修补、DeepFakes、内容创作。

2. 生成模型 vs. 判别模型 (Generative vs. Discriminative)

- **判别模型 (Discriminative Models)**:
 - 学习**条件概率** $P(Y|X)$ 。
 - 关注点: 寻找类别之间的决策边界 (如分类猫和狗)。
- **生成模型 (Generative Models)**:
 - 学习**联合概率** $P(X, Y)$ 或**数据分布** $P(X)$ 。
 - 关注点: 理解数据是如何生成的。
 - **用途**: 采样 (采样/内容合成)、密度估计 (异常检测)、缺失数据填补 (逆问题)、半监督/自监督学习。

3. 生成模型的分类体系 (Taxonomy)

生成模型主要分为两大类:

1. **显式密度模型 (Explicit Density)**: 通过数学公式直接定义或逼近概率密度函数。
 - **可处理密度 (Tractable)**: 如自回归模型 (AR)、概率PCA (PPCA)。
 - **近似密度 (Approximate)**:
 - 变分法: 如变分自编码器 (VAE)。
 - 马尔可夫链: 如扩散模型 (Diffusion)。
2. **隐式密度模型 (Implicit Density)**: 不直接学习概率密度，而是学习一个生成过程。
 - **直接采样**: 如生成对抗网络 (GAN)。

第二部分：显式密度 - 可处理模型 (P9 - P16)

4. 概率主成分分析 (Probabilistic PCA, PPCA)

- **定义**: 将PCA看作一个受约束的高斯分布模型，属于线性-高斯框架。
- **生成视角**: $x = Wz + \mu + \epsilon$
 - z 是隐变量, $z \sim \mathcal{N}(z|0, I)$ 。
 - 条件分布: $p(x|z) = \mathcal{N}(x|Wz + \mu, \sigma^2 I)$ 。
- **边缘分布**: $p(x) = \int p(x|z)p(z)dz = \mathcal{N}(x|\mu, C)$, 其中协方差矩阵 $C = WW^T + \sigma^2 I$ 。
- **后验分布**: $p(z|x) = \mathcal{N}(z|M^{-1}W^T(x - \mu), \sigma^{-2}M)$, 其中 $M = W^TW + \sigma^2 I$ 。

5. 自回归模型 (Autoregressive Models, AR)

- **核心原理**: 利用概率链式法则将联合分布分解。
 - $p(x) = \prod_{i=1}^n p(x_i|x_1, \dots, x_{i-1})$
- **代表架构**:
 - **RNNs**: 处理序列数据 (文本、音频)。
 - **PixelRNN / PixelCNN**: 逐像素生成图像。
- **优缺点**:
 - **优点**: 似然度计算是精确且可处理的 (Tractable), 训练容易。
 - **缺点**: 生成过程是序列化的, 计算速度慢。

第三部分：显式密度 - 近似模型 (P17 - P20)

6. 自动编码器 (Autoencoder) 与 VAE

- **普通自编码器 (AE)**: 通过最小化均方误差 $E(w) = \frac{1}{2} \sum ||y(x_n, w) - x_n||^2$ 将输入映射到自身。即使有非线性层, 某种意义上也等同于线性PCA。
- **变分自编码器 (VAE)**:
 - **编码器** $q_\phi(z|x)$: 将输入映射到隐空间 z (捕捉 pose, light 等变化因子)。
 - **解码器** $p_\theta(x|z)$: 从隐变量重建输入。
 - **目标函数 (ELBO)**: $\log p(x) \geq \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)] - D_{KL}(q_\phi(z|x)||p(z))$
 - 第一项: 重建损失 (Reconstruction Loss)。

- 第二项：正则化项 (Regularization Term)。
 - 优缺点：原理扎实，隐空间具有可解释性；但生成的图像由于L2损失和高斯假设通常较模糊。
 - VQ-VAE：引入离散码本 (Discrete Codebook) 解决VAE输出模糊的问题。
-

第四部分：隐式生成模型与前沿技术 (P21 - P25)

7. 生成对抗网络 (GANs)

- 核心原理：博弈论中的极小极大博弈 (Minimax Game)。
- 两个网络相互竞争：
 - 生成器 (G)：生成假数据以欺骗判别器。
 - 判别器 (D)：区分真实数据 x 和假数据 $G(z)$ 。
- 目标函数：
$$\min_G \max_D V(G, D) = \mathbb{E}_{x \sim p_{data}} [\log D(x)] + \mathbb{E}_{z \sim p_z} [\log(1 - D(G(z)))]$$
- 优缺点：
 - 优点：生成的图像非常清晰、高质量。
 - 缺点：训练不稳定（难以达到纳什均衡）；存在模式崩塌 (Mode Collapse)；没有显式的似然密度。

8. 扩散模型 (Diffusion Models) - SOTA

- 地位：现代生成式AI的主流范式（如Stable Diffusion, DALL-E）。
 - 过程：
 - 前向过程 (固定)：逐渐向数据添加高斯噪声，直到变成纯噪声 x_T 。
 - 反向过程 (学习)：学习一个神经网络来逐级去噪重建图像。
 - 优点：样本质量极高，训练稳定性优于GAN。
-

第五部分：总结对比 (P26 - P27)

模型	类型	优点 (Pros)	缺点 (Cons)
自回归模型 (AR)	显式 (可处理)	精确的似然估计	采样速度慢 (逐个生成)

模型	类型	优点 (Pros)	缺点 (Cons)
变分自编码器 (VAE)	显式 (近似)	采样快, 隐空间平滑	样本模糊
生成对抗网络 (GAN)	隐式	样本非常锐利清晰	训练不稳定, 模式崩塌
扩散模型 (Diffusion)	显式 (近似)	质量高, 训练稳定	采样速度慢 (需多次迭代)

复习建议:

1. **重点掌握** VAE 的 ELBO 公式分解及其每一项的物理意义。
2. **理解** GAN 的对抗损失函数, 注意 min 和 max 分别对应的对象。
3. **对比** 为什么 AR 模型和 Diffusion 模型生成速度慢 (都是迭代/序列生成) , 而 VAE 生成快 (单次前向传播) 。
4. **注意** PPCA 是生成模型中唯一一个有解析闭式解 (在特定条件下) 且属于线性范畴的模型。