09 EM算法

隐变量学习、有效性与VAE联系

学习目标

- 说明EM处理的隐变量参数估计问题，区分观测、隐状态与模型参数
- 解释E/M两步的分工，以及精确后验和下界改善为何支持似然不下降
- 完成GMM一轮更新，核对责任度、参数与目标变化，说明实践边界
- 理解VAE中的近似推断、生成模型与ELBO，说明重构、表示和生成的用途
- 重建ELBO、KL与Q的关系，并完成精确EM的单调性证明

1 EM的用途与隐变量问题

从可见观测估计含隐状态的模型参数。

EM用于含隐变量的参数估计

EM（Expectation-Maximization，期望最大化）是一种参数估计算法：模型包含未观测的**隐状态**，训练要从可见数据估计模型参数。

- **高斯混合模型**（Gaussian Mixture Model，GMM）用多个高斯分量解释观测数据
  - 本例只有四个观测值 $x = (0,2,8,10)$，设定两个分量
  - 每个样本来自哪个分量没有记录；两个分量的权重、均值和方差也需要估计
- EM反复利用当前模型推断隐状态，再利用这些推断更新参数
  - 算法输出估计的模型参数；后验软分配可用于解释样本归属
  - GMM规定数据模型，EM提供学习这一模型的一条求解路线

GMM的观测、隐标签与生成模型

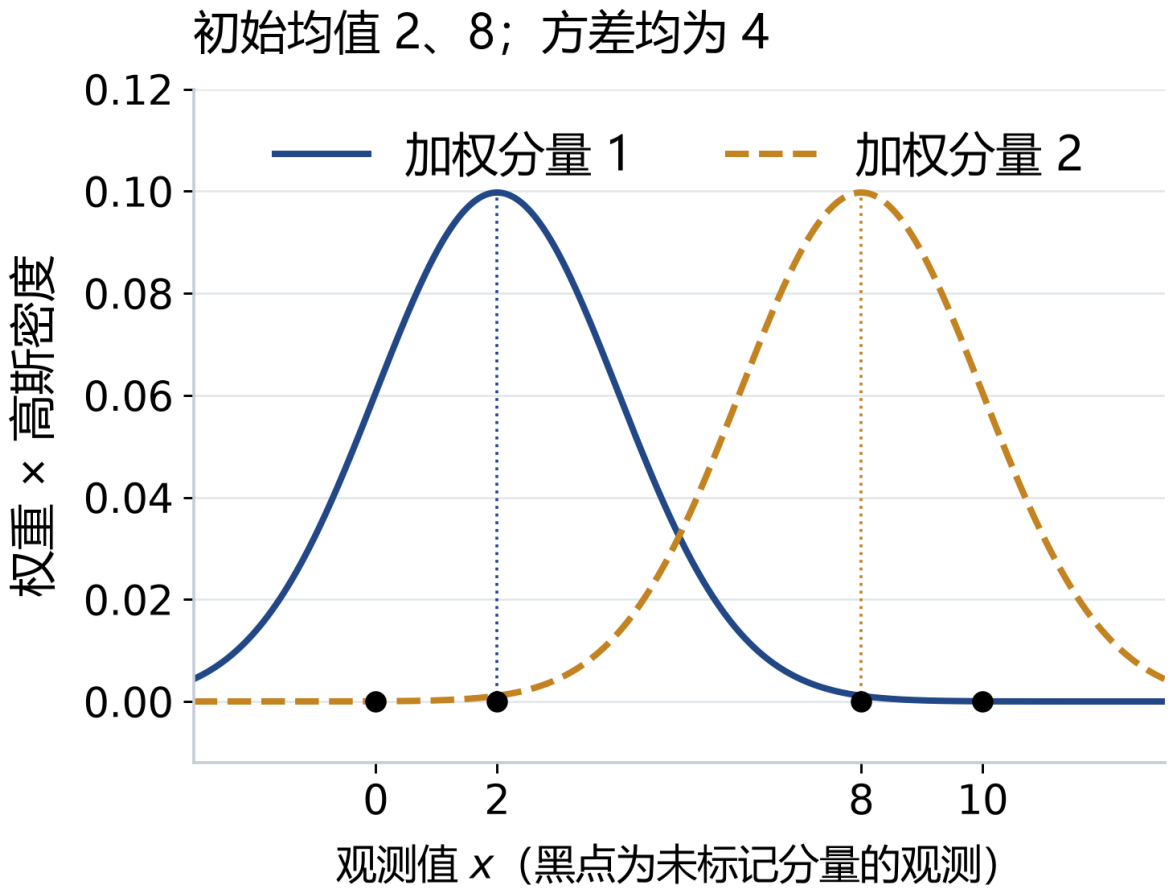
对象	本例含义
观测 x_i	已知的第i个数值
隐标签 $z_i \in \{1,2\}$	未知的分量身份，每个样本各有一个
参数 $\theta = (\pi_k, \mu_k, \sigma_k^2)_{k=1}^2$	各样本共享的权重、均值与方差

- 先按权重选择分量，再从该高斯分量产生观测
  - $\pi_k > 0$、$\sum_k \pi_k = 1$、$\sigma_k^2 > 0$；$\mathcal{N}(x \mid \mu, \sigma^2)$表示高斯密度
  - $p_\theta(x, z = k) = \pi_k \mathcal{N}(x \mid \mu_k, \sigma_k^2)$
- 分量身份未观测，对所有可能身份求和得到观测密度
  - $p_\theta(x) = \sum_k \pi_k \mathcal{N}(x \mid \mu_k, \sigma_k^2)$
  - 联合项与观测项都是关于连续x的密度，密度值不是单点事件概率

未知归属与参数估计的相互依赖

初始模型仍使用同一观测 $x = (0,2,8,10)$：$\pi = (0.5,0.5)$、$\mu = (2,8)$、$\sigma^2 = (4,4)$。

- **完整数据**同时包括观测与隐标签；标签若已知，可按分量计数并分别估计均值与方差
  - 完整数据目标为 $\sum_i \log p_\theta(x_i, z_i)$
- 标签未知，一个观测可由两条分量曲线共同解释，观测目标对可能分量求和
  - $\ell(\theta) = \sum_i \log \left[\sum_k \pi_k \mathcal{N}(x_i \mid \mu_k, \sigma_k^2)\right]$
- **参数决定后验软权重，参数更新又需要这些权重**
  - EM先用当前模型计算贡献，再固定贡献更新参数；直接梯度优化也可能可用



2 EM的机制与有效性

当前后验形成加权目标，下界解释有效性。

E步：以当前后验表示软归属

- 初始化给出旧参数 θ^t，E步固定它，为每个样本计算隐状态的后验分布
  - $q_i^{t+1}(z_i) = p_{\theta^t}(z_i \mid x_i)$
  - q_i是隐状态上的归一化分布，表示当前模型给各解释的软权重
  - t表示当前轮次；q_i^{t+1}是这一轮E步的输出，仍由旧参数θ^t计算
- GMM中的后验概率称为**责任度**
  - $r_{ik} = \frac{\pi_k^t \mathcal{N}(x_i|\mu_k^t,(\sigma_k^2)^t)}{\sum_j \pi_j^t \mathcal{N}(x_i|\mu_j^t,(\sigma_j^2)^t)}$
  - 初始GMM的观测2：两项加权密度约0.099736、0.001108，归一化得责任度约0.989、0.011
  - 这些值就是该观测的q；每行和为1，隐标签仍未被观察到
- E表示期望：用这些概率计算完整数据对数似然的期望，交给M步

M步：固定后验更新模型参数

- 后验权重把未知隐状态的贡献汇总为Q**函数**
  - $Q_{\text{data}}(\theta, \theta^t) = \sum_i E_{q_i^{t+1}} [\log p_\theta(x_i, z_i)]$
  - 对观测2，期望是约0.989倍的第一分量对数联合项加约0.011倍的第二分量对数联合项
  - Q再汇总全部四个观测；M步中固定数据与未舍入旧权重，调整候选参数
- M步在同一允许参数集合内最大化这个总目标，以下假定最大值可取得
  - $\theta^{t+1} = argmax_\theta Q_{\text{data}}(\theta, \theta^t)$
  - θ是候选新参数，θ^t只确定旧后验。GMM由此得到加权计数、均值与方差。
- 用新参数重新计算后验，进入下一轮E步
  - 小写q表示分布，大写Q表示目标函数；M步中途不改用新后验

ELBO与后验差值

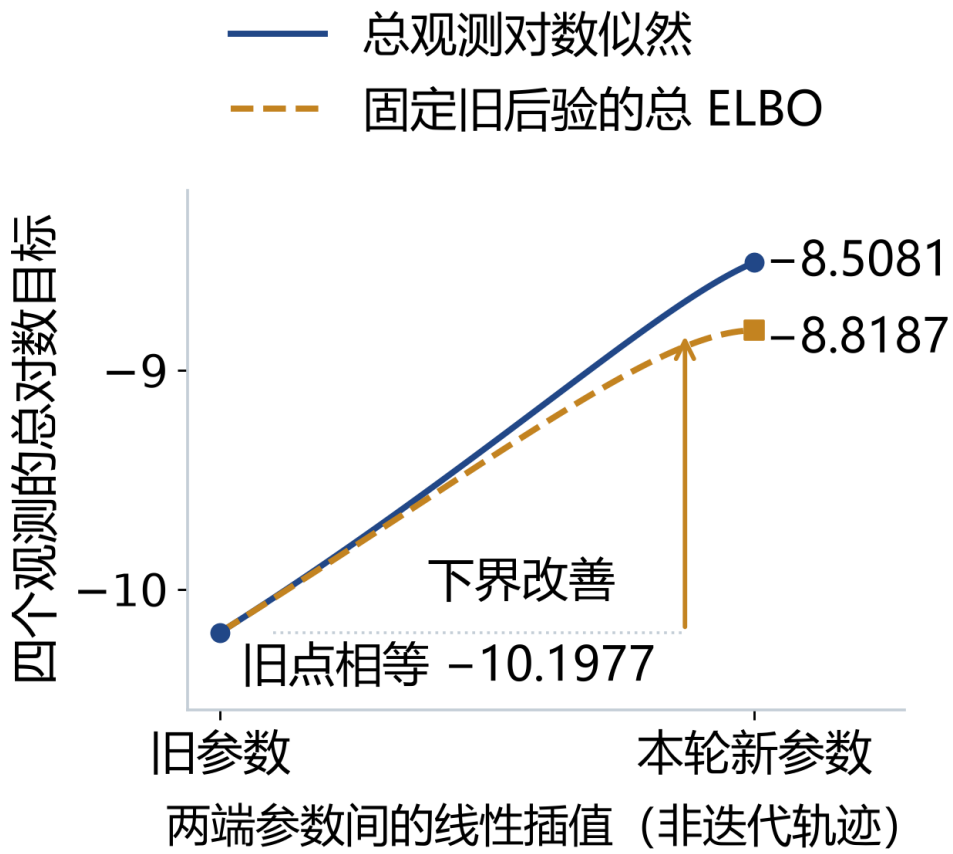
证据下界（Evidence Lower Bound，ELBO）是观测对数似然的下界，将M步优化的Q与观测目标联系起来。

- 对一个观测x，固定其候选隐状态分布q，ELBO定义为：
  - $\mathcal{L}(q,\theta) = E_q[\log p_\theta(x,z) - \log q(z)]$
  - 对数凹性给出“对数的平均≤平均的对数”；Jensen完整代入在附录
  - “证据”指观测密度 $p_\theta(x)$；ELBO使用联合模型与q计算，不要求先算真实后验的归一化项
  - $H(q) = -\sum_z q(z)\log q(z)$是**离散熵**；ELBO等于期望对数联合项加熵，固定q时熵不变
- 下界与观测对数似然之差是q与真实后验的**KL散度**
  - $\log p_\theta(x) = \mathcal{L}(q,\theta) + D_{KL}(q \| p_\theta(z \mid x))$
  - $D_{KL}(q\|p) = E_q \log(q/p)$，非负且不对称；它在这里衡量分布差异
  - 本例观测2取旧模型精确后验作q：KL为零，ELBO等于该观测的对数密度
  - 对四个观测求和，得到下一页的总目标对照

精确EM为何不降低似然

同一四点GMM，固定旧参数的精确后验q。

- **旧点相等**：精确E步使后验差值为零
  - 总ELBO与总观测对数似然都约为−10.1977
- **固定后验，改善下界**：M步提高总Q，也就提高总ELBO
  - 总ELBO与总Q只差固定后验的熵之和，因此改善幅度相同
- **新参数处**：新似然仍不低于同一固定后验的下界
  - 本轮总ELBO约−8.8187，总观测对数似然约−8.5081；均不低于旧值



保证要求精确后验、M步不降低同一总Q、旧新参数属于同一允许集合且相关目标有限。不保证严格增加、全局最优或参数序列收敛；完整证明保留在附录。

3 GMM示例与实践边界

固定数据完成一轮更新，核查结果及求解边界。

固定输入与后验责任度

固定观测 $x = (0,2,8,10)$，$n = 4$；旧权重为 $(0.5, 0.5)$，旧均值为 $(2, 8)$，旧方差为 $(4, 4)$。

一维高斯密度用均值和方差计算：

$$\mathcal{N}(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]$$

对 $x = 2$，两项“权重×高斯密度”约为0.099736与0.001108；除以两项之和，得到后验概率约为0.989与0.011。

观测值	分量1责任度	分量2责任度
0	0.999	0.001
2	0.989	0.011
8	0.011	0.989
10	0.001	0.999

每行和为1。表格显示三位小数，参数更新使用未舍入值；M步固定这组旧后验。

责任度如何形成加权学习目标

- 完整数据中，一个样本只贡献给其所属分量；标签未知时，按各分量后验概率加权
  - 联合项为 $p_\theta(x_i, z_i = k) = \pi_k \mathcal{N}(x_i \mid \mu_k, \sigma_k^2)$
  - 取对数，再按固定责任度求期望，得到：
  - $Q(\theta, \theta^t) = \sum_i \sum_k r_{ik} \left[\log \pi_k + \log \mathcal{N}(x_i \mid \mu_k, \sigma_k^2)\right]$
- 同一责任度汇总成分量的**软计数**
  - $N_k = \sum_i r_{ik}$是该分量的后验期望样本数，不要求是整数
  - 本例由对称性得到 $N_1 = N_2 = 2$
- Q中的对数权重项给出混合权重更新，高斯对数项给出加权均值和方差更新
  - 三类参数使用同一组责任度；本例每个软计数均大于零

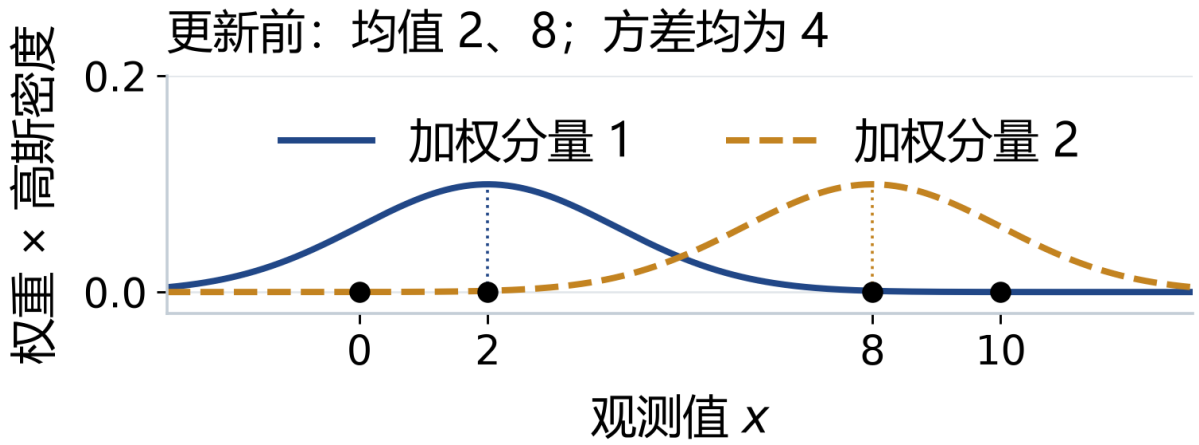
混合权重与加权均值更新

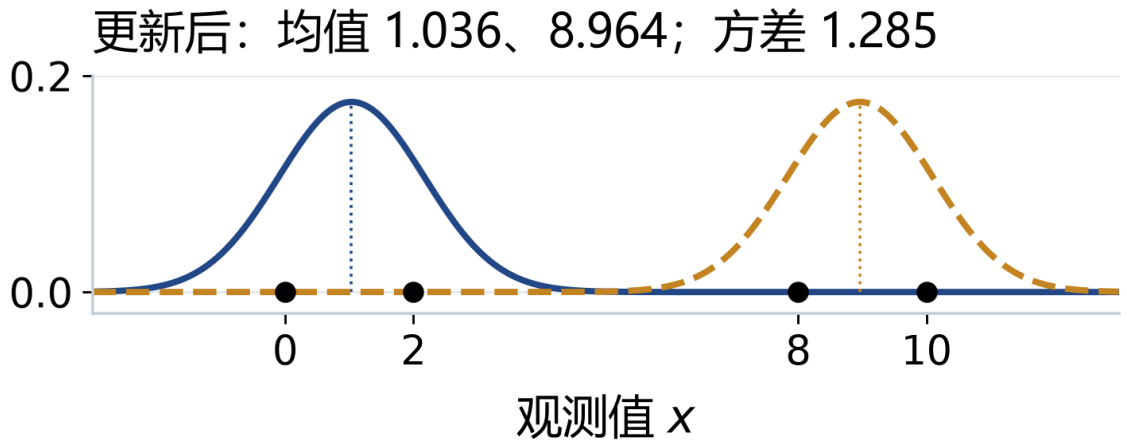
- 分量权重由软计数占总样本数的比例估计
  - $\pi_k^{new} = N_k/n$
  - 本例两分量均为 $2/4 = 0.5$，满足非负且和为1
- 均值由责任度加权的观测值平均得到
  - $\mu_k^{new} = \frac{\sum_i r_{ik}x_i}{N_k}$
  - 分量1的加权和约为2.0714494422，除以2得 $\mu_1^{new} \approx 1.0357$；分量2得 $\mu_2^{new} \approx 8.9643$
  - 左侧样本对分量1的权重大，右侧样本对分量2的权重大，均值分别向1与9移动

使用新均值更新方差

- 方差是观测值到新均值的平方偏差的责任度加权平均
  - $\left(\sigma_k^2\right)^{new} = \frac{\sum_i r_{ik}(x_i - \mu_k^{new})^2}{N_k}$
  - 使用本轮新均值与本轮E步旧责任度，方差更新前不重算后验
- 本例两个分量的新方差都约为1.2845
  - 先算到新均值的平方偏差，再乘责任度、求和、除以软计数2
  - 分量1主要受0、2影响；8、10虽权重小，平方偏差较大，仍有贡献
  - 分量2由对称性得到相同方差

一轮结果、使用与下一轮输入





状态	混合权重	均值	方差	总观测对数似然
更新前	(0.5,0.5)	(2,8)	(4,4)	−10.1977
更新后	(0.5,0.5)	约(1.0357,8.9643)	约(1.2845,1.2845)	−8.5081

均值向责任度较大的观测移动；方差减小，曲线变窄、峰值升高。

未舍入参数的自然对数目标增加约1.6896；本轮数值核查不代替一般证明。

新模型可评价密度、计算后验；下一轮重算责任度，本轮M步固定旧责任度。

初始化、停止与退化风险

- 初始化影响最终结果
  - 多个起点可能得到不同驻点或局部解；在同一模型约束下排除退化与异常，再比较目标
- 停止规则检查计算是否稳定
  - 相邻轮非负似然增量小于给定阈值时可停止；超过舍入误差的下降应检查步骤
  - 最大迭代次数是计算限制；很小的密度可用对数域计算减少下溢
- 无约束GMM可能**方差塌缩**，似然没有有限上界
  - 一个正权重分量对准单个样本并令方差趋零；其他固定分量为其余样本保持正密度，总似然可无界增长
  - 方差下限、惩罚与保护性更新改变问题或更新规则，保证须对应实际目标

目标单调且有有限上界时，目标值收敛；这不自动保证参数收敛或达到全局最优。

4 变分思想与VAE用途

后验难处理时采用近似推断，连接VAE学习与使用。

精确后验难处理时的近似推断

本例GMM只需归一化两个密度，精确后验能直接计算。复杂隐变量模型可能需要难以处理的积分，经典精确E步因而难以执行。

- 在可处理的分布族 $\mathcal{Q}$ 中选择近似后验q，继续优化同一ELBO
  - 固定模型参数，在允许族中改善q；相关期望应能计算或估计，并支持优化
  - 固定参数时观测证据不变，**提高ELBO等价于减少q与真实后验的KL差值**，因此这一优化在寻找更接近后验的软解释
  - 连续隐变量按概率密度对z积分，离散隐状态的求和相应换为积分；ELBO仍写成关于q的期望
  - 固定q，更新模型参数；这给出**变分EM**的交替分工
- 近似q通常与真实后验有正KL差值
  - 即使两步都提高ELBO，也不能直接沿用精确EM的旧参数处相等关系
  - 两步都不降低实际ELBO时保证该目标不下降，观测似然未必逐步提高

VAE的生成模型与近似后验

VAE（Variational Autoencoder，变分自编码器）用神经网络参数化生成模型与近似后验，并通过ELBO学习；潜变量z通常连续。

对象	本讲采用的含义
$p(z)$	固定先验，例如标准高斯，规定生成前的潜变量分布
$p_\theta(x \mid z)$	**概率解码器**，给定z产生观测分布
$q_\phi(z \mid x)$	**概率编码器**，由观测x给出近似后验分布

- 联合模型为 $p_\theta(x, z) = p(z)p_\theta(x \mid z)$
  - 生成方向从z到x，推断方向从x到z；编码器用于近似难计算的真实后验
  - θ控制解码器，ϕ控制编码器，均为跨样本共享参数；z是每样本隐状态
- 同一编码器处理不同输入，避免每个输入都重新建立独立的迭代推断过程
  - 神经网络增强表达能力，但精确后验通常仍难计算

VAE的ELBO如何支持学习

将 $p_\theta(x,z) = p(z)p_\theta(x \mid z)$ 代入已有ELBO，并合并 $\log p(z) - \log q_\phi(z \mid x)$，得到：

$$\mathcal{L}(\theta,\phi;x) = E_{q_\phi(z|x)}[\log p_\theta(x \mid z)] - D_{KL}(q_\phi(z \mid x) \| p(z))$$

- 重构期望项鼓励由编码所得的z解释输入x
  - 解码器给输入较高的条件概率或密度，提高这一项；具体重构损失取决于观测分布
- 与先验的KL项使近似后验与先验匹配
  - 它连接编码所得潜变量与生成时的先验；不要求每个输入的后验完全等于先验
- 该目标无需显式计算真实后验，可用小批量与采样估计来训练θ和ϕ
  - 给定输入时，重参数化将潜变量采样写为编码器参数与独立噪声的可微变换，使编码器参数可通过采样期望的梯度更新

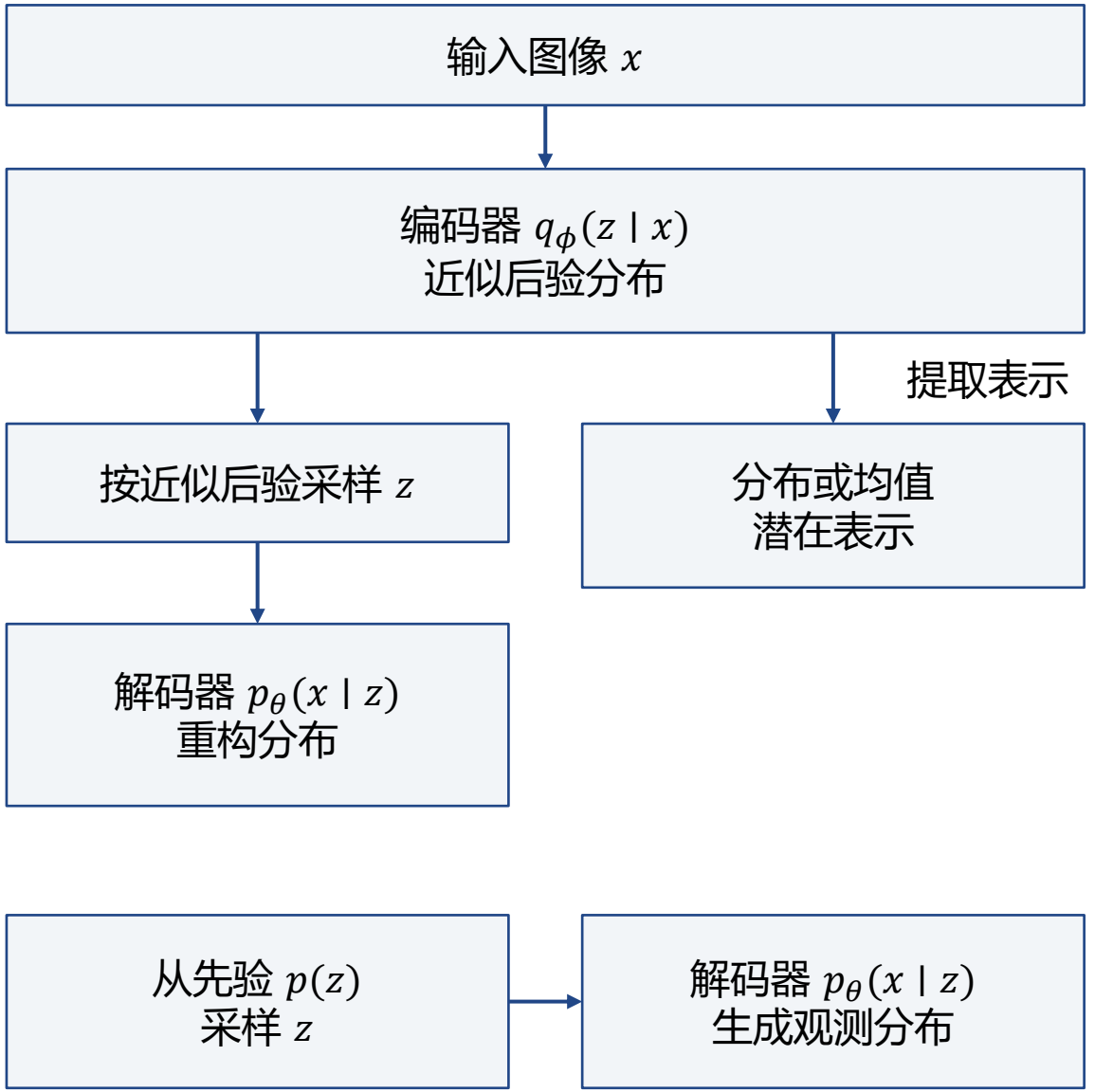
这里的KL比较近似后验与先验；ELBO与证据的差值比较近似后验与真实后验，两者对象不同。

VAE的重构、表示与生成用途

以图像x为例，已训练VAE提供两条不同的数据路径。

- 重构：输入经近似推断后解码，用于观察模型如何解释已有数据
- 表示：编码器的分布或均值可作为潜在表示，是否适合具体任务还需评价
- 生成：从先验采样并解码，得到新的观测样本；无需先输入一张待重构图像

潜变量不必对应可解释的人类语义，重构和生成质量也需要单独检查。



EM、变分EM与VAE的联系和区别

比较对象	隐状态处理	参数学习与保证
精确EM	当前模型的精确后验	固定后验改善总Q；满足条件时观测似然不下降
变分EM	允许分布族中的近似后验	固定对象交替改善ELBO；保证对应实际目标与更新条件
典型VAE训练	共享编码器给出近似后验	采样/小批量估计ELBO，联合梯度更新θ与ϕ

- 三者都联系隐状态推断、生成模型参数学习与下界目标
  - VAE把近似推断参数化，并使用适合神经网络的随机梯度训练
- 精确EM的单调性证明依赖精确后验与M步不降
  - 这些条件不自动适用于VAE；随机估计提高不等于实际ELBO或似然每步提高

VAE沿用隐变量与变分下界思想，具体训练过程需要按其自身算法判断。

本讲总结

- EM把含隐状态的参数学习组织成后验推断与参数更新
  - 精确E步计算旧后验，M步固定它改善Q；GMM责任度据此形成加权统计量
- 精确后验与下界改善解释似然不下降
  - 附录用Jensen、KL与Q关系补全证明；条件性保证不包含全局最优或参数必然收敛
- 完整GMM一轮更新把软分配转为权重、均值和方差
  - 结果核查、初始化、停止与退化边界共同约束算法使用
- VAE用近似后验与生成模型学习同一类下界目标
  - 编码支持重构和表示，先验采样支持生成；典型联合随机梯度训练不继承精确EM的逐步似然保证

5 附录：完整推导与单调性证明

完整推导与单调性证明，供课后必修研读。

5.1 后验加权梯度与 ELBO 构造

由梯度识别后验权重，再以Jensen构造下界并说明等号条件。

观测对数似然的梯度展开

- **归属与参数的相互依赖**
  - 观测对数似然可求导；其中各隐标签的后验概率由当前参数决定，并随参数变化
- **求导条件**
  - 隐状态有限，$p_\theta(x_i) > 0$，联合项 $p_\theta(x_i, z_i)$ 关于 θ 可微且为正
- **从对数求和到加权梯度**
  - $$\nabla_\theta \log \sum_k p_\theta(x_i, k) = \frac{\sum_k \nabla_\theta p_\theta(x_i, k)}{p_\theta(x_i)}$$
  - $$= \sum_k \frac{p_\theta(x_i, k)}{p_\theta(x_i)} \nabla_\theta \log p_\theta(x_i, k)$$
  - 第一行对“隐标签求和后的密度”取对数再求导；第二行用 $\nabla p = p \nabla \log p$，将联合项的导数写为对数联合项的梯度。$p_\theta(x_i, k)$ 是 $p_\theta(x_i, z_i = k)$ 的简写

后验加权梯度与交替更新

- **识别当前后验权重**
  - 系数 $r_{ik}(\theta) = p_\theta(z_i = k \mid x_i) = p_\theta(x_i, k)/p_\theta(x_i)$ 是隐标签的后验概率；观测对数似然的梯度因此是各 $\log p_\theta(x_i, k)$ 梯度的后验加权和
- **交替结构的动机**
  - 参数决定责任度，参数更新又利用责任度形成加权统计量。直接梯度优化仍可行；EM 则交替计算后验并固定后验更新参数

Jensen 不等式的加权形式

- **加权平均的条件**
  - 设 $a_z > 0$，$q(z) \geq 0$，且 $\sum_z q(z) = 1$
- **对数凹性与 Jensen 形式**
  - $\log(\sum_z q\,(z) a_z) \geq \sum_z q(z) \log a_z$
  - 左侧先按 q 加权平均再取对数，右侧先取对数再按 q 加权平均；对数凹性保证左侧不小于右侧
  - 对数曲线上两点之间的连线不高于曲线；因此在同一加权横坐标处，曲线高度不低于两点高度的加权平均

ELBO 的构造与后验分布

候选分布条件

- 沿用有限隐状态与正联合项的条件：$q(z) \geq 0$，$\sum_z q(z) = 1$，且 $p_\theta(x,z) > 0$ 的状态须有 $q(z) > 0$

隐变量模型中的代入

- 以下 x 表示单个观测；独立样本的数据集目标对各样本结果求和。将 $a_z = p_\theta(x,z)/q(z)$ 代入 Jensen，左侧为观测对数似然，右侧为 ELBO：

$$\log p_\theta(x) = \log \sum_z q(z) \frac{p_\theta(x,z)}{q(z)}$$

$$\geq \sum_z q(z) \log \frac{p_\theta(x,z)}{q(z)} = \mathcal{L}(q,\theta)$$

- 比值在 $q(z) > 0$ 的状态上计算；支撑条件确保乘除 q 不遗漏正的联合项。本例 GMM 的正权重、正方差使两个标签的联合项均为正

等号与后验选择

- 对数严格凹，等号成立要求所有具有正 q 概率的状态满足 $p_\theta(x,z)/q(z) = c$。由归一化得 $p_\theta(x) = c \sum_z q(z) = c$，因此 $q(z) = p_\theta(x,z)/p_\theta(x) = p_\theta(z \mid x)$

5.2 ELBO 展开与 KL 散度分解

展开ELBO的期望与熵，并由后验定义推导KL差值。

ELBO 的期望项与熵项

- **下界定义的展开**
  - 将对数比值写为 $\log p_\theta(x,z) - \log q(z)$，再分别按 q 求期望
- **完整数据对数似然的期望**
  - 对每种隐状态的 $\log p_\theta(x,z)$ 按 q 加权平均，得到期望项；观测 x 在这一计算中保持固定
- **分布 q 的熵**
  - $H(q) = -E_q[\log q(z)]$ 衡量离散隐状态分布的不确定性，约定 $0\log 0 = 0$；在相同有限状态集合上，质量均匀时最大，集中在一个状态时为零
  - 改变 q 会同时改变两项，更新 q 时须优化它们的和；固定 q 更新参数时，熵项保持不变

$$\begin{aligned}\mathcal{L}(q,\theta) &= \sum_z q(z)\log p_\theta(x,z) - \sum_z q(z)\log q(z) \\ &= E_q[\log p_\theta(x,z)] + H(q)\end{aligned}$$

这里按 q 对隐状态求期望；数据集的目标再对各样本的 ELBO 求和。

观测对数似然与 ELBO 的差值

- **分解条件**：隐状态有限，$p_\theta(x) > 0$，q 归一化，且 KL 有限；后验概率为零的状态，q 也赋予零概率，求和中零质量项取零
- **代入后验定义**：用 $p_\theta(z \mid x) = p_\theta(x, z)/p_\theta(x)$ 展开 KL
  - $D_{KL}(q \| p_\theta(z \mid x)) = \sum_z q(z) \log \frac{q(z) p_\theta(x)}{p_\theta(x,z)}$
- **提出对数证据并识别 ELBO**
  - $D_{KL} = \log p_\theta(x) - \sum_z q(z) \log \frac{p_\theta(x,z)}{q(z)}$
  - $\log p_\theta(x) = \mathcal{L}(q, \theta) + D_{KL}(q \| p_\theta(z \mid x))$

$\log p_\theta(x)$ 与 z 无关，提出求和后乘以 $\sum_z q(z) = 1$，因此只保留一份。分解本身是恒等式：KL 非负使 ELBO 成为下界，KL 为零时 ELBO 等于观测对数似然。

5.3 E/M 更新与观测对数似然单调性

固定旧后验，连接加权目标与下界，证明总似然不下降。

精确 E 步与数据集 Q 函数

精确 E 步与 Q 函数

- 对独立样本，用旧参数 θ^t 分别计算精确后验，再对各样本的完整数据对数似然取后验期望、求和：

$$q_i^{t+1}(z_i) = p_{\theta^t}(z_i \mid x_i)$$

- $$Q_{\text{data}}(\theta, \theta^t) = \sum_i E_{q_i^{t+1}} [\log p_\theta(x_i, z_i)]$$

- $E_{q_i^{t+1}}$ 按样本 i 的旧参数后验对隐标签求期望；观测 x_i 保持固定。GMM 中这些概率就是责任度，隐标签仍未观测

M 步中 ELBO 与 Q 的等价优化

- **固定后验的 M 步**
  - 记 $q^{t+1} = \prod_i q_i^{t+1}$，总下界为 $\mathcal{L}_{\text{data}}(q^{t+1}, \theta) = \sum_i \mathcal{L}_i\,(q_i^{t+1}, \theta)$
  - 在同一允许参数集合上，若最大值可取得，M 步可写为：
  - $$\theta^{t+1} = \text{argmax}_\theta \mathcal{L}_{\text{data}}(q^{t+1}, \theta)$$ $$= \text{argmax}_\theta Q_{\text{data}}(\theta, \theta^t)$$
  - 由 $\mathcal{L}_{\text{data}} = Q_{\text{data}} + \sum_i H(\,q_i^{t+1})$ 可知，固定后验后，两目标只相差一个与共享参数 θ 无关的常数，因此最大化它们得到相同的最优参数集合；q 表示分布，Q 表示参数目标函数
- **旧参数处的等式与 GMM 目标**
  - 精确 E 步使旧参数处 KL=0、观测对数似然等于 ELBO；GMM 的 Q 展开为责任度加权项

观测对数似然不下降的证明链

从旧参数到新参数的证明链

- $\ell(\theta) = \sum_i \log p_\theta(x_i)$ 是数据集观测对数似然；q^{t+1} 由各样本在旧参数下的精确后验组成，证明中保持固定，总下界是各样本 ELBO 之和
- 旧、新参数属于同一允许参数集合，各样本在两点的观测密度均为正，所用对数似然、ELBO 与 KL 均有限

$$\ell(\theta^t) = \mathcal{L}_{\text{data}}(q^{t+1}, \theta^t)$$

- $$\leq \mathcal{L}_{\text{data}}(q^{t+1}, \theta^{t+1})$$
$$\leq \ell(\theta^{t+1})$$

单调性证明的依据与结论

三段关系的依据

- 旧参数处的等式：精确 E 步使各样本的 KL 为零，因此总 ELBO 等于旧参数处的观测对数似然
- 中间不等式：M 步固定同一个 q^{t+1}，旧参数仍是可选点，因此最大化总 ELBO 不会降低其值；证明只需总 ELBO 不下降，不要求逐样本提高或每次都达到最大值
- 新参数处的不等式：同一固定后验下，各样本 ELBO 都不超过新参数处的观测对数似然，求和后仍成立

结论与范围

- 因此，一次 E/M 更新后的数据集观测对数似然不低于更新前的值
- 不推出严格增加、参数必然收敛或全局最优