

基于高斯混合模型案例的EM算法教学设计

杜 芳, 方晓峰

火箭军工程大学基础部数学教研室, 陕西 西安

收稿日期: 2024年12月6日; 录用日期: 2025年1月14日; 发布日期: 2025年1月22日

摘 要

EM (Expectation Maximization) 算法是统计学中的核心算法, 也是本校近代数理统计课程教学过程中的一个重难点。论文采用案例式、启发式、研讨式教学方法, 以基于高斯混合模型(GMM)的轴承退化阶段划分问题为例, 引导学生发现隐变量模型极大似然估计(MLE)存在的困难, 设计问题链启发学生探寻参数估计的数值方法, 并总结出EM算法的一般过程。基于matlab编程可视化EM算法下的GMM模型参数更新过程, 对比MLE目标函数和EM迭代目标函数, 分析EM算法的内涵思想并结合图形进行直观展示, 并且挖掘其中蕴含的思政元素, 在知识传授的同时实现价值塑造。

关键词

EM算法, 高斯混合模型, 教学设计

Teaching Design of EM Algorithm Based on Gaussian Mixture Model Case

Fang Du, Xiaofeng Fang

Teaching Section of Mathematics, Basic Department of Rocket Force University of Engineering, Xi'an Shaanxi

Received: Dec. 6th, 2024; accepted: Jan. 14th, 2025; published: Jan. 22nd, 2025

Abstract

Expectation maximization (EM) algorithm is a core algorithm in statistics and also a key and difficult point in the teaching process of modern mathematical statistics courses in our school. The paper adopts a case-based and heuristic teaching method, taking the Gaussian Mixture Model (GMM) based bearing degradation stage division problem as an example, guiding students to discover the difficulties of maximum likelihood estimation (MLE) in the latent variable model, designing a problem chain to inspire students to explore numerical methods for parameter estimation, and summarizing the general process of EM algorithm. Based on Matlab programming, the parameter update

process of GMM based on EM algorithm is visualized. Comparing the MLE objective function and EM iteration objective function, the intrinsic thought of EM algorithm is analyzed and visually displayed with graphics. The ideological and political elements are also explored, so as to achieve value shaping while knowledge transmission.

Keywords

EM Algorithm, Gaussian Mixture Model, Teaching Design

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着国防和军队改革的深入推进, 全军兴起加强军事人才建设热潮, 也对军校研究生教育提出了新的要求: 针对服务实战中人才培养的需要, 从学术型人才培养向应用型人才培养转变, 保证国防和军队改革后的战斗力[1]。而当前军校研究生培养面临着课程内容理论性较强、科学研究工作与人才培养的结合不紧、学生自身积极主动性不够等问题。如何从科学研究和服务部队两个方面出发, 建设以学生能力素质提升为主线, 以创新能力培养为核心的课程, 是军校研究生教育的重点任务[2]。

近代数理统计是面向我校研究生开设的一门选修课程, 课程聚焦于统计学中的概念、理论和方法, 为学生依据实际问题进行建模或通过统计推断发现问题的内在规律提供科学方法, 支撑研究生的专业方向课程学习, 启迪研究生分析问题和解决问题的思维[3]。在学情方面, 选修该门课程的学生已经选定研究方向, 具备了一定的文献查阅能力、问题分析能力和科研探索能力, 但是数理统计基础参差不齐, 容易对数理统计中的抽象原理望而生畏, 同时非常关心如何应用数理统计方法解决实际问题, 希望从课程中学习到的能应用于专业领域的知识和方法[4]。

在工程实践中, 存在大量对观测不完整数据或者含有隐变量的模型做参数估计的问题。直接采用极大似然估计方法通常无法获得参数估计的解析表达式, EM 算法是针对此类问题的一种有效的数值求解方法[5]。EM 算法是统计学中的核心算法, 是高斯混合模型、隐马尔可夫模型、深度生成模型等众多机器学习方法的理论基础, 也是近代数理统计课程教学中的一个重难点。

基于这样的国防需求、课程目标和学情特点, 我们采用案例式、启发式、研讨式教学方法进行 EM 算法的教学设计, 在提高课程的专业应用性的同时, 注重学生多维能力的培养。本次课以基于 GMM 的轴承退化阶段划分问题为引例, 引导学生发现在对不完整观测数据或隐变量模型做极大似然估计时面临的困境以及设计数值解法的必要性。拆解问题并组织学生进行课前思考和课上讨论, 使其逐步探索总结出 EM 算法的迭代更新过程, 进而应用算法解决引例问题, 再结合 matlab 编程使学生对迭代过程有更加直观的认识。最后, 对比 MLE 目标函数和 EM 迭代目标函数, 深入分析 EM 算法的思想内涵, 证明 EM 算法的有效性和收敛性。具体实施过程在 2~5 节展开介绍。

2. 基于 GMM 的轴承退化阶段划分问题

2.1. 问题阐述

轴承寿命预测对于开展机械设备健康管理, 提高旋转机械运行安全性、可靠性具有重要意义, 其中轴承退化阶段的划分能够有效提高其寿命预测的精度[6]。现实应用中通常缺乏大量故障标记数据, 经常

采用无监督的聚类算法进行轴承退化阶段的划分, 高斯混合模型(GMM)就是一种有效的聚类算法:

$$p(x, \theta) = \sum_{k=1}^K P(y^k = 1) p_{X|Y}(x | y^k = 1) \quad (1)$$

其中随机变量 Y 表示轴承所属退化阶段, 其取值为 K 维 one-hot 编码向量 $(y^1, y^2, \dots, y^K)$, $y^k = 1$ 代表轴承属于第 k 个退化阶段, 每个退化阶段的轴承服从多元高斯分布, 即

$$p_{X|Y}(x | y^k = 1) = (2\pi)^{-\frac{n}{2}} |\Sigma_k|^{-\frac{1}{2}} e^{-\frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k)}, k = 1, 2, \dots, K \quad (2)$$

为简化后续表达, 将(1)式记为

$$p(x, \theta) = \sum_{k=1}^K \alpha_k p_k(x, \theta_k) \quad (3)$$

其中 $\theta = (\alpha_1, \theta_1, \alpha_2, \theta_2, \dots, \alpha_K, \theta_K)$ 为总体参数, 且 $\sum_{k=1}^K \alpha_k = 1$, 于是 $Y \sim PN(1: \alpha_1, \alpha_2, \dots, \alpha_K)$ 。

轴承退化阶段划分的目的是通过观测数据 $\tilde{x} = (x_1, x_2, \dots, x_N)$, 推断出每个成分的参数 θ_k 以及每个轴承 $x_i (i = 1, 2, \dots, N)$ 所属的退化阶段 y_i 。事实上, 只要(3)式中的所有参数已知, 即可得到 y_i 的条件分布为:

$$P_{Y|X}(y | x, \theta) = \frac{P(y, \theta) P_{X|Y}(x | y, \theta)}{\sum_y P(y, \theta) P_{X|Y}(x | y, \theta)} \quad (4)$$

所以解决轴承退化阶段划分的核心问题还是要通过观测 $\tilde{x} = (x_1, x_2, \dots, x_N)$ 估计总体参数 $\theta = (\alpha_1, \theta_1, \alpha_2, \theta_2, \dots, \alpha_K, \theta_K)$ 。

2.2. GMM 参数估计的困难

课前向学生发布预习问题 1~3, 课堂中组织学生交流汇报。

问题 1. 已知观测数据 $\tilde{x} = (x_1, x_2, \dots, x_N)$ 时, 请尝试采用 MLE 方法估计总体参数。

不难发现对数似然函数具有如下形式:

$$\sum_{i=1}^N \log p(x_i, \theta) = \sum_{i=1}^N \log \sum_{k=1}^K \alpha_k p_k(x_i, \theta_k) \quad (5)$$

其中包含求和项的对数, 在关于参数求偏导时, 导数形式复杂, 很难或者无法求得参数估计的解析表达式。而当问题无法解析求解时, 我们通常会尝试迭代的数值算法, EM 算法就是一种针对不完整观测数据或隐变量模型参数估计的数值迭代算法。

3. EM 算法迭代过程探索

数值迭代算法, 通常预先给定一个初始值, 通过分析目标函数在当前取值下的性质, 构造出下一步的参数迭代公式。这里我们采用类似的方式, 考虑下面两个问题。

3.1. 模型参数给定时隐变量(轴承退化阶段)后验分布的推断

问题 2. 当已知参数为 $\theta^t = (\alpha_1^t, \theta_1^t, \alpha_2^t, \theta_2^t, \dots, \alpha_K^t, \theta_K^t)$ 时, 分析观测为 $\tilde{x} = (x_1, x_2, \dots, x_N)$ 时, 退化阶段变量 Y 的后验分布是什么?

对学生的讨论结果进行总结, 对于问题 2, 有:

$$P(y_i^k = 1 | x_i, \theta^t) = \frac{\alpha_k^t p_k(x_i, \theta_k^t)}{\sum_{k=1}^K \alpha_k^t p_k(x_i, \theta_k^t)} \quad (6)$$

将 $P(y_i^k = 1 | x_i, \theta^t)$ 记为 y_i^{kt} , 则 $Y_i \sim PN(1: y_i^{1t}, y_i^{2t}, \dots, y_i^{Kt})$, 即隐变量的后验分布和先验分布一样, 属于多项分布。

3.2. 隐变量(轴承退化阶段)已知时(3)式的 MLE

问题 3. 当隐变量也可观测时, 根据观测数据 $(\tilde{x}, \tilde{y}) = (x_1, \dots, x_N, y_1, \dots, y_N)$ 能否对(3)式中的参数做 MLE?

总结学生讨论结果, 轴承特征随机变量 X 和退化阶段随机变量 Y 的联合分布如下:

$$p(x, y, \theta) = P(y, \theta) p_{X|Y}(x | y, \theta) = \alpha_k^{y_k} \prod_{k=1}^K [p_k(x, \theta_k)]^{y_k} = \prod_{k=1}^K [\alpha_k p_k(x, \theta_k)]^{y_k} \quad (7)$$

已知观测数据 $(\tilde{x}, \tilde{y}) = (x_1, \dots, x_N, y_1, \dots, y_N)$ 时, 似然函数为:

$$L(\theta) = p(\tilde{x}, \tilde{y}, \theta) = \prod_{i=1}^N p(x_i, y_i, \theta) = \prod_{i=1}^N \prod_{k=1}^K [\alpha_k p_k(x_i, \theta_k)]^{y_i^k} \quad (8)$$

对似然函数求对数, 得对数似然函数为:

$$\log L(\theta) = \sum_{i=1}^N \sum_{k=1}^K y_i^k \log \alpha_k + y_i^k \log p_k(x_i, \theta_k) \quad (9)$$

对数似然函数关于参数求偏导, 可得:

$$\frac{\partial \log L(\theta)}{\partial \theta_k} = \sum_{i=1}^N y_i^k \frac{\partial \log p_k(x_i, \theta_k)}{\partial \theta_k} \quad (10)$$

其中 y_i^k 取值为 0 或 1, 相当于只利用样本中属于第 k 个退化阶段的轴承特征, 来估计 GMM 中第 k 个成分参数 θ_k 。

对于成分占比参数 α_k , 极大化(9)式相当于求解如下约束优化问题:

$$\max_{\alpha_1, \dots, \alpha_K} \sum_{k=1}^K \sum_{i=1}^N y_i^k \log \alpha_k \quad \text{s.t.} \quad \sum_{k=1}^K \alpha_k = 1 \quad (11)$$

利用拉格朗日乘子法, 可得 $\alpha_k = \frac{1}{N} \sum_{i=1}^N y_i^k$, 即样本中属于第 k 个退化阶段的个体所占比例。

3.3. EM 算法的迭代过程

问题 2 和问题 3 是比较容易解决的, 也就是说在总体参数给定的情况下, 可以根据观测轴承推断其所属退化阶段的后验分布, 在轴承退化阶段也可观测的情况下, 能够通过 MLE 得到总体参数的显式估计。如果交替进行这两步, 就可以得到参数的一种迭代更新过程, 但是问题 2 中给出的是分布, 而问题 3 中用到的是观测值, 如何衔接这两个过程呢?

联系随机变量的分布和随机变量的观测值, 有两种简单途径: 第一种是从分布中随机抽样得到观测值, 第二种是按照分布对所有可能得观测值求期望。两种途径各有优缺点, 抽样法能保证可行的数值计算, 但是需要设计抽样方法缓解以偏概全现象, 期望法比较全面, 但是当分布形式复杂时, 难以获得显式表示。当采用第二种途径时, 对应的迭代过程就是 EM 算法:

E 步: 当参数 $\theta^t = (\alpha_1^t, \theta_1^t, \alpha_2^t, \theta_2^t, \dots, \alpha_K^t, \theta_K^t)$ 时, 计算完整观测下的似然函数 $p(x, y, \theta)$ 或者对数似然函数 $\log p(x, y, \theta)$ 关于条件分布 $P(y | x, \theta^t)$ 的期望:

$$E_Y(p(x, Y, \theta) | x, \theta^t) \quad \text{或} \quad E_Y(\log p(x, Y, \theta) | x, \theta^t) \quad (12)$$

上述期望可以看作是参数 θ 的函数, 记为 $Q(\theta, \theta^t)$ 。

M 步: 将 $Q(\theta, \theta^t)$ 极大化, 得到参数迭代公式

$$\theta^{t+1} = \arg \max_{\theta} Q(\theta, \theta^t) \quad (13)$$

4. 基于 EM 算法的引例问题求解

4.1. GMM 参数迭代公式推导

回到 GMM 中, (8)式中的似然函数关于条件分布 $P(\tilde{y} | \tilde{x}, \theta^t)$ 的期望为:

$$E_Y(L(\theta) | \tilde{x}, \theta^t) = \prod_{i=1}^N \prod_{k=1}^K [\alpha_k p_k(x_i, \theta_k)]^{E_Y(y_i^k | x_i, \theta^t)} = \prod_{i=1}^N \prod_{k=1}^K [\alpha_k p_k(x_i, \theta_k)]^{y_i^{kt}} \quad (14)$$

(9)式中的对数似然函数关于条件分布 $P(\tilde{y} | \tilde{x}, \theta^t)$ 的期望为:

$$E_Y(\log L(\theta) | \tilde{x}, \theta^t) = \sum_{i=1}^N \sum_{k=1}^K y_i^{kt} \log \alpha_k + y_i^{kt} \log p_k(x_i, \theta_k) \quad (15)$$

记 $E_Y(\log L(\theta) | \tilde{x}, \theta^t)$ 为 $Q(\theta, \theta^t)$, 则其关于参数的偏导为:

$$\frac{\partial Q(\theta, \theta^t)}{\partial \theta_k} = \sum_{i=1}^N y_i^{kt} \frac{\partial \log p_k(x_i, \theta_k)}{\partial \theta_k} \quad (16)$$

将 $p_k(x_i, \theta_k)$ 替换为(2)式中的多元高斯分布可得:

$$\frac{\partial Q(\theta, \theta^t)}{\partial \mu_k} = \sum_{i=1}^N y_i^{kt} \Sigma_k^{-1} (\mu_k - x_i) \quad (17)$$

$$\frac{\partial Q(\theta, \theta^t)}{\partial \Sigma_k} = \sum_{i=1}^N y_i^{kt} \left[-\Sigma_k^{-1} + \Sigma_k^{-1} (x_i - \mu_k)(x_i - \mu_k)^T \Sigma_k^{-1} \right] \quad (18)$$

令偏导为 0, 即可获得如下参数更新公式:

$$\mu_k^{t+1} \leftarrow \frac{\sum_{i=1}^N y_i^{kt} x_i}{\sum_{i=1}^N y_i^{kt}} \quad \Sigma_k^{t+1} \leftarrow \frac{\sum_{i=1}^N y_i^{kt} (x_i - \mu_k^{t+1})(x_i - \mu_k^{t+1})^T}{\sum_{i=1}^N y_i^{kt}} \quad (19)$$

参数 α_k 更新方式与 3.1 中类似, $\alpha_k^{t+1} \leftarrow \frac{1}{N} \sum_{i=1}^N y_i^{kt}$ 。

这些迭代公式和单高斯分布参数极大似然估计的公式非常类似, 差别是 GMM 迭代公式中为每个观测值赋予了权重 y_i^{kt} 。

4.2. 基于 matlab 的 GMM 参数迭代过程展示

通过 matlab 编程实现上述迭代过程, 图 1 展示了包含两个成分的一维高斯混合模型的参数和类别划分迭代过程。其中黑色曲线代表真实的 GMM 模型概率密度曲线, 红色和蓝色曲线分别代表学习到的两个一维高斯分布成分, 底部的红色“*”和蓝色“。”代表两种数据划分结果。可以看到随着迭代次数的增加, 学习到的两个成分期望逐渐逼近真实数据分布中的期望, 数据划分结果也逐渐接近真实数据类型。

5. EM 算法的理论分析

前面几部分启发式地探寻了 EM 算法的迭代更新步骤, 并且通过一维高斯混合模型示例直观展示了

EM 算法的效果。下面我们从理论角度分析证明 EM 算法的思想内涵和有效性。

5.1. EM 算法的思想

首先, 比较 EM 迭代优化目标 $Q(\theta, \theta^t) = E_Y(\log p(x, Y, \theta) | x, \theta^t)$ 和极大似然估计的优化目标 $\log p(x, \theta)$ 。

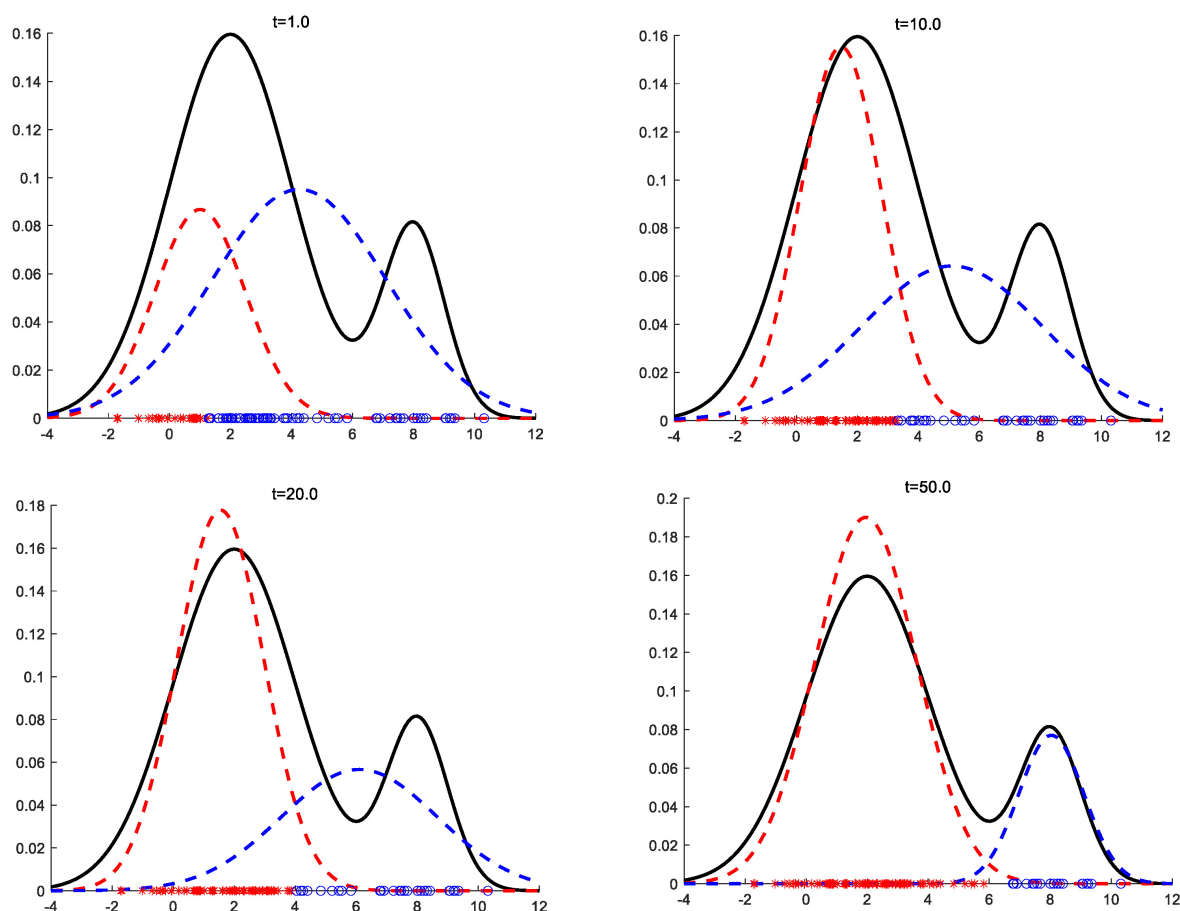


Figure 1. Iterative update process of one-dimensional GMM

图 1. 一维混合高斯模型的迭代更新过程

以连续型隐变量为例, 也就是比较 $\log p(x, \theta)$ 和 $\int \log p(x, y, \theta) p(y | x, \theta^t) dy$, 我们选择从 $\log p(x, \theta)$ 出发逐步向 $\int \log p(x, y, \theta) p(y | x, \theta^t) dy$ 靠拢:

$$\log p(x, \theta) = \log \int p(x, y, \theta) dy = \log \int \frac{p(x, y, \theta)}{p(y | x, \theta^t)} p(y | x, \theta^t) dy \quad (20)$$

其中 $\log x$ 为凹函数, 根据 Jensen 不等式, 可得:

$$\begin{aligned} \log p(x, \theta) &\geq \int \log \frac{p(x, y, \theta)}{p(y | x, \theta^t)} p(y | x, \theta^t) dy \\ &= \int \log p(x, y, \theta) p(y | x, \theta^t) dy - \int \log p(y | x, \theta^t) p(y | x, \theta^t) dy \\ &= Q(\theta, \theta^t) + H[p(y | x, \theta^t)] \end{aligned} \quad (21)$$

其中 $H[p(y|x, \theta^t)]$ 和优化参数 θ 无关, (21) 式说明 EM 迭代优化目标 $Q(\theta, \theta^t)$ 是 MLE 优化目标 $\log p(x, \theta)$ 的下界, 也就是说 EM 算法中每个 E 步, 都是在当前参数 θ^t 下求 $\log p(x, \theta)$ 的下界函数, 再用下界函数的极大值点来更新参数, 过程如图 2 所示。

虽然无法直接求得 $\log p(x, \theta)$ 的极大值点, 但是按照这样的学习过程, 在迭代过程中参数会朝着极大值点不断靠近, 最终能够达到极大值点的任意小邻域内。这样的学习过程也蕴含着人生哲理: 不积跬步无以至千里, 不积小流无以成江海, 在学习和工作中, 每天积累一点, 我们会朝着目标不断靠近。

5.2. EM 算法的收敛性

从图 2 中直观来看, EM 算法之所以有效, 是因为随着参数的迭代更新, $\log p(x, \theta)$ 在不断提高, 这一结果是可证明的。

定理: EM 算法在每一次迭代后, 都有 $\log p(x, \theta^{t+1}) \geq \log p(x, \theta^t)$ 。

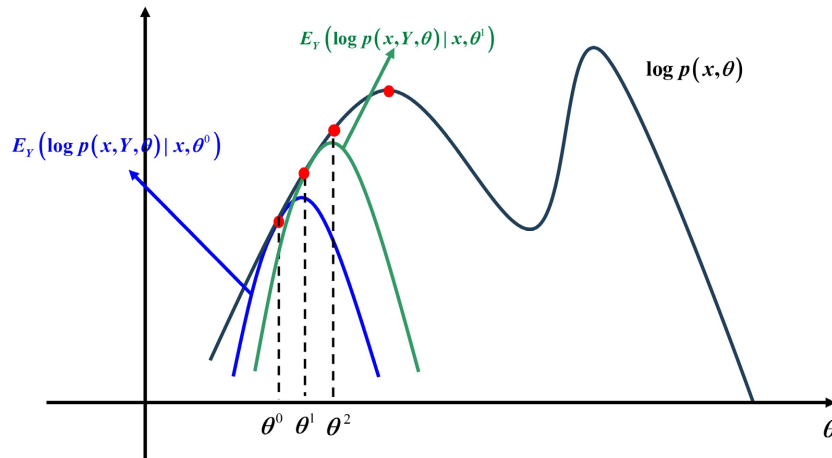


Figure 2. Parameter learning process of EM algorithm
图 2. EM 算法参数学习过程

证明: 根据条件概率公式, 有:

$$p(x, \theta) = \frac{p(x, y, \theta)}{p(y|x, \theta)} \quad (22)$$

两边取对数, 则有:

$$\log p(x, \theta) = \log p(x, y, \theta) - \log p(y|x, \theta) \quad (23)$$

两边同时关于条件分布 $P(y|x, \theta^t)$ 求期望, 可得:

$$\begin{aligned} \int \log p(x, \theta) p(y|x, \theta^t) dy &= \log p(x, \theta) \\ &= \int \log p(x, y, \theta) p(y|x, \theta^t) dy - \int \log p(y|x, \theta) p(y|x, \theta^t) dy \\ &= Q(\theta, \theta^t) - \int \log p(y|x, \theta) p(y|x, \theta^t) dy \end{aligned} \quad (24)$$

将 θ^t 和 θ^{t+1} 代入 $\log p(x, \theta)$ 再相减, 可得:

$$\begin{aligned} \log p(x, \theta^{t+1}) - \log p(x, \theta^t) &= [Q(\theta^{t+1}, \theta^t) - Q(\theta^t, \theta^t)] \\ &\quad - [\int \log p(y|x, \theta^{t+1}) p(y|x, \theta^t) dy - \int \log p(y|x, \theta^t) p(y|x, \theta^t) dy] \end{aligned} \quad (25)$$

因为 $\theta^{t+1} = \arg \max_{\theta} Q(\theta, \theta^t)$, 所以等式右边第一项 $Q(\theta^{t+1}, \theta^t) - Q(\theta^t, \theta^t) \geq 0$, 将等式右边第二项记为 K , 对其做等价变换, 有:

$$K = \int \log \frac{p(y|x, \theta^{t+1})}{p(y|x, \theta^t)} p(y|x, \theta^t) dy \quad (26)$$

根据 Jensen 不等式, 有:

$$K \leq \log \int \frac{p(y|x, \theta^{t+1})}{p(y|x, \theta^t)} p(y|x, \theta^t) dy = \log \int p(y|x, \theta^{t+1}) dy = \log 1 = 0 \quad (27)$$

综上所述, $\log p(x, \theta^{t+1}) - \log p(x, \theta^t) \geq 0 \Leftrightarrow \log p(x, \theta^{t+1}) \geq \log p(x, \theta^t)$ 。

再根据单调有界原理, 若 $\log p(x, \theta)$ 有上界, 则 EM 算法收敛到 $\log p(x, \theta)$ 的稳定点。从图 2 也可以看出, EM 算法不能保证收敛到全局最优点, 而且收敛位置受到初值影响。较为可行的办法是选取几个不同的初值进行迭代, 然后在多个估计间加以选择, 减轻初值选取对结果的影响[7]。

6. 教学效果分析

学生通过查阅资料和课前讨论, 针对教师课前给出的 3 个问题做了充分准备, 并且在授课过程中主动发言, 课堂参与的主动性和积极性被充分调动。此外学生的科研兴趣也得到了激发, 在课后进一步追踪 EM 算法的研究进展及其在前沿机器学习方法中的应用。批改课后大作业发现, 作业的结构完整度、统计方法 and 专业应用的契合度都有明显提升, 学生基本都能通过编程实现算法。此外, 通过微信群向选修课程的 12 名学生发出调查问卷, 请他们对本节课的教学效果进行评价, 统计结果分析见表 1。

Table 1. Feedback on class effectiveness of EM algorithm

表 1. EM 算法课堂效果反馈

调查项目	备选项及对应占比
对教学模式是否满意?	A. 非常满意(占比 91.7%)
	B. 满意(占比 8.3%)
	C. 不满意(占比 0%)
对教学内容是否满意?	A. 非常满意(占比 100%)
	B. 满意(占比 0%)
	C. 不满意(占比 0%)
对能力提升是否有帮助?	A. 非常有用(占比 83.3%)
	B. 比较有用(占比 16.7%)
	C. 不太有用(占比 0%)
对专业研究是否有帮助?	A. 非常有帮助(占比 83.3%)
	B. 比较有帮助(占比 16.7%)
	C. 帮助不大(占比 0%)

7. 结语

本节课以与学生专业相关的轴承退化阶段划分问题为引例, 结合课前问题, 引导学生发现隐变量模型极大似然估计中的困难, 从而探索迭代数值算法, 总结 EM 算法的一般步骤, 并应用其解决引例问题,

培养了学生分析问题、解决问题的能力,锻炼了学生的探索思维,提高了学生的创新意识。最后通过理论推导,分析 EM 算法的内涵和收敛性,培养学生严谨的科研态度,并且挖掘算法中蕴含的思政元素,在知识传授、能力培养的同时实现价值塑造。学生课堂上主动讨论,课后积极探索并认真完成大作业,对于课堂效果满意度较高,为课程的进一步建设提供了方向。

致 谢

感谢相关文献对本文的启发以及审稿专家提出的宝贵意见。

参考文献

- [1] 赵慎,王红云,郭希维,张自宾.军校研究生教育应用型培养模式探讨[J].中国教育技术装备,2018(22): 92-93+98.
- [2] 安涛,张腾,何宇廷.聚焦备战打仗型军校研究生培养模式的探索与实践[J].大学教育,2024(14): 111-114.
- [3] 肖枝洪,黄守成.人工智能时代下研究生应用数理统计优质课程建设[J].大学数学,2024,40(2): 41-46.
- [4] 宫春梅.基于创新能力培养的研究生《数理统计》课程教学研究[J].创新创业理论研究与实践,2020(16): 25-26.
- [5] Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **39**, 1-22.
<https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>
- [6] 陈东楠,胡昌华,郑建飞,郑红倩,裴洪.基于一种新型健康指标的轴承退化阶段自适应划分方法[J].哈尔滨工程大学学报,2025(9): 1-13.
- [7] 茆诗松,王静龙,濮晓龙.高等数理统计[M].第3版.北京:高等教育出版社,2022.