

基于变分贝叶斯的数据分类算法

张文倩¹, 王 瑛¹, 张红梅^{2✉}, 宋增杰³

(1. 空军工程大学装备管理与安全工程学院, 西安, 710051; 2. 空军工程大学理学院, 西安, 710051;

3. 西安交通大学数学与统计学院, 西安, 710049)

摘要 随着互联网技术的高速发展, 数据库的规模和复杂度不断增长, 传统的分类方法已经不能满足复杂数据的分类需求, 针对此类问题, 提出了一种基于变分贝叶斯的数据分类算法。该算法在传统贝叶斯推断上引入变分近似理论, 结合最大期望算法思想, 利用统计物理中的平均场理论, 并以混合高斯模型为例进行了实验仿真。实验结果证明, 随机生成数据在经过 382 次迭代后, 能明显看出由 3 组高斯模型混合而成, 似然函数的下界随迭代次数增加不断上升, 在 350 次迭代后曲线与预想一样趋于平缓, 并且在误差允许的范围内得到接近真实数据的均值和逆协方差矩阵, 实现其分类处理。在保证高精度的要求下计算速度更快、效率更高、更符合实际工程的应用背景。

关键词 变分贝叶斯; 分类算法; 最大期望算法

DOI 10.3969/j.issn.1009-3516.2017.02.015

中图分类号 TP391 **文献标志码** A **文章编号** 1009-3516(2017)02-0089-06

A Data Classification Algorithm Based on Variational Bayesian

ZHANG Wenqian¹, WANG Ying¹, ZHANG Hongmei², SONG Zengjie³

(1. Equipment Management and Safety Engineering College, Air force Engineering University, Xi'an 710051, China; 2. Science College, Air force Engineering University, Xi'an 710051, China;

3. School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710051, China)

Abstract: With the rapid development of Internet technology, the size and complexity of the database are continually growing, the traditional classification method can no longer meet the demand of the classification of complex data. For this reason, a data classification algorithm based on variational Bayesian is proposed. This paper introduces the variational approximation theory on the basis of traditional Bayesian inference, combines with the thought of maximum expected algorithm, utilizes the mean field theory in the statistical physics, and simulates taking Gaussian mixture model as an example. The experimental results show that the randomly generated data are composed of the three Gaussian models mixed after 382 iterations, the lower bound of likelihood function rises with the increase of iteration number, the curve becomes flat as expectation after 350 iterations, and the mean value and the inverse of covariance matrix close to the real data are obtained in the range of allowable error. Under the requirement of high precision, the calcula-

收稿日期: 2016-09-18

基金项目: 国家自然科学基金(71171199)

作者简介: 张文倩(1993—), 女, 山西临猗人, 硕士生, 主要从事系统工程研究。E-mail: wenqian_z@163.com

通信作者: 张红梅(1967—), 女, 山西临猗人, 教授, 主要从事网络信息处理研究。E-mail: zhm_plum@163.com

引用格式: 张文倩, 王瑛, 张红梅, 等. 基于变分贝叶斯的数据分类算法[J]. 空军工程大学学报(自然科学版), 2017, 18(2): 89-94. ZHANG Wenqian, WANG Ying, ZHANG Hongmei, et al. A Data Classification Algorithm Based on Variational Bayesian[J]. Journal of Air Force Engineering University(Natural Science Edition), 2017, 18(2): 89-94.

tion speed is faster, calculation efficiency is higher, and all of these accord with the demands of actual engineering application background.

Key words: variational Bayesian; classification algorithm; EM algorithm

随着互联网技术的飞速发展,数据收集变得越来越容易,导致数据库规模越来越大、复杂性及维度也越来越高。如何从大量的数据中提取隐含的、事先未知的、并且潜在有用的数据,是挖掘技术需要解决的问题^[1]。目前数据挖掘已广泛应用于商务、金融、医疗、通讯等行业,并展现出了其强大的知识发现的能力,但对于复杂数据的挖掘,传统的数据处理方法难以得到好的效果,使得对复杂数据的挖掘研究意义重大,引起了学术界的广泛关注^[2]。

目前复杂数据挖掘研究的热点主要集中在数据分类上。分类学习研究是数据挖掘的一个重要方向,其目的是根据有类标的训练样本学习一个分类器。相关算法研究成果非常丰富,其中代表性的分类算法包括决策树分类算法、贝叶斯分类算法、传统的基于关联规则分类算法、支持向量机分类算法等^[3]:决策树分类算法能够很好地处理噪声数据,但只对规模较小训练样本集有效;贝叶斯分类算法精度高、速度快,错误率低,但分类不够准确;传统的基于关联规则分类算法准确率高,但容易受硬件内存的制约;支持向量机分类算法准确率高、复杂度低,但速度慢等^[4-5]。每一种算法都有其优点和缺陷,但针对上述算法处理复杂数据存在的问题,本文提出了一种基于变分贝叶斯的数据分类算法,快速准确地实现了复杂数据的分类处理。

1 变分贝叶斯原理

变分贝叶斯方法是由变分法与贝叶斯理论结合而成的一类用于贝叶斯估计的方法,是在传统贝叶斯推断与EM算法的基础上引入变分近似理论而提出的^[6-7]。其中EM算法是采用直接估计参数最优值的方式,而变分贝叶斯方法则是通过一个恰当的先验分布,计算其超参数值,最终计算出联合概率分布函数的各个参数^[8]。这种方法既能利用先验信息,又可通过迭代使计算复杂度显著降低,因此被广泛用于含有隐变量的统计模型的近似估计,包括随机系统的参数估计,盲源分离和语音增强领域。

统计模型中,已知一组观测数据 D ,确定这组观测数据的参数与潜变量 $Z = \{Z_1, Z_2, \dots, Z_n\}$ 的后验分布 $P(Z|D)$ 。考虑到后验概率 $P(Z|D)$ 的高维复杂性,我们用一种简单的模型形式 $Q(Z)$ 来代替它,即 $P(Z|D) \approx Q(Z)$ 。对于 $Q(Z)$ 与 $P(Z|D)$ 之间差

异性的度量,信息论中已经有现成的度量指标——相对熵^[4],即 KL 散度,它是用来度量两个随机分布 P 与 Q 之间的距离。 $D(P||Q)$ 中真实分布用 P 描述,则 Q 分布与 P 分布的 KL 散度为:

$$\ln P(D) = D_{KL}(Q||P) - \sum_Z Q(Z) \ln \frac{Q(Z)}{P(Z,D)} = D_{KL}(Q||P) + L(Q) \quad (1)$$

要使2个分布之间的距离最短,在边缘对数似然函数不变的前提下,则需要使得下界 $L(Q)$ 达到极大值,见图1。

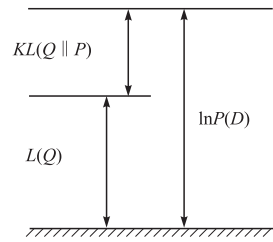


图1 对数似然函数与证据下界的关系

Fig. 1 The relationship between Logarithm likelihood function and evidence lower bound

除此之外,考虑下界的计算复杂度,则需要经过分析 Q 的取值来降低 $L(Q)$ 的计算难度。为了方便描述,我们将获得的后验 $P(Z|D)$ 的近似值,以及计算得出的下界 $L(Q)$,统称为变分自由能^[9]:

$$L(Q) = \sum_Z Q(Z) \ln P(Z,D) - \sum_Z Q(Z) \ln Q(Z) = E_Q[\ln P(Z,D)] + H(Q) \quad (2)$$

为了方便表达,我们定义等式右边首项为能量,定义接着的下一项为信息熵。

而对于求解高维复杂积分的函数的近似简化形式,统计学中的物理分支已经给出了现有的理论——平均场理论^[10]。简单讲就是:在一个已知的宏观系统里,处在个体之间局部的相互作用会导致影响宏观层面的结果,使得其间产生较为平稳固定的作用。于是我们可以作出后验条件独立的假设。即:

$$\forall i, p(Z|D) = p(Z_i|D) p(Z_{-i}|D) \quad (3)$$

为了方便计算,将其化为对数形式,最终的积分表达式为:

$$L(Q(Z)) = \int \left(\prod_i Q_i(Z_i) \right) \ln P(Z,D) dZ - \int \left(\prod_k Q_k(Z_k) \right) \sum_i \ln Q_i(Z_i) dZ \quad (4)$$

式中: $Q(Z) = \prod_i Q_i(Z_i)$ 且满足 $\forall i, \int Q_i(Z_i) dZ_i = 1$ 。

另外考虑到划分 $Z = \{Z_i, Z_{-i}\}$, 其中 $Z_{-i} = Z \setminus Z_i$, 先进行能量项的计算:

$$E_{Q(Z)}[\ln P(Z, D)] = \int Q_i(Z_i) \ln Q_i^*(Z_i) dZ_i + \ln C \quad (5)$$

定义 $Q_i^*(Z_i) = \frac{1}{C} \exp \langle \ln P(Z, D) \rangle_{Q_{-i}(Z_{-i})}$, C 为归一化常数。再考虑熵量:

$$H(Q(Z)) = - \sum_i \int Q_i(Z_i) \ln Q_i(Z_i) dZ_i \quad (6)$$

此时得到泛函,

$$L(Q(Z)) = -D_{KL}(Q_i(Z_i) || Q_i^*(Z_i)) + H[Q_{-i}(Z_{-i})] + \ln C \quad (7)$$

表达式转换为一个泛函问题的求解, 在符合约束条件的情况下求函数极值, 表达式如下:

$$\forall i, \frac{\partial}{\partial Q_i(Z_i)} \{-D_{KL}[Q_i(Z_i) || Q_i^*(Z_i)] - \lambda_i (\int Q_i(Z_i) dZ_i - 1)\} = 0 \quad (8)$$

显然在 K_L 散度为 0 的情况下 $L(D)$ 值达到最大, 表达式如下:

$$Q_i(Z_i) = Q_i^*(Z_i) = \frac{1}{C} \exp \langle \ln P(Z_i, Z_{-i}, D) \rangle_{Q_{-i}(Z_{-i})} \quad (9)$$

式中: C 为归一化常数, $C = \int \exp \langle \ln P(Z_i, Z_{-i}, D) \rangle_{Q_{-i}(Z_{-i})} dZ_{-i}$, $Q(Z_i)$ 为除划分 Z_i 自身外其他所有划分的对数期望的联合概率密度函数^[11]。又可以写为:

$$\ln Q_i(Z_i) = \langle \ln P(Z_i, Z_{-i}, D) \rangle_{Q_{-i}(Z_{-i})} + \text{const} \quad (10)$$

2 应用实例仿真

以高维数据中的混合高斯模型^[12-13]为例, 其可表示为 M 个高斯分布的线性组合:

$$P(x | \boldsymbol{\pi}, \boldsymbol{\mu}, \mathbf{T}) = \sum_{i=1}^M \pi_i N(x | \mu_i, T_i) \quad (11)$$

式中: $\boldsymbol{\pi} = \{\pi_i\}_{i=1}^M$, $\boldsymbol{\mu} = \{\mu_i\}_{i=1}^M$, $\mathbf{T} = \{T_i\}_{i=1}^M$; π_i 为混合系数, 满足 $0 < \pi_i < 1$ 且 $\sum_{i=1}^M \pi_i = 1$; μ_i 为均值; T_i 为逆协方差矩阵系数。

假设观测数据集 $D = \{x_n\}_{n=1}^N$ 中的数据由上述混合分布独立采样得到, 则关于模型参数 $(\boldsymbol{\pi}, \boldsymbol{\mu}, \mathbf{T})$ 的似然函数为:

$$P(D | \boldsymbol{\pi}, \boldsymbol{\mu}, \mathbf{T}) = \prod_{n=1}^N \left[\sum_{i=1}^M \pi_i N(x_n | \mu_i, T_i) \right] \quad (12)$$

引入隐变量 $s = \{s_m\}_{m=1, n=1}^{M \times N}$ 描述各数据与各高斯分量之间隶属关系, 即对每一个数据 x_n, s_m :

若 x_n 来自第 j 个高斯分量, 则:

$$s_m = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad (13)$$

于是, 各观测变量与非观测变量之间的依赖关系^[14]可表示为图 2。

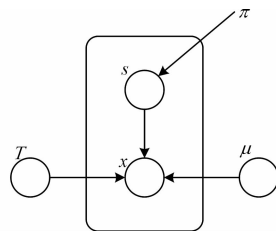


图 2 观测与非观测变量之间的依赖关系

Fig. 2 Dependency between observable variable and unobservable variable

以混合系数为条件的所有随机变量的联合分布:

$$P(D, \boldsymbol{\mu}, T, s | \boldsymbol{\pi}) = P(D | \boldsymbol{\mu}, T, s) P(s | \boldsymbol{\pi}) P(\boldsymbol{\mu}) P(T) \quad (14)$$

则先验分布如下:

数据的先验:

$$P(D | \boldsymbol{\mu}, T, s) = \prod_{n=1}^N \prod_{i=1}^M N(x_n | \mu_i, T_i)^{s_{ni}} \quad (15)$$

隐变量的先验:

$$P(s | \boldsymbol{\pi}) = \prod_{n=1}^N \prod_{i=1}^M \pi_i^{s_{ni}} \quad (16)$$

均值的先验:

$$P(\boldsymbol{\mu}) = \prod_{i=1}^M N(\mu_i | 0, \beta I) \quad (17)$$

逆协方差矩阵的先验:

$$P(T) = \prod_{i=1}^M W(T_i | \nu, V) \quad (18)$$

因为直接最大化边缘似然难于操作, 因此考虑最大化其下界 记 $\theta = \{\boldsymbol{\mu}, T, s\}$, 则边缘似然函数可表示为:

$$P(D | \boldsymbol{\pi}) = \int P(D, \theta | \boldsymbol{\pi}) d\theta \quad (19)$$

引入 $Q(\theta)$ 作为参数 θ 的真实后验分布 $P(\theta | D, \boldsymbol{\pi})$ 的近似值, 再利用 Jensen 不等式可得:

$$\ln P(D | \boldsymbol{\pi}) = \ln \int Q(\theta) \frac{P(D, \theta | \boldsymbol{\pi})}{Q(\theta)} d\theta \geq \int Q(\theta) \ln \frac{P(D, \theta | \boldsymbol{\pi})}{Q(\theta)} d\theta = L(Q) \quad (20)$$

只需寻找恰当的 Q 值, 使得下界 $L(Q)$ 尽可能的大。

事实证明, 由于对数似然函数 $\ln P(D | \boldsymbol{\pi})$ 与 Q 无关, 所以最大化 $L(Q) \Leftrightarrow$ 最小化 $K_L(Q | P)$ 。为简化计算, 假定 $Q(\theta)$ 关于各变量 θ_i 可以因式分解:

$$Q(\theta) = \prod_i Q_i(\theta_i) \quad (21)$$

对于 $Q_j = Q_j(\theta_j)$, 将下界 $L(Q)$ 分解为含 Q_j 和不含 Q_j 的两部分并关于 Q_j 优化:

$$L(Q) = \int Q_j \left\{ \int \ln P(D, \theta | \pi) \prod_{i \neq j} Q_i d\theta_i \right\} d\theta_j - \int Q_j \ln Q_j d\theta_j + \text{const}$$

记为:

$$\langle \ln P(D, \theta | \pi) \rangle_{i \neq j} = \int \ln P(D, \theta | \pi) \prod_{i \neq j} Q_i d\theta_i \quad (22)$$

$$\ln \tilde{P}(D, \theta_j | \pi) = \langle \ln P(D, \theta | \pi) \rangle_{i \neq j} + \text{const} \quad (23)$$

则变分下界可改写为:

$$L(Q) = \int Q_j \ln \tilde{P}(D, \theta_j | \pi) d\theta_j - \int Q_j \ln Q_j d\theta_j + \text{const} = -K_L(Q_j \| \tilde{P}) + \text{const} \quad (24)$$

要最大化 $L(Q)$, 只需最小化 $D_{KL}(Q_j | \tilde{P})$, 于是 $Q_j = \tilde{P}(D, \theta_j | \pi)$ 时 $L(Q)$ 取最大, 即:

$$\ln Q_j^*(\theta_j) = \ln \tilde{P}(D, \theta_j | \pi) = \langle \ln P(D, \theta | \pi) \rangle_{i \neq j} + \text{const} \quad (25)$$

值得注意的是, $Q_j^*(\theta_j)$ 作为概率密度函数还应满足以下条件:

$$\int Q_j^*(\theta_j) d\theta_j = 1 \quad (26)$$

由此可确定归一化常数, 因此最优变分后验 $Q_j^*(\theta_j)$ 应满足下面的表达式为:

$$Q_j^*(\theta_j) = \frac{\exp \langle \ln P(D, \theta | \pi) \rangle_{i \neq j}}{\int \exp \langle \ln P(D, \theta | \pi) \rangle_{i \neq j} d\theta_j} \quad (27)$$

2.1 算法说明

1) 利用变分贝叶斯方法求解高斯混合模型, 需要确定的未知量为隐变量 s , 各高斯分量的均值和逆协方差矩阵 $\{\mu, T\}$, 混合系数 π 与模型阶数 M ;

2) $\{s, \mu, T\}$ 的求解利用最优变分后验满足式(25):

隐变量 s 的最优变分先验:

$$Q_s(s) = \prod_{n=1}^N \prod_{i=1}^M p_{in}^{s_{in}} \quad (28)$$

均值 μ 的最优变分先验:

$$Q_\mu(\mu) = \prod_{i=1}^M N(\mu_i | m_\mu^{(i)}, T_\mu^{(i)}) \quad (29)$$

逆协方差矩阵 T 的最优变分先验:

$$Q_T(T) = \prod_{i=1}^M W(T_i | \nu_T^{(i)}, V_T^{(i)}) \quad (30)$$

3) π 通过对变分下界利用拉格朗日乘子法得到; 仅考虑变分下界 $L(Q)$ 中含 π_i 的项:

$$\langle \ln P(s) \rangle_s = \sum_{i=1}^M \sum_{n=1}^N \langle s_{in} \rangle \ln \pi_i \quad (31)$$

再利用约束条件 $\sum_{i=1}^M \pi_i = 1$, 可构造如下形式的目标函数:

$$L(\pi, \lambda) = \sum_{i=1}^M \sum_{n=1}^N \langle s_{in} \rangle \ln \pi_i + \lambda \left(\sum_{i=1}^M \pi_i - 1 \right) \quad (32)$$

求解如下 2 个方程:

$$\begin{cases} \frac{\partial L}{\partial \pi_j} = \sum_{n=1}^N \langle s_{jn} \rangle \frac{1}{\pi_j} + \lambda = 0 \\ \sum_{i=1}^M \pi_i = 1 \end{cases} \rightarrow \pi_j = \frac{1}{N} \sum_{n=1}^N \langle s_{jn} \rangle = \frac{1}{N} \sum_{n=1}^N p_{jn} \quad (j = 1, \dots, M) \quad (33)$$

4) M 在算法迭代过程中通过删减混合系数较小的分量自动确定

5) 变分下界 $L(Q)$ 的计算:

$$\begin{aligned} L(Q) &= \langle \ln P(D, \mu, T, s) \rangle_{\mu, T, s} - \langle \ln Q(\mu, T, s) \rangle_{\mu, T, s} = \\ &= \langle \ln P(D | \mu, T, s) \rangle_{\mu, T, s} + \langle \ln P(s) \rangle_s + \\ &= \langle \ln P(\mu) \rangle_\mu + \langle \ln P(T) \rangle_T - \langle \ln Q_s(s) \rangle_s - \\ &= \langle \ln Q_\mu(\mu) \rangle_\mu - \langle \ln Q_T(T) \rangle_T \end{aligned} \quad (34)$$

将各个先验分布与变分后验分布的表达式代入上式即可计算出变分下界^[15]。

2.2 算法流程

1) 初始参数设置。分量个数、混合系数、误差容忍度、最大迭代次数、分量移除阈值。

先验分布: β, ν, V

后验分布: $p, m_\mu^{(i)}, T_\mu^{(i)}, \nu_T^{(i)}, T_T^{(i)}$

2) E step。计算:

$$Q_s(s) = \prod_{n=1}^N \prod_{i=1}^M p_{in}^{s_{in}}; Q_\mu(\mu) = \prod_{i=1}^M N(\mu_i | m_\mu^{(i)}, T_\mu^{(i)});$$

$$Q_T(T) = \prod_{i=1}^M W(T_i | \nu_T^{(i)}, V_T^{(i)})$$

3) M step。计算:

$$\pi_j = \frac{1}{N} \sum_{n=1}^N p_{jn}; j = 1, 2, \dots, M \quad (35)$$

4) 依据 π_j 的大小删减分量, 并估计变分下界 $L(Q)$ 。

5) 终止条件。下界 $L(Q)$ 变化小或达最大迭代次数时终止; 否则, 转步骤 2)。

2.3 算法图示

在高斯混合模型算法图 3 中, E 步骤中由于 KL 散度不断减小, 似然下界 $L(Q)$ 不断上升, 其中红色横线代表似然下界, 黑色横线代表边缘似然函数。M 步骤中由于参数的更新, 随着迭代次数增加, 不断产生新的变分下界。这两步骤交替进行, 直到最终下界趋于稳定停止。

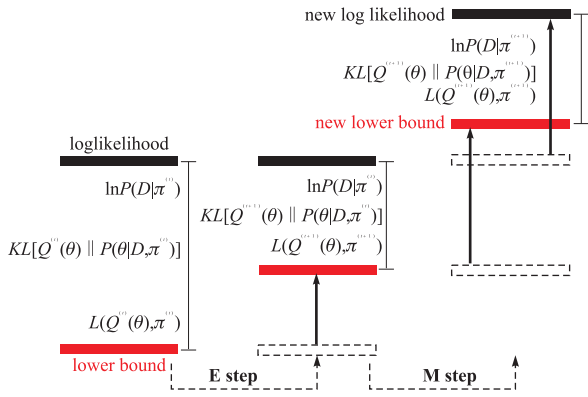


图 3 高斯混合模型中变分贝叶斯算法图

Fig. 3 Variational Bayesian algorithm diagram of Gaussian mixture model

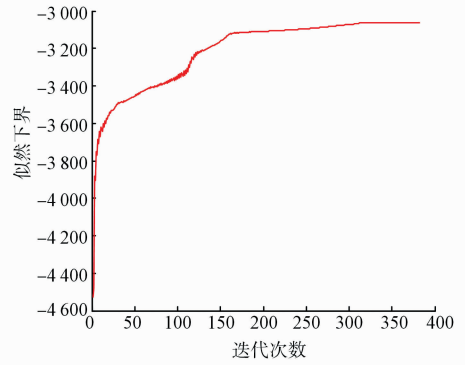


图 6 似然下界上升曲线

Fig. 6 Likelihood lower bound rising curve

3 实验结果及分析

3.1 实验结果

模拟一个 3-高斯混合模型(协方差相同),原始数据见图 4,最终结果见图 5,似然下界上升曲线见图 6,真实和模拟数据下的均值与协方差矩阵见表 6、表 7。

表 1 真实和模拟数据下的均值

Tab. 1 Mean value under the real and simulated data

真实	$[0, -2]$	$[0, 0]$	$[0, 2]$
		$[0.04, -2.01]$	
模拟	$[-0.06, -0.06]$	$[-0.03, 1.98]$	

表 2 真实和模拟数据下的协方差矩阵

Tab. 2 Covariance matrix under the real and simulated data

真实	$\begin{pmatrix} 2 & 0 \\ 0 & 0.2 \end{pmatrix}$
模拟	$\begin{pmatrix} 2.06 & -0.01 \\ -0.01 & 0.19 \end{pmatrix} \begin{pmatrix} 2.26 & 0.07 \\ 0.07 & 0.22 \end{pmatrix}$
	$\begin{pmatrix} 1.96 & 0.10 \\ 0.10 & 0.23 \end{pmatrix}$

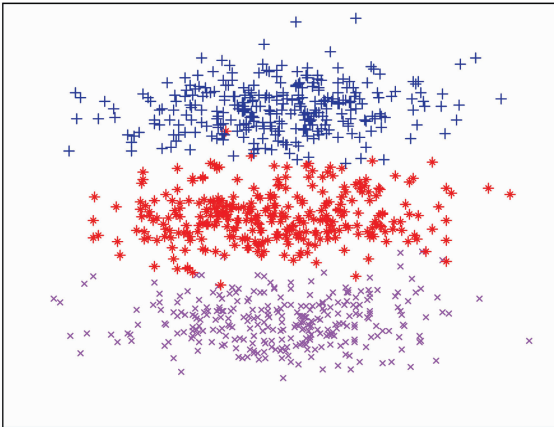


图 4 原始数据

Fig. 4 Original data

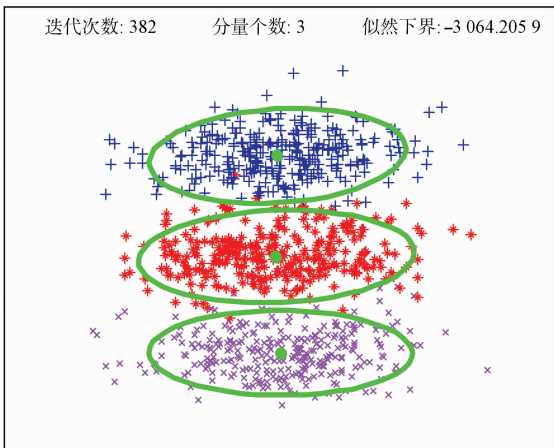


图 5 最终结果

Fig. 5 Final result

3.2 实验评价

本次实验中我们先随机生成数据如图 4, 然后通过模拟可以从图 5 看到经过 382 次迭代, 能较明显地看出是由 3 组高斯模型混合而成, 并且以均值为圆心, 逆协方差矩阵为半径得到的椭圆结果在误差允许的范围与真实数据相差不大, 模拟结果比较理想。除此之外, 如图 6 似然函数的下界不断随迭代次数增加不断上升, 在 350 次迭代后曲线趋于平缓, 与该方法预想结果基本一样。真实和模拟数据下的均值和协方差矩阵分别见表 1 和表 2, 已知一组观测数据 D , 使用变分贝叶斯方法来确定高斯混合模型中的未知量包括隐变量 s , 各高斯分量的均值和逆协方差矩阵 $\{\mu, \mathbf{T}\}$, 混合系数 π 与模型阶数 M , 实现复杂数据的分类是具有可行性和高效性的, 值得广泛推广。

4 结语

本文提出了一种基于变分贝叶斯的数据分类算

法,并进行了实验仿真。仿真结果表明,该方法可以精确、快速地实现复杂数据的分类及参数估计。与此同时,仍有许多待解决的问题值得深入研究:

1)传统的变分贝叶斯方法对模型的推导是繁琐而复杂的,在推导边缘概率的时候,我们只需要关心所求参数的马尔科夫毯上的节点,因此我们希望对数联合概率在一些参数下的期望可以简化。

2)若近似后验分布在整个概率密度函数空间取值,则要对其进行约束。应用变分贝叶斯学习原理,在工程领域中有较好的实用价值,今后的研究工作将围绕进一步提高算法的实时性和应用价值展开。

参考文献(References):

- [1] 吴庆耀. 高维数据的若干分类问题及算法研究[D]. 哈尔滨:哈尔滨工业大学,2013.
WU Q Y. Research on Classification Problems and Algorithms for High Dimensional data [D]. Harbin: Harbin Institute of Technology,2013. (in Chinese)
- [2] CHENG F, LI J Y. A Multispectral Remote Sensing Data Spectral Unmixing Algorithm Based on Variational Bayesian ICA[J]. Journal of the Indian Society of Remote Sensing,2013,41(2):259-268.
- [3] BISHOP C M. Pattern Recognition and Machine Learning[M]. New York: Springer, 2006.
- [4] TARA B, ELIZABETH J, CROSS K, et al. Variational Bayesian Mixture of Experts Models and Sensitivity Analysis for Nonlinear Dynamical Systems[J]. Mechanical Systems and Signal Processing,2016,S66-67:178-200.
- [5] 李伶俐. 数据挖掘中分类算法综述[J]. 重庆师范大学学报(自然科学版),2011(4):44-47.
LI L L, A Review on Classification Algorithms in Data Mining[J]. Journal of Chongqing Normal University (Natural Science), 2011(4):44-47. (in Chinese)
- [6] BAS, ALAMEDA-PINEDA X, XOMPERO A, et al. An On-line Variational Bayesian Model for Multi-Person Tracking from Cluttered Scenes [J]. Computer Vision and Image Understanding, <http://dx.doi.org/10.1016/j.cviu.2016.07.006>.
- [7] BEAL M J. Variational Algorithms for Approximate Bayesian inference[D]. London: University of London, 2003.
- [8] JAAKKOLAT S, JORDAN M I. Bayesian Parameter Estimation via Variational Methods[J]. Statistics and Computing, 2000, 10(1):25-37.
- [9] WATANABE K, SHIGA M, WATANABE S. Upper Bound for Variational Free Energy of Bayesian Networks[J]. Machine Learning, 2009, 75(2):199-215.
- [10] LI C F, YIN J Y, BAI C S. Variational Bayesian Independent Component Analysis for Spectral Unmixing in Remote Sensing Image[J]. Arabian Journal of Geosciences, 2013, 6(4):1119-1129.
- [11] LI C F, YIN J Y. A Multispectral Remote Sensing Data Spectral Unmixing Algorithm Based on Variational Bayesian ICA[J]. Journal of the Indian Society of Remote Sensing, 2013, 41(2):259-268.
- [12] SHENC, XU D, HUANG W, et al. An Interacting Multiple Model Approach for State Estimation with Non-Gaussian Noise Using a Variational Bayesian Method[J]. Asian Journal of Control, 2015, 17(4):1424-1434.
- [13] 徐定杰,沈忱,沈锋. 混合高斯分布的变分贝叶斯学习参数估计[J]. 上海交通大学学报,2013(7):1119-1125.
XU D J, SHEN C, SHEN F. Variational Bayesian Learning for Parameter Estimation of Mixture of Gaussians. [J]. Journal of Shanghai Jiaotong University, 2013(7):1119-1125. (in Chinese)
- [14] ARAKAWA A, TANIGUCHI M, HAYASHI T, et al. Variational Bayesian Method of Estimating Variance Components[J]. Animal Science Journal. 2016, 87(7):863-872.
- [15] LING H, STONE D B. Time-Varying Forecasts by Variational Approximation of Sequential Bayesian Inference[J]. Quantitative Finance, 2015(1):43-67.

(编辑:姚树峰)