山东大学机器学习课程

实验报告

——实验三：最大似然估计与贝叶斯参数估计

姓名：刘梦源

学院：计算机科学与技术学院

班级：计算机 14.4

学号：201400301007

一、实验目的：

（1）熟悉最大似然估计与贝叶斯参数估计的原理
（2）实现最大似然估计与贝叶斯参数估计的程序设计，并比较他们的异同
（3）反思归纳，将参数估计的方法融会贯通到今后的机器学习知识中，从而针对不同的问题需求选择恰当的参数估计方法

二、实验环境：

（1）硬件环境：

英特尔® 酷睿™ i7-7500U 处理器
512 GB PCIe® NVMe™ M.2 SSD
8 GB LPDDR3-1866 SDRAM

（2）软件环境：

Windows10 家庭版 64 位操作系统
Matlab R2016a

三、实验内容

1、最大似然估计

（1）假定：

①待估参数θ是确定的未知量
②按类别把样本分成 M 类 X1，X2，X3，… XM 其中第 i 类的样本共 N 个 Xi =(X1,X2,… XN)T ，并且是独立从总体中抽取的。
③Xi 中的样本不包含 （i≠j）的信息，所以可以对每一类样本独立进行处理。
④ 第 i 类的待估参数 θi=(θ1, θ2.... θn)T

根据以上四条假定，可以利用第 i 类学习样本来估计第 i 类的概率密度，其它类的概率密度由其它类的学习样本来估计。

（2）一维时：

A. 正态总体均值的最大似然估计即为学习样本的算术平均
B. 正态总体方差的最大似然估计与样本的方差不同。当 N 较大的时候，二者的差别不大。
C. 多维情况：n 个特征时，估计值的公式如下：

$$\hat{\mu} = \frac{1}{n}\sum_{k=1}^{n} x_k$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{n} (X_k - \hat{\mu})(X_k - \hat{\mu})^T$$

2、贝叶斯参数估计

估计步骤:

1. 确定θ的先验分布P(θ),待估参数为服从某一分布的概率密度。

2. 计算 $p(\theta|D)$：

$$p(\theta|D)=\frac{p(D|\theta)p(\theta)}{\int p(D|\theta)p(\theta)d\theta}$$

$$p(D|\theta)=\prod_{k=1}^{n}p(x_k|\theta)$$

3. 计算 $p(x|D)$：

$$p(x|D)=\int p(x|\theta)p(\theta|D)d\theta$$

四、实验结果

3.2 exercise 1

（a）w1 类别的均值方差的最大似然估计

$\hat{\mu}$

X1	X2	X3
−0.0709	−0.6047	−0.9110

$\hat{\sigma}^2$

0.9062	4.2007	4.5419

(b)w1 类别任意两个特征值的最大似然估计

x1 和 x2

$\hat{\mu}$：

-0.0709	-0.6047

$\hat{\sigma}^2$：

0.9062	0.5678
0.5678	4.2007

x2 和 x3

$\hat{\mu}$

-0.6047	-0.9110

$\hat{\sigma}^2$

4.2007	0.7337
0.7337	4.5419

x1 和 x3

$\hat{\mu}$

-0.0709	-0.9110

$\hat{\sigma}^2$

0.9062	0.3941
0.3941	4.5419

(c) w1 类别三个特征值的最大似然估计

$\hat{\mu}$

-0.0709	-0.6047	-0.9110

$\hat{\sigma}^2$

0.9062	0.5678	0.3941
0.5678	4.2007	0.7337
0.3941	0.7337	4.5419

(d) 假设三维数据的高斯模型是可分离的。W2 类别中的均值和协方差矩阵中的三个参数的最大似然估计。

均值：-0.1126　　0.4299　　0.0037

方差：0.0539　0.0460　　0.0073

因为是独立分布，我们的对协方差矩阵的最大似然估计也就变成了：

0.0539	0	0
0	0.0460	0
0	0	0.0073

(e) 对 $\hat{\mu}$ 来说，不论是否考虑多个特征向量还是单个特征向量，均没有变化。因为最大似然估计中，均值的影响不考虑特征值之间的作用。

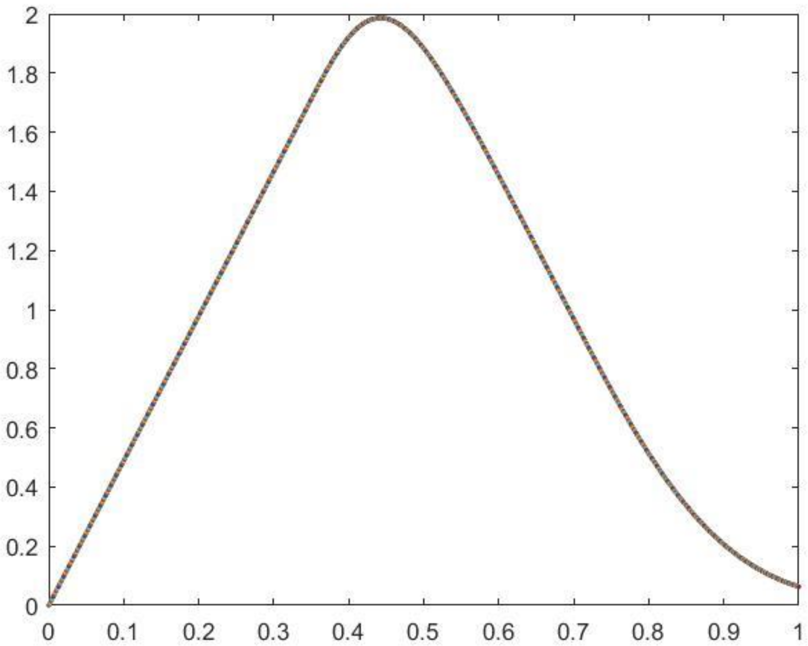
(f) 对 $\hat{\sigma}^2$ 来说，单个特征值时，方差的最大似然估计为一个数，多个特征值时变为了协方差矩阵，当特征值相互独立时，矩阵变为对角线矩阵，其余元素用零补齐。这也是可以解释的，因为协方差矩阵的非对角线元素本来描述的就是特征值之间的关系。

3.2 exercise 2

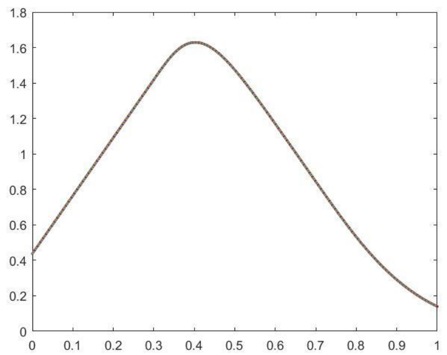
课本 P74 中有这样一段描述，“总的来说，如果我们对参数向量 θ 的真实值并不十分有把握的话，那么该方程指导我们应该把 $p(x|\theta)$ 对所有的 θ 求平均，这样得到的结果将令人满意。”

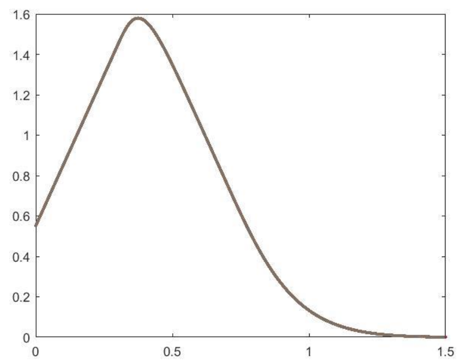
所以在这样的前提下，我们为了避免不确定的先验概率对实验结果造成过大的影响，这些影响甚至大过样本本身，我们不妨假设 θ 在一定范围内均匀分布，我们不妨假设 μ，δ 都在 0 到 1 之间均匀分布。

得到后验如下图所示。



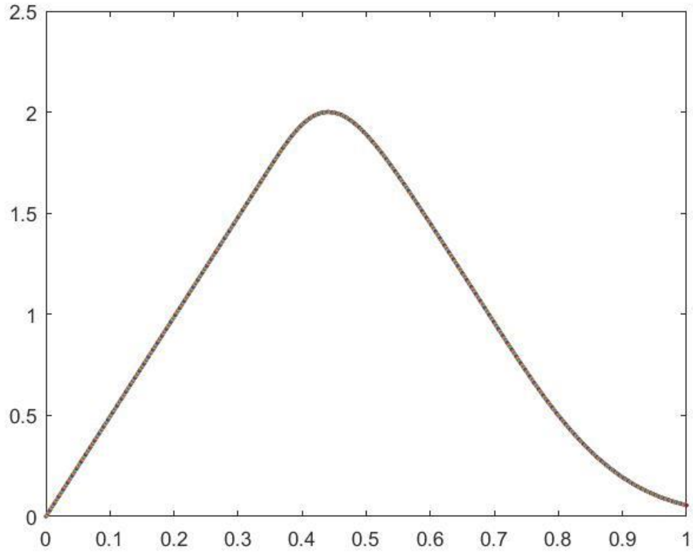
为探究先验概率对后验概率的影响，我们改变先验θ的分布。得到图像如下所示。



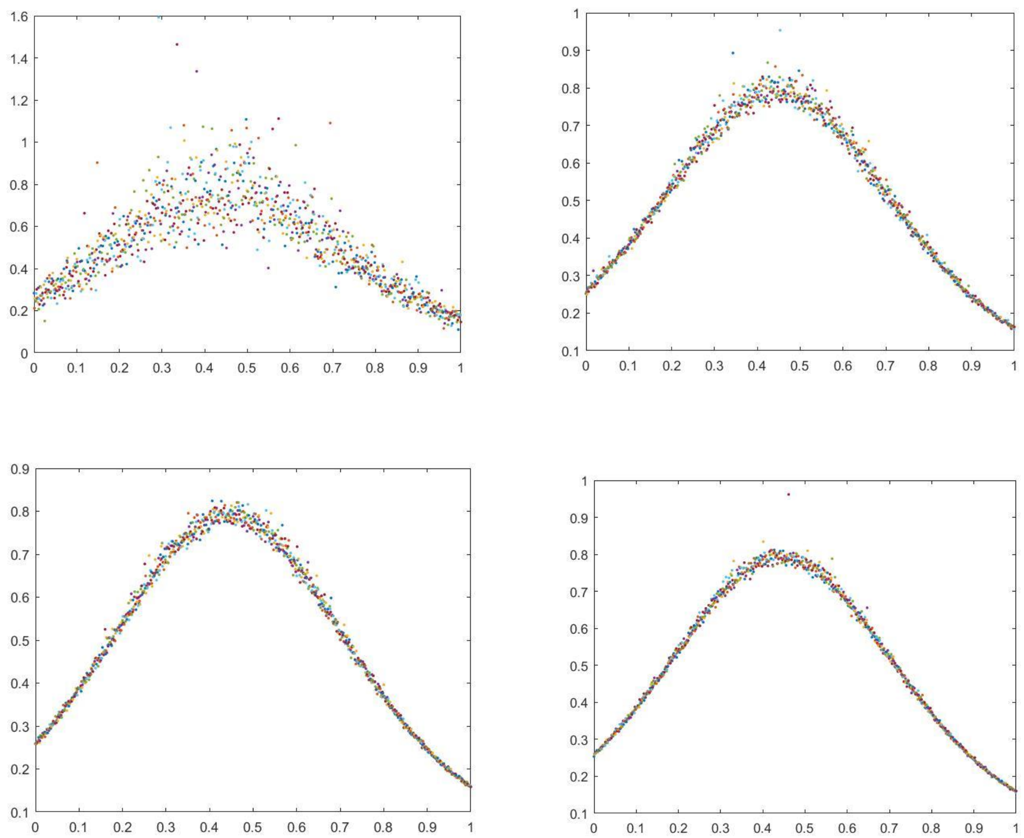


发现当调整先验的取值范围，我们依然得到近似的图像，图像的峰值差距不大，峰值点差距不大，形状基本一样，拖尾基本保持。

进一步研究，我们把参数的均匀分布替换成正态分布，但仍然尊重样本进行训练，只不过加入了主观影响。得到如图。不难发现，图像会像先验假设的分布偏移，但明显影响不大。可以得到结论，先验的假设分布只要不太离谱，贝叶斯参数估计得到的结果还是可以保持准确的。



我们还做了蒙塔卡罗仿真验证，参数服从正态分布，但是采取高斯随机点，得到下面图像。



可以发现，用高斯分布随机点代替积分的方法，也取得了很好的效果。只不过需要注意的是，蒙特卡洛选取的仿真点越多，平均下来越接近实际的结果，而蒙特卡洛也被称为是现实问题中很好的贝叶斯参数估计方法。

至此，最大似然估计和贝叶斯参数估计得到了比较理想的实验。