机器学习 课程实验报告

|  |  |  |  |
| --- | --- | --- | --- |
| 学号： | 姓名： | | 班级： |
| 实验题目： | | | |
| 实验学时：2 | | 实验日期： | |
| 实验目的：  考虑不同维数下的高斯概率密度模型，编写程序，对表格中的类的三个特征，求解最大似然估计μ和σ2 | | | |
| 硬件环境：  Intel Core（TM） i5-4210U  RAM 8G  64位操作系统，基于x64处理器 | | | |
| 软件环境：  Win 10 64bit  MATLAB | | | |
| 实验步骤与内容：   1. 对三个特征Xi，分别独立求解 2. 处理二维数据的情形 3. 处理三维数据的情形 4. 假设三维高斯模型可分离，对W2求解 5. 设计函数   function [ u,o2 ] = calculate( matrix, dim ,len)  %matrix 数据  %dim 是一个数组，存放要用到的维数  %len 数据的条数  % 初始化μ和σ  u=zeros(1,size(dim,2));  o2=zeros(size(dim,2),size(dim,2));  x = matrix(:,dim);  for index = 1:len  u = u + x(index,:);  end  u = u / len;  for index = 1:len  o2 = o2 + (x(index,:)-u)'\*(x(index,:)-u);  end  o2 = o2 / len;  end   1. 提取表格数据   %这里使用MATLAB的“导入”加载实验数据  %load lab2\_data as lab2data  w1=[];  w2=[];  c1=1;  c2=2;  selected = lab2data(:,4)==c1;  w1 = lab2data(selected,1:3);  selected = find(lab2data(:,4)==c2);  w2 = lab2data(selected,1:3);   1. 分别求解   %[result\_u,result\_o] = calculate(w1,[1],10);  %[result\_u,result\_o] = calculate(w1,[2],10);  %[result\_u,result\_o] = calculate(w1,[3],10);  %这里假设使用W1的第1,2维度数据的组合求解μ和Σ，数据条数为10  %[result\_u,result\_o] = calculate(w1,[1,2],10);  %[result\_u,result\_o] = calculate(w1,[2,3],10);  %[result\_u,result\_o] = calculate(w1,[1,3],10);  %[result\_u,result\_o] = calculate(w1,[1,2,3],10);  %[result\_u,result\_o] = calculate(w2,[1,2,3],10);  数据整理如下：     1. 解释数据   对于μ，由于求解公式中，每个维度独立求解，因此无论是独立还是多个维度同时，所得结果都相同  对于σ2，协方差矩阵中的（x,y）意义为维度x和y间的相关关系，由于无论是二维矩阵还是三维矩阵，对于相同的i值，（Xi,Yi）表示的意义及涉及的数据不变，因此结果也是相同的。至于W2，由于源数据不同，所得μ和σ2自然也与W1各不相同。 | | | |
| 结论分析与体会：  通过这次实验，我对最大似然估计有了更加详尽的认知，了解了最大似然估计的基本原理，学会了在高斯分布情况下针对不同未知数据求已知数据的方法公式，并了解了其推算过程。可以看出，协方差的最大似然估计结果是比较直观和令人满意的。之后的实验也将继续使用MATLAB。 | | | |