**机器学习实验报告**

**（三）**

****

姓名：崔玉峰

学号：201600301079

班级：2016级4班

1. **实验目的**

2.考虑对于表格中的数据进行Parzen窗估计和设计分类器。窗函数为一个球形的高斯

函数，如下所示:

ψ0((x- x;)/h)xexp[-(x- x;)'(x- x;)/(2h2)]

1. 编写程序,使用Parzen窗估计方法对一个任意的测试样本点x进行分类。对分类器的训练则使用表格中的三维数据。同时令h=1,分类样本点为(0.5,1. 0,0.0）,(0. 31,1.51,-0.50)',(-0.3,0.44,-0.1)。
2. 现在我们令h=0.1,重复(a).

3.考虑不同维数的空间中,使用k近邻概率密度估计方法的效果。

(a)编写程序，对于一 维的情况 ，当有n个数据样本点时，进行k-近邻概率密度估计。对表格中的类别a中的特征工,用程序画出当k=1,3,5时的概率密度估计结果。

(b)编写程序，对于二维的情况，当有n个数据样本点时，进行k近邻概率密度估计。

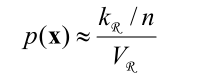
对表格中的类别wy中的特征(x,x2)' ,用程序画出当k=1,3,5时的概率密度估计结果。

(c)对表格中的3个类别的三维特征,使用k-近邻概率密度估计方法。并且对下列点处

的概率密度进行估计:(-0.41, 0.82, 0.88),(0.14，0.72，4.1),(-0.81,0.61， - 0.38)

1. **实验思路**

概率密度估计函数：



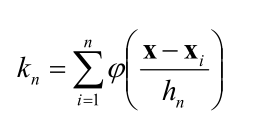
1. **Parzen窗方法：**

**本题中数据维度是3维**

① 在本题中区间为一个高斯球体：

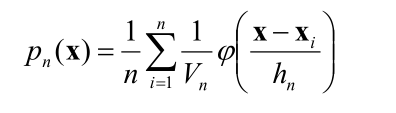
****

② 高斯球內窗函数为：





③ 则通过Parzen窗方法估计的概率密度函数为：



④ 将上述式子带入最后此题的Parzen窗估计式为：



**分类器：**

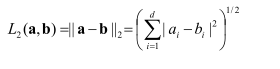
**通过编程实现上述表达式，并将样本点带入，就能求出每个样本在各类别下的Parzen窗估计概率，可以将该概率作为似然，并且假设每个类别先验概率相同，通过贝叶斯公式可知，各类别下的Parzen窗估计概率越大，该样本点就越属于该类别。**

1. **K近邻估计**
2. 通过找到样本点X的K个最近邻，可以确定概率密度估计函数的体积:



其中h为X的K个最近邻中最远的距离。d为样本的维数

1. 通过欧拉距离求出样本间的距离：

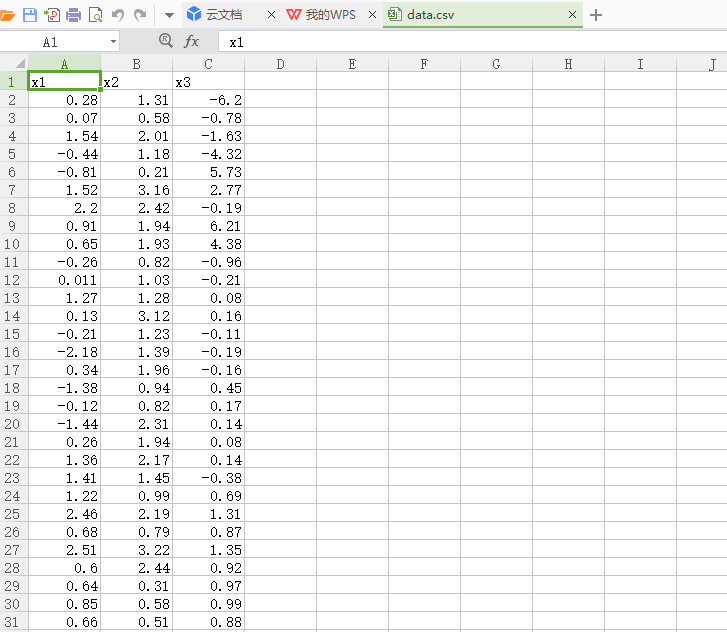


1. K近邻概率密度估计：



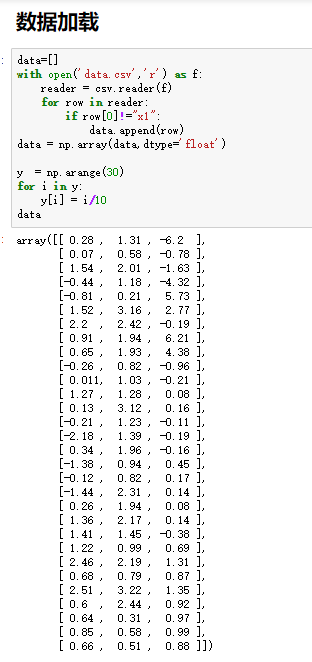
**通过编程实现上述式子就可以求得K近邻估计值，然后通过取大量的样本点求出K近邻估计值就可以画出概率密度估计曲线。**

1. **具体实现**
2. **实验环境：**
3. 编程语言：Python
4. 软件环境 ：Jupyter Notebook
5. 硬件环境 ： PC
6. **实验准备：**
7. 将上机实验所用到的数据，手动录入成csv文件方便程序编程使用，可以方便通过程序读入读出。



1. 准备Numpy库：方便进行矩阵的运算操作。
2. **实验实现：**
3. **数据加载：**

将数据30条样本读入，每个样本三维（x1,x2,x3）,前十个样本属于类，中间十个样本属于类，最后十个样本属于类，将这三十个样本存入一个数组data中，并用一个数组y存储每个样本的类别。



1. **第二题求解（Parzen窗估计）**
2. **实验总结：**

实验的代码，会以ipynb文件的形式上传可以随时进行运行和查看。

实验完成了上机实验的全部题目，并且测试通过，通过实验基本掌握了对于极大似然估计的实现和推导，不同维度下求解极大似然估计出的均值方差之间的异同。极大似然估计是重要的参数估计算法运用广泛。在开始进行实验时进展缓慢，对样本数，样本维数，样本类别数等最基本的概念也搞不清楚，在不断查询资料和回顾老师所讲的知识的过程中进步了很多，最后在弄懂了公式后编写代码，测试结果的速度也就变快，最终完成了整个实验，收获了很多。

1. **实验代码：**
2. **参考文献**

*[美]RichardO.Duda PeterE.Hart DavidG.Stork 著 模式分类 第二版*