

山东大学机器学习课程

实验报告

——实验二：贝叶斯分类器的设计与实现

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）了解分参数估计的原理和思想。

（2）用parzen窗和knn方法设计非参数估计器。

（3）编写程序，通过具体实验来分析不同非参数估计方法的异同，非参数估计和参数估计之间的异同，他们的优缺点分别是什么，以及在两种非参数化估计中，h和k对估计的结果分别是什么。

**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

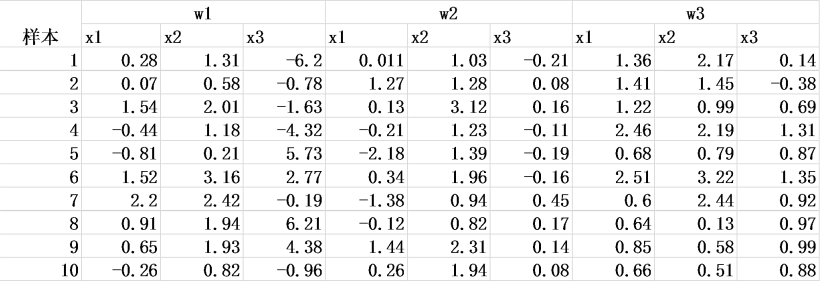
8 GB LPDDR3-1866 SDRAM

（2）软件环境：

Windows10家庭版64位操作系统

Matlab R2016a

**三、实验内容**



1. **Parzen窗非参数话估计法：**

利用课本公式P135 （11）：



1. **Knn非参数估计法：**

利用课本公式P145（32）：



对于一维情况，V为k近邻的距离的绝对值的2倍，对于二维情况，V为以k近邻的距离为半径的圆的面积，对于三位情况，V为以k近邻的距离为半径的球的体积。

1. **实验结果**
2. **Parzen窗非参数话估计法：**
3. **令h=1**

**（0.5，1.0，0.0）：**

**（0.31，1.51，-0.50）：**

**（-0.3，0.44，-0.1）：** 

**（b）令h=0.1**

**（0.5，1.0，0.0）：**

****

**（0.31，1.51，-0.50）：**

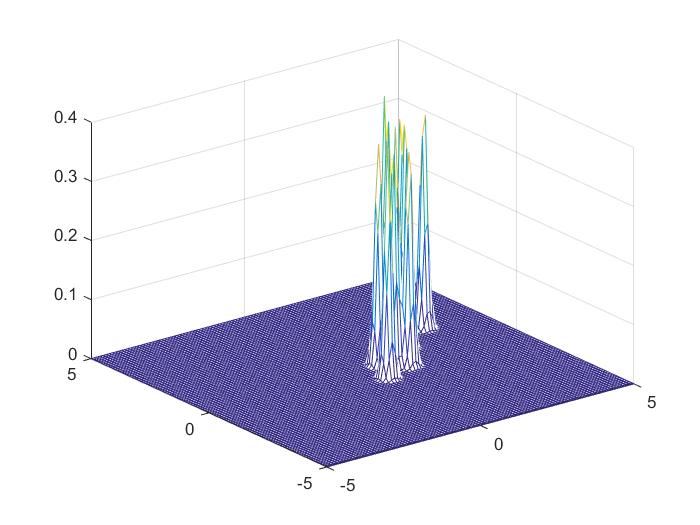
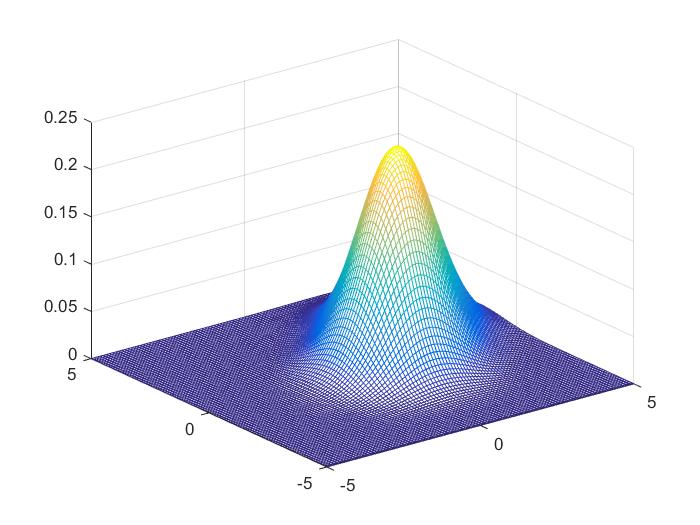
****

**（-0.3，0.44，-0.1）：**

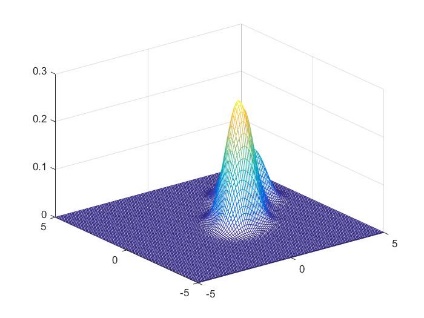
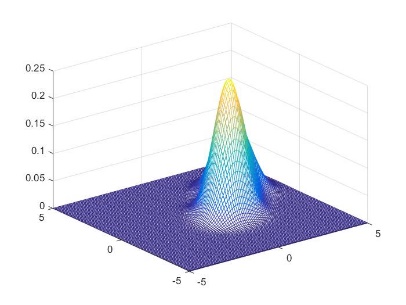
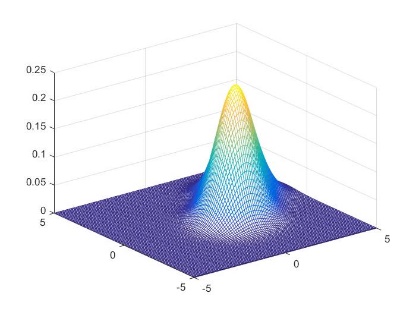


发现，h在取1和0.1时，这三个测试点都是第二类的概率密度比较大，因此把这三个测试点都分到了第二类。

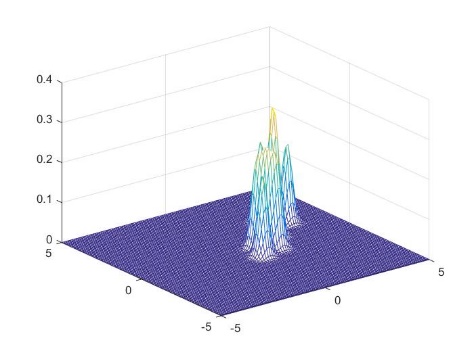
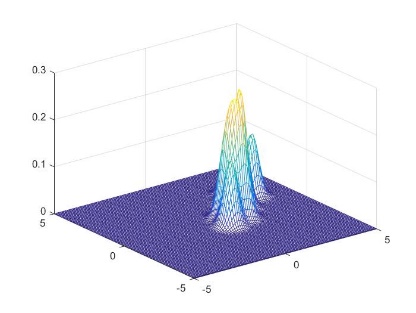
为进一步研究h的大小对非参数估计的影响，我们取第一类样本点的前两个特征值做实验。画出概率密度高程图象如下所示。可以发现当h取1时得到的较为平滑，也更为理想。

**H=1**  **H=0.1**

再调整h的大小，得到下列图像：



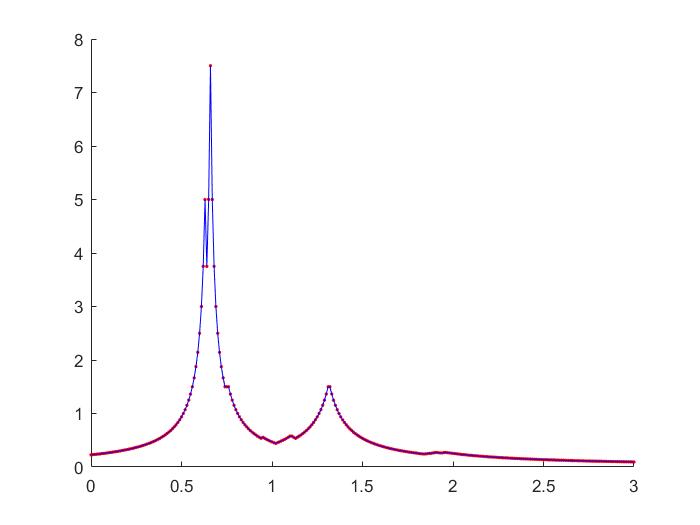
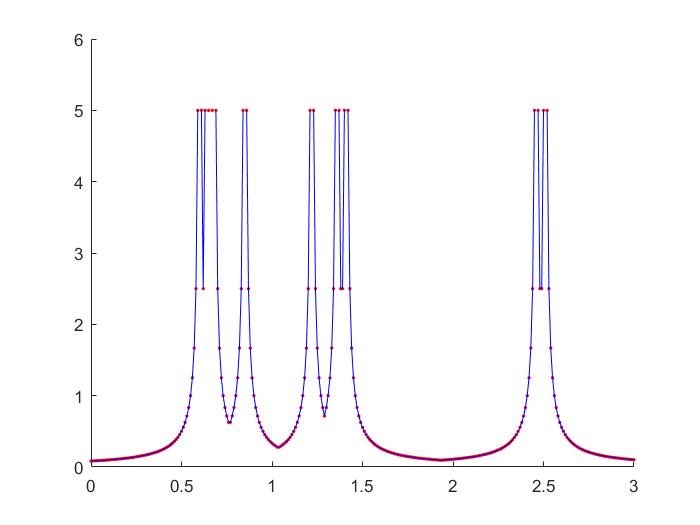
**h=0.8**   **h=0.6 h=0.45**



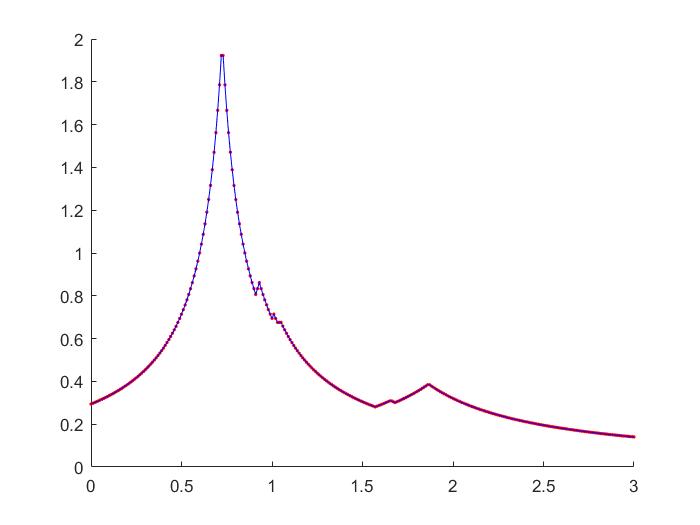
**h=0.3 h=0.2**

**（2）Knn非参数估计法：**

（a）对类别w3中的特征x1进行概率密度估计：

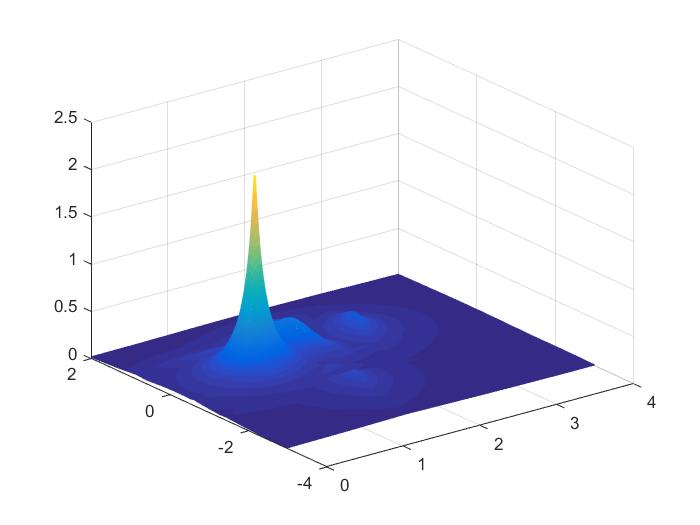
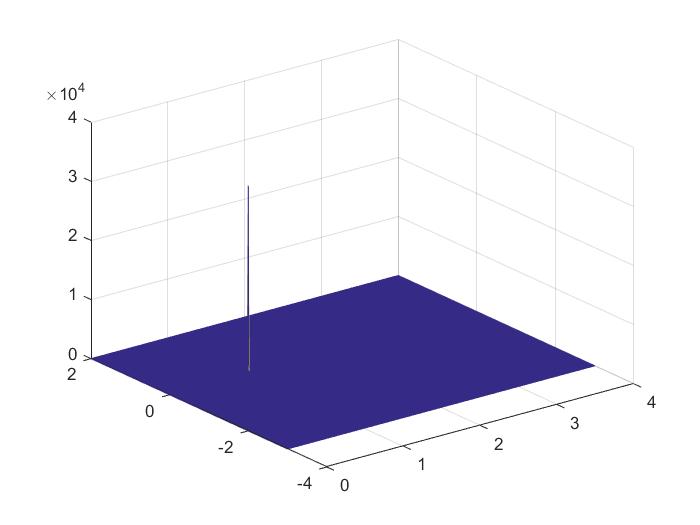


K=1 k=3

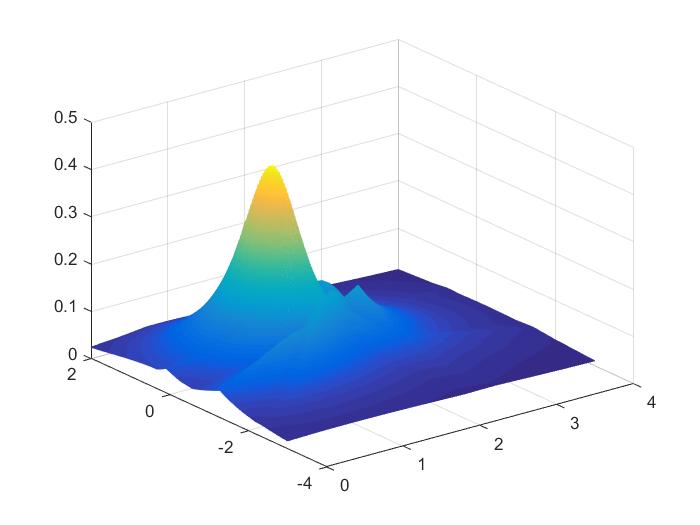


K=5

1. 对类别w2中的特征x1，x2进行概率密度估计：



K=1 k=3



K=5

1. 对三个测试点进行概率密度估计：

不妨取k=3：

x=[-0.41,0.82,0.88]在三类中的概率密度分别为



x=[0.14,0.72,4.1] 在三类中的概率密度分别为

x= [-0.81 0.61 -0.38]在三类中的概率密度分别为

至此，本次实验报告得到较为完美的结果，实验目的基本达到。