



山东大学  
SHANDONG UNIVERSITY

# 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

样本	w1			w2			w3		
	x1	x2	x3	x1	x2	x3	x1	x2	x3
1	0.28	1.31	-6.2	0.011	1.03	-0.21	1.36	2.17	0.14
2	0.07	0.58	-0.78	1.27	1.28	0.08	1.41	1.45	-0.38
3	1.54	2.01	-1.63	0.13	3.12	0.16	1.22	0.99	0.69
4	-0.44	1.18	-4.32	-0.21	1.23	-0.11	2.46	2.19	1.31
5	-0.81	0.21	5.73	-2.18	1.39	-0.19	0.68	0.79	0.87
6	1.52	3.16	2.77	0.34	1.96	-0.16	2.51	3.22	1.35
7	2.2	2.42	-0.19	-1.38	0.94	0.45	0.6	2.44	0.92
8	0.91	1.94	6.21	-0.12	0.82	0.17	0.64	0.13	0.97
9	0.65	1.93	4.38	1.44	2.31	0.14	0.85	0.58	0.99
10	-0.26	0.82	-0.96	0.26	1.94	0.08	0.66	0.51	0.88

### (1) Parzen 窗非参数估计法：

利用课本公式 P135 (11)：

$$p_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{V_n} \varphi\left(\frac{x-x_i}{h_n}\right)$$

### (2) Knn 非参数估计法：

利用课本公式 P145 (32)：

$$p_n(x|w_i) = \frac{k_i/n}{V}$$

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。

# 四、实验结果

(1) Parzen 窗非参数估计法:

(a) 令  $h=1$

	1	2	3
(0.5, 1.0, 0.0) :	0.0502	0.1880	0.1588

	1	2	3
(0.31, 1.51, -0.50) :	0.0612	0.1927	0.0902

	1	2	3
(-0.3, 0.44, -0.1) :	0.0558	0.1509	0.0728

(b) 令  $h=0.1$

(0.5, 1.0, 0.0) :

1	2	3
3.5015e-22	2.7011e-07	3.1924e-19

(0.31, 1.51, -0.50) :

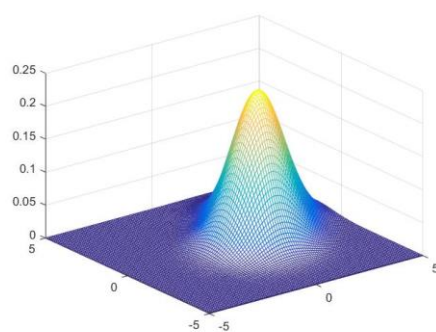
1	2	3
1.1457e-22	4.7893e-08	8.6167e-28

(-0.3, 0.44, -0.1) :

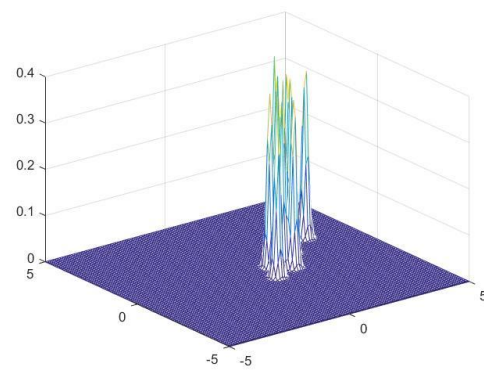
1	2	3
1.4514e-14	1.5096e-06	4.2459e-42

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。

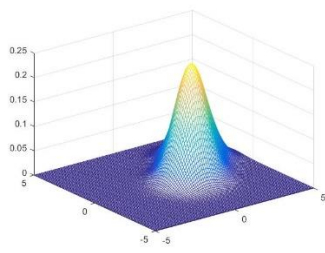


H=1

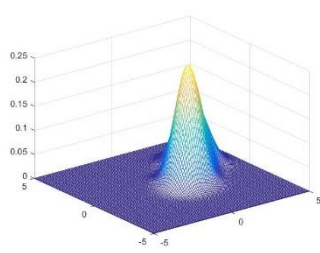


H=0.1

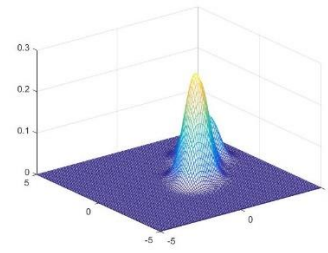
再调整  $h$  的大小，得到下列图像：



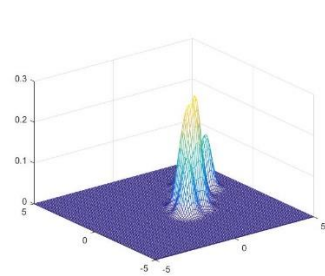
$h=0.8$



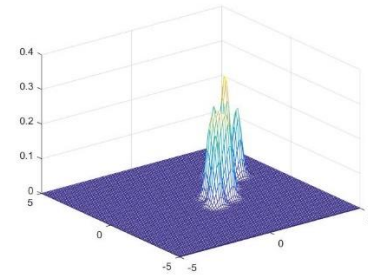
$h=0.6$



$h=0.45$



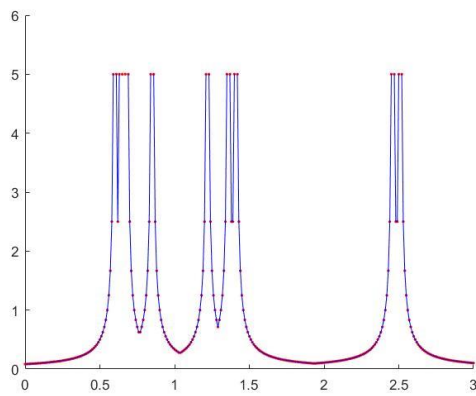
$h=0.3$



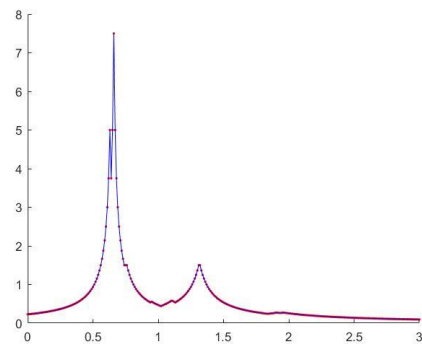
$h=0.2$

## (2) Knn 非参数估计法:

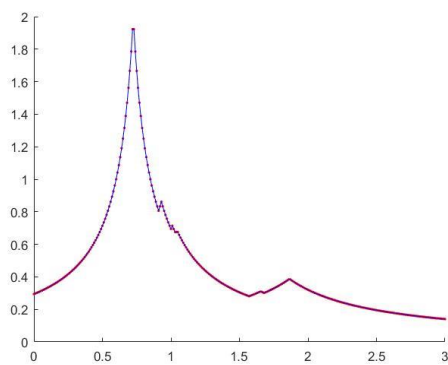
(a) 对类别  $w_3$  中的特征  $x_1$  进行概率密度估计：



$K=1$

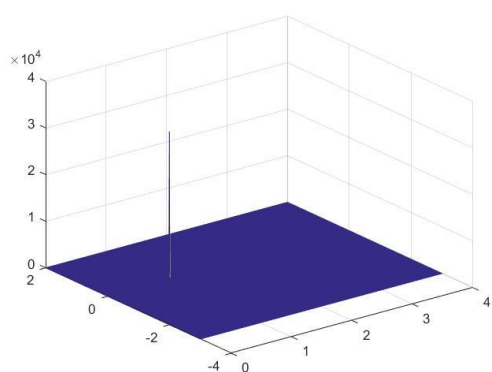


$k=3$

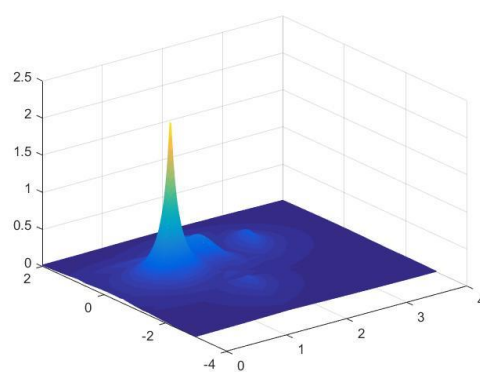


$K=5$

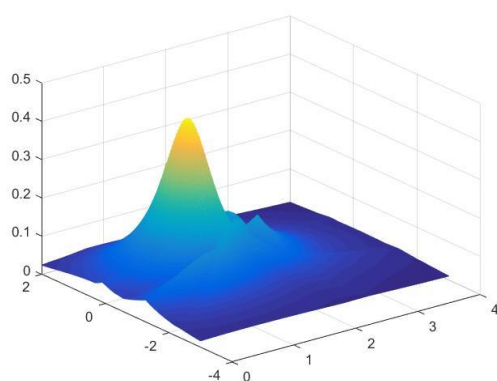
(b) 对类别 w2 中的特征 x1, x2 进行概率密度估计：



K=1



k=3



K=5

(c) 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

0.0021	0.0553	0.0358
--------	--------	--------

$x = [0.14, 0.72, 4.1]$  在三类中的概率密度分别为

0.0043	9.6478e-04	0.0021
--------	------------	--------

$x = [-0.81, 0.61, -0.38]$  在三类中的概率密度分别为

0.0026	0.0869	0.0085
--------	--------	--------

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。