PATTERN RECOGNITION AND CLASSIFICATION

WEEK 3 ASSIGNMENT ANSWERS

ARYA RAJIV CHALOLI

PES1201700253

1. Why do we need Non-Parametric techniques for pattern recognition?

   Answer:

   - In most pattern recognition applications, the assumption that all the forms of the underlying density functions are known is suspect.
   - The common parametric forms rarely fit the densities actually encountered in practice.
   - All classical parametric densities are unimodal whereas most practical problems involve multimodal densities.
   - The chances that a high-dimensional density might bring about a significant change is rare, as they are often simply represented as the product of one-dimensional functions.
   - For the above reasons, nonparametric procedures are adopted as they can be used with arbitrary distributions and without the assumption that the forms of the underlying densities are known.

2. What are the methods of density estimation?

   Answer:

   There are two common ways of obtaining sequences of regions for a density estimation:

   - The Parzen-window method:

This method intends to shrink an initial region by specifying the volume V_n as some function of n, such as $V_n = 1/\sqrt{n}$. It then must be shown that the random variables k_n and k_n/n behave properly, or more to the point, that $p_n(x)$ converges to p(x).

- The KN-Nearestneighbor method:

  The second method is to specify k_n as some function of n, such as $k_n = \sqrt{n}$. Here the volume V_n is grown until it encloses k_n neighbors of x. This is the kn-nearestneighbor estimation method.

3. What is a metric? Give the properties of a metric along with an example.

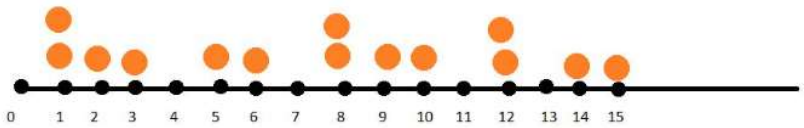
   Answer:

   - A metric D(a,b) is a function that gives a generalized scalar distance between two argument patterns.
   - A metric must have four properties for all vectors a, b and c
     - non-negativity: D(a,b) ≥ 0
     - reflexivity: D(a,b) = 0 if and only if a = b
     - symmetry: D(a,b) = D(b,a)
     - triangle inequality: D(a,b) + D(b,c) ≥ D(a,c).
   - An example of a metric is the Euclidean formula for distance in d dimensions,

$$D(a,b) = (\sum_{k=1}^{d} (a_k - b_k)^2)^{1/2}$$

4. Use PARZEN WINDOW approach and K-N NEAREST NEIGHBOR approach to solve the given density estimation and plot the graph for steps of .5. Take window size h= 2 and k=3

Density allocation (Orange dots are entities)



Answer:

	A	B	C	D
1	PARZEN WINDOW			
2	x	kn	h	d=kn/h
3	0	2	2	1
4	1	3	2	1.5
5	2	4	2	2
6	3	1	2	0.5
7	4	2	2	1
8	5	2	2	1
9	6	2	2	1
10	7	3	2	1.5
11	8	3	2	1.5
12	9	4	2	2
13	10	2	2	1
14	11	3	2	1.5
15	12	2	2	1
16	13	3	2	1.5
17	14	2	2	1
18	15	2	2	1