SNS COLLEGE OF ENGINEERING

Kurumbapalayam (Po), Coimbatore – 641 107

An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A' Grade
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING-IoT Including CS & BCT

COURSE NAME :19SB701 PATTERN RECOGNITION TECHNIQUES IN CYBER CRIME

IV YEAR / VII SEMESTER

Unit III- **NONPARAMETRIC TECHNIQUE AND NON-METRIC METHODS**

Topic :Density Estimation - Parzen Windows

Parzen Windows (also known as the Parzen-Rosenblatt window method) is a non-parametric technique used for density estimation.

It provides a way to estimate the probability density function (PDF) of a random variable based on a given sample.

Unlike parametric methods, which assume a specific form for the underlying distribution, Parzen Windows makes no such assumption, making it flexible but also computationally intensive, especially for large datasets.

Key Concepts in Parzen Windows Density Estimation:

Non-Parametric Density Estimation:

Parzen Windows is a non-parametric method, meaning it does not assume a specific parametric form (e.g., normal distribution) for the data's underlying distribution.

The PDF is estimated directly from the data using a kernel function that places "windows" around each data point.

Kernel Function:

A kernel function $K(u)$K(u) is used to place a window around each data point.

The choice of kernel function affects the smoothness and shape of the estimated density.

Commonly used kernel functions include

- Uniform Kernel: $K(u) = \frac{1}{2}$ for $|u| \leq 1$, and 0 otherwise.
- Gaussian Kernel: $K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$.

The kernel is centered at each data point, and its width is controlled by a parameter called the bandwidth hh.

Bandwidth Parameter (hh):

The bandwidth hh determines the width of the kernel function.

A small hh results in a very "spiky" density estimate (high variance, low bias), whereas a large hh produces a smoother estimate (low variance, high bias).

Choosing an appropriate hh is critical and often involves cross-validation or other selection methods.

Density Estimation Formula:

Given a sample of nn data points $\{x1,x2,...,xn\}${x 1 ,x 2 ,...,x n }, the estimated density f^(x)f^ (x) at a point xx using Parzen Windows is:

where dd is the

$$\hat{f}(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

Each data point contributes to the density estimate at xx based on its distance from xx and the chosen kernel function.

Dimensionality:

In multi-dimensional cases, the kernel function and bandwidth need to be extended to account for each dimension.

The complexity of the method increases significantly with dimensionality, often referred to as the "curse of dimensionality."

Advantages and Disadvantages:

Advantages:

No assumption about the underlying distribution.

Can model complex distributions and multimodal distributions.

Disadvantages:

Computationally expensive, especially for large datasets.

Sensitive to the choice of bandwidth hh.

Example:

Suppose we have a one-dimensional dataset with n=5n=5 points: {2,3,4,5,6}{2,3,4,5,6}.

To estimate the density at a point xx using a uniform kernel with bandwidth h=1h=1, the Parzen window estimate might look like this:

$$\hat{f}(x) = \frac{1}{5 \times 1} \sum_{i=1}^{5} K\left(\frac{x - x_i}{1}\right)$$

where the uniform kernel $K(u)$K(u) assigns equal weight to points within a distance of 1 from xx.

If x=4x=4, the kernel will be non-zero for points at 3, 4, and 5, contributing to the density estimate.

This estimate gives an idea of the "concentration" of data points around x=4x=4.

MCQ

1.What type of method is Parzen Windows in the context of density estimation?

A) Parametric

B) Non-Parametric

C) Semi-Parametric

D) Deterministic

Answer: B

2.In the Parzen Windows method, what is the role of the bandwidth parameter *h*h?

A) It determines the number of data points used in the estimation.

B) It controls the width of the kernel function and affects the smoothness of the density estimate.

C) It specifies the dimensionality of the data.

D) It decides the shape of the data distribution.

Answer: B

3. Which of the following is a commonly used kernel function in Parzen Windows?

A) Exponential Kernel

B) Sigmoid Kernel

C) Uniform Kernel

D) Poisson Kernel

Answer: C

Any Query????

Thank you……