INDIAN INSTITUTE OF INFORMATION TECHNOLOGY ALLAHABAD

Subject – Data Mining and Warehousing

Topic – A One-Class Classification Decision Tree Based on Kernel Density Estimation

Report By -

IIT2018108 – Ravi Kumar Sharma

Introduction :

OCC is of major concern in several domains where it may be expensive

and/or technically difficult to collect data on a range of behaviors or phenomenons. For example, it may be quite affordable to gather data on the

representatives of a given pathology in medicine, or positive operating scenarios of machines in the industry. The related complementary occurrences are, by contrast, scarce and/or expensive to raise.

One-Class Support Vector Machine (OCSVM) and Support Vector Data Description (SVDD) are among the most common OCC methods. OCSVM

aims at finding the hyper-plane that separates the target instances from the origin with the wider margin, while SVDD aims at enclosing these instances within a single hyper-sphere of minimal volume.

Result :

ClusterSVDD	**OC-Tree**
• Detects target hyper-sphere(s). • Requires to set the number of hyper-sphere(s) as a parameter. • Relies on two parameters: k, ν_{SVDD}. • Results in a classification model whose predictions are based on the whole set of training attributes.	• Detects target hyper-rectangle(s). • Does not require indications about the number of hyper-rectangle(s) to detect. • Relies on four parameters: $\gamma, \beta, \alpha, \nu$. • Results in a classification model whose predictions are based on a subset of training attributes.

Table 1: Comparison of ClusterSVDD & OC-Tree

	# Classes	# Features	# Instances
Australian	2	14	690
Diabetes	2	8	268
Ionosphere	2	34	351
Iris	3	4	150
Satimage	6	36	4435
Segment	7	19	2310

Table 2: Benchmark datasets [33, 34]

NOTE : For detail wise result see the main .ipynb file.