

CLUSTER BASED

Un oggetto è considerato un outlier se:

- 1) Non appartiene a nessun cluster
- 2) δ lontano rispetto al cluster più vicino
- 3) Appartiene a un cluster piccolo o sperso

CBLOF: Cluster-Based LOF

Questo fattore tiene conto della grandezza del cluster a cui appartiene l'oggetto analizzato, sia la distanza fra l'oggetto e il cluster più vicino.

Due parametri α e β :

$$\begin{cases} \sum_{i=1}^b |C_i| \geq |D| \cdot \alpha \\ \frac{|C_b|}{|C_{b+1}|} \geq \beta \end{cases}$$

Dove la prima equazione ci dà modo di indicare il numero di punti che non sono outlier e quindi appartengono a un cluster (es: $\alpha = 90\%$ punti non outlier)
la seconda ci aiuta a definire la differenza di dimensione fra cluster piccoli e grandi (SC e LC)

Large Cluster: $LC = \{C_i \mid i \leq b\}$

Small Cluster: $SC = \{C_j \mid j > b\}$

Dove b definisce il confine fra cluster grandi e piccoli

Definitions CBLWF(o)

$$CBLWF(o) = \begin{cases} |C_i| \cdot \min(\text{dist}(o, C_i)) & o \in C_i, C_i \in SC, C_i \in LC \\ |C_i| \cdot \text{dist}(o, C_i) & o \in C_i, C_i \in LC \end{cases}$$

