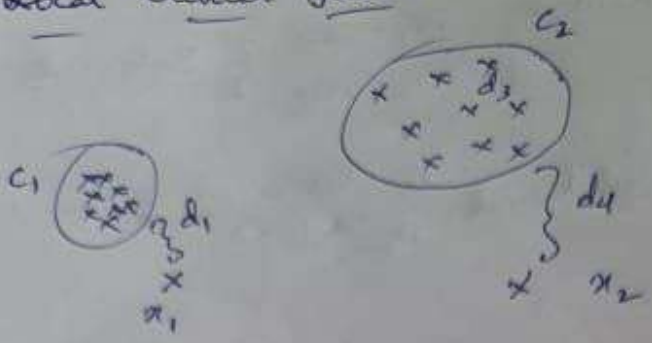
Local Outlier Factor

C_1, C_2 → Clusters

x_1, x_2 → Outliers



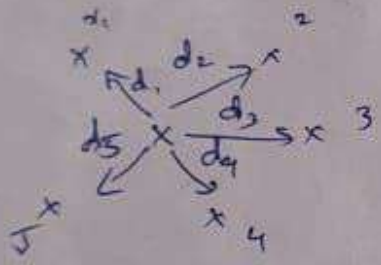
x_i → taken mean distance from nearest neighbour

d_1 and d_4 is very large, so consider as outlier

If $d_3 > d_1$ then C_2 is also considered as outlier

So the idea is to find Local density

① k-distance (x_i) → distance from k^{th} nearest neighbour of x_i from x_i



② Reachability distance (x_i, x_j)

↳ $\max$ (k distance (x_j) , dist (x_i, x_j))

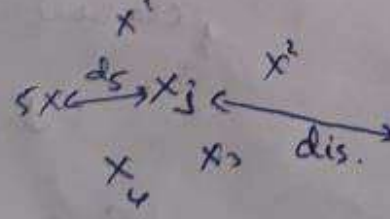
$k = 5$

if $x_j \in N(x_j)$

return k distance (x_j)

else

return dist (x_i, x_j)



③ Local Reachability density $lrd(x_i)$

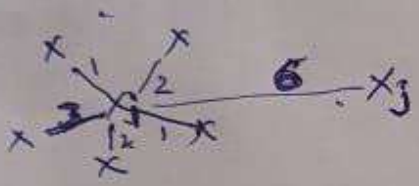
$$= \frac{1}{\sum_{x_j \in N(x_i)} \left\{ \frac{reach\_dist(x_i, x_j)}{|N(x_i)|} \right\}}$$

$N(x_i)$

(4) Local Outlier factor (x_i)

$$= \frac{\sum_{x_j \in N(x_i)} lrd(x_i)}{|N(x_j)|} \times \frac{1}{lrd(x_i)}$$

if LOF(x) is large then it is outlier



Reachability dist $(x_i, x_j) = 6$

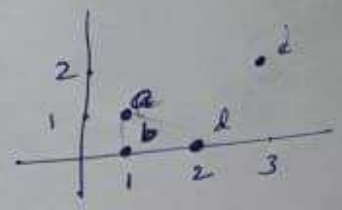
$$LRD(x_i) = \frac{1}{\sum_{x_j \in N(x_i)} \left\{ \frac{\text{reach-dist}(x_i, x_j)}{|N(x_i)|} \right\}}$$

$$= \frac{1}{\frac{1+2+2+1+3}{5}} = \frac{5}{9} \approx 0.5$$

$$LOF = \frac{\sum_{x_j \in N(x_i)} lrd(x_i)}{|N(x_i)|} \times \frac{1}{}$$

Ex :- a(1, 1), b(0, 0) c(3, 2) d(2, 0)

(1, 0)
(3, 2)
(2, 0)
(1, 1)
(3, 1)
(3, 0)



dis (a, b) = 1 dis (a, c) = 3 dis (a, d) = 2
dis (b, c) = 4 dis (b, d) = 1
dis (c, d) = 3

N (a) = (b, d) N (b) = (a, d) N (c) = (a, d)
N (d) = (a, b)

① 2 distance (a) = d = 2
2 distance (b) = add = 1
2 distance (c) = add = 3
2 distance (d) = a = 2

② rd (a, b) = max (1, 1) = 1.
↳ max (2distance (b), dis (a, b))

rd (a, c) = max (3, 3) = 3
rd (a, d) = max (2, 2) = 2
rd (b, c) = max (3, 4) = 4
rd (b, d) = max (2, 1) = 2
rd (c, d) = max (2, 3) = 3.

③ LRD =
$$LRD(a) = \frac{1}{\frac{\sum_{x_j \in N(a)} reach\_dist(a, x_j)}{n}}$$

$$LRD(a) = \frac{1}{\frac{rd(a,b) + rd(a,d)}{2}} = \frac{2}{1+2} = \frac{2}{3} = 0.66$$

$$LRD(b) = \frac{2}{rd(b,a) + rd(b,d)} = \frac{2}{1+2} = \frac{2}{3} = 0.66$$

$$LRD(c) = \frac{2}{rd(c,a) + rd(c,d)} = \frac{2}{3+3} = \frac{1}{3} = 0.33$$

$$LRD(d) = \frac{2}{rd(d,a) + rd(d,b)} = \frac{2}{2+2} = \frac{1}{2} = 0.5$$

④ $$LOF(a) = \frac{\sum_{x_j \in N(a)} lrd(x_j)}{|N(x_j)|} \times \frac{1}{lrd(x_i)}$$

$$= \frac{lrd(b) + lrd(d)}{2} \times \frac{3}{2}$$

$$= \frac{\frac{2}{3} + \frac{1}{2}}{2} \times \frac{3}{2} \Rightarrow \frac{\frac{7}{6 \times 2}}{2} \times \frac{3}{2}$$

$$\Rightarrow \frac{7}{8} = 0.875$$

$\Rightarrow$

$LOF(b) = 1.33$ $LOF(c) = 2$ $LOF(d) = 0.875$