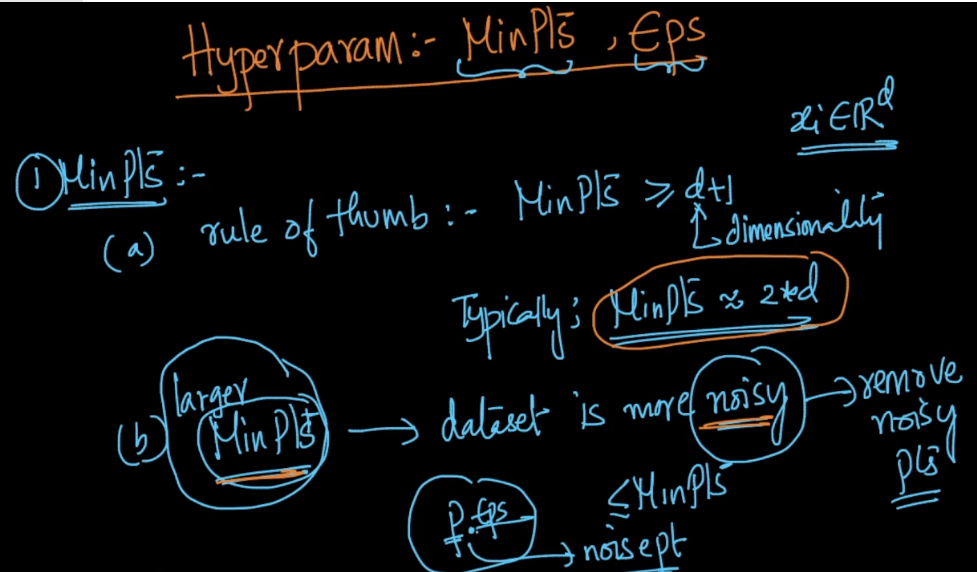
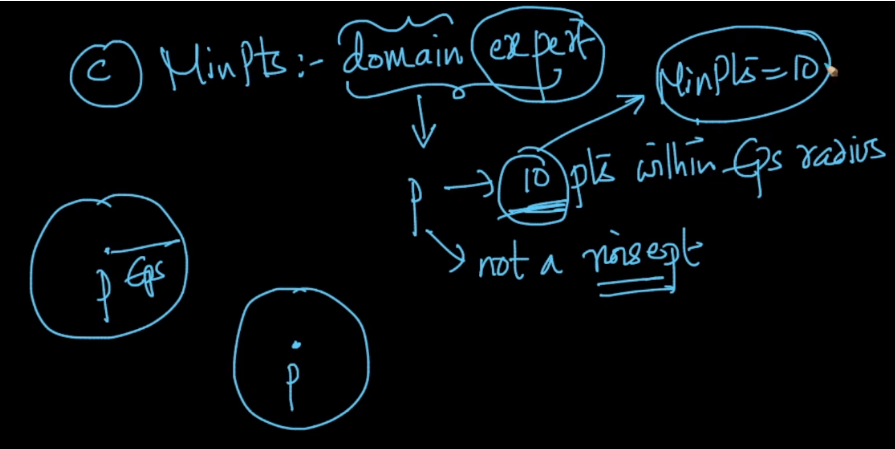
How to select hyperparameter (Min pts, epsilon) in DBSCAN

1. Min pts:
2. Rule of thumb is min pts >= d+1, where d is the dimensionality or no. of features.
3. Choose large value for min pts if dataset is more noisy, because larger the values of min pts, the large no. of noise pts will be detected and removed.
4. From Domain Expert/knowledge: Suppose domain expert says that 10 points within eps are not noisy, then we choose 10 as min pts.





1. Calculating EPS:

Let’s say we obtain min pts as 4, so we calculate EPS as:

1. Find distance from point to the 4th NN of that point.
2. Sort all the distances in increasing order.
3. Plot a graph where on x-axis there are points index and on y axis there is distances in increasing order.

Now at certain point you will see that after which there is drastic increase in distances, so we’ll pick that point’s distance as epsilon.

This drastic change in distance means that next point is very far from current point, and there is high probability that after that all the points are noisy.

