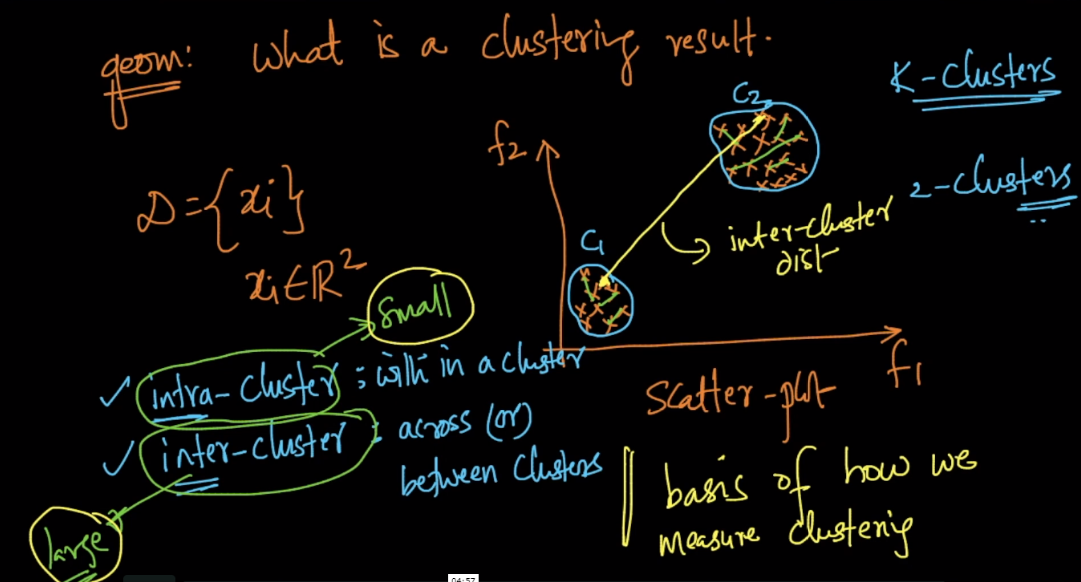


As we’ve seen for classification and regression there are several error metrics, now question what is metrics for clustering.

Let’s know some terms:

**Intra-cluster:** distance of points within clusters, and it should be small.

**Inter-cluster:** distance of points across cluster or between clusters, and it should be large.



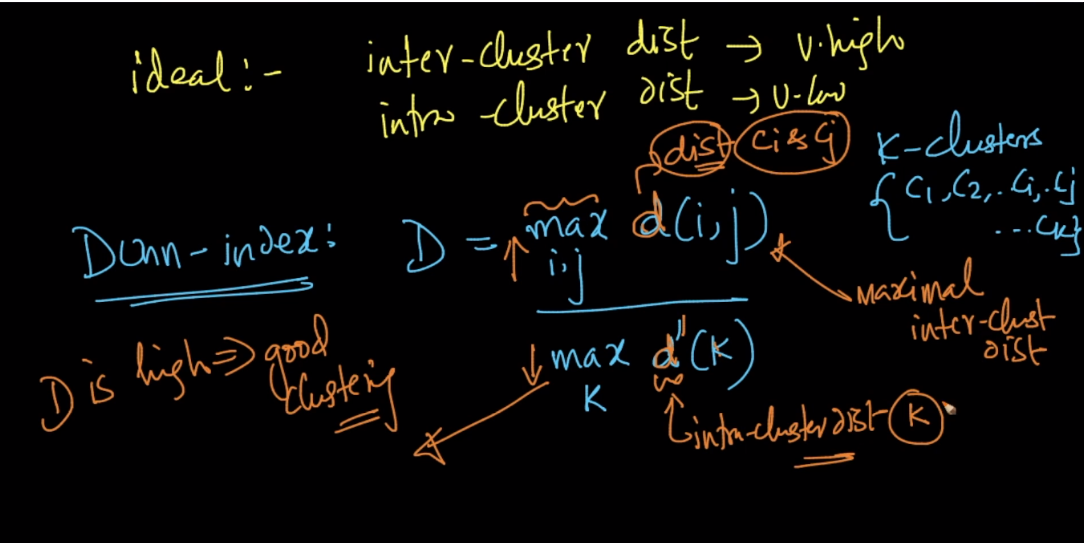
There is one metric called dunn-index which is good if it’s high, it’s formula is:

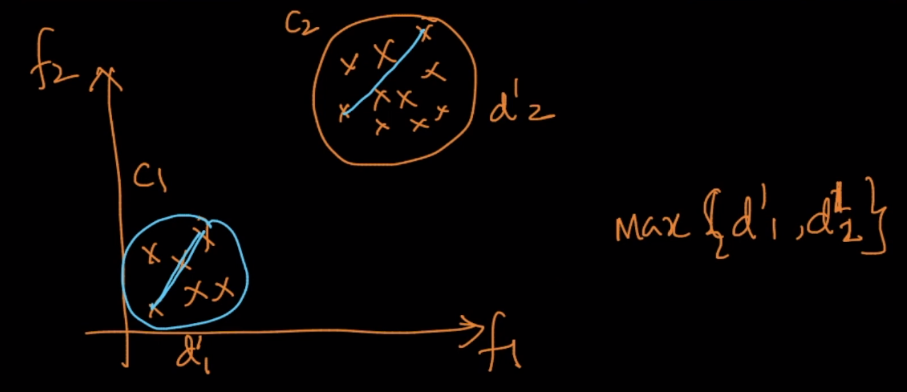
Dunn-index(D) = min d(I, j) / max d’(k)

Here d(I, j) is the min distance between clusters: it’s calculated as from every point of cluster c1, calculate distance with every point of cluster c2. Pick the minimum distance.

d’(k) is the max distance between the points within a cluster: it’s calculated as find distance of every point with every point within that cluster. Pick the max distance from whichever the cluster is maximum. Ex suppose from c1 obtained distance is 4, and from c2 obtained distance is 6, then we pick 6.

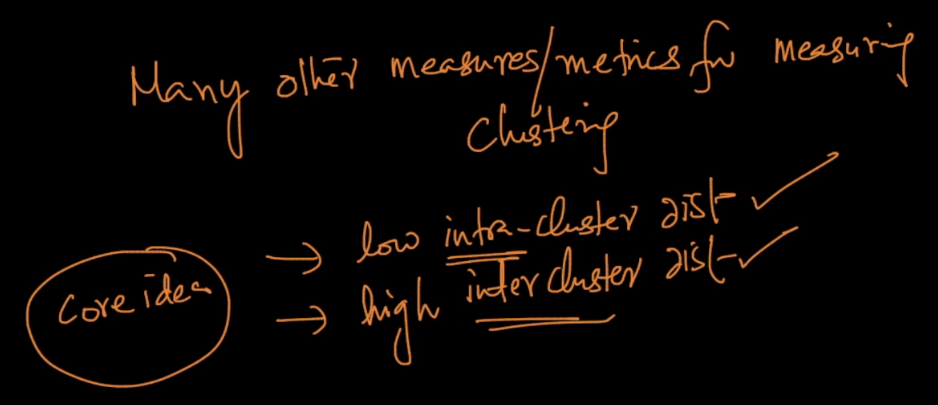






There are several other metrics also for measuring clustering, but the core-idea for all metrics is similar that:

* There should be low intra-cluster distance.
* There should be high inter-cluster distance.



**Why we choose min distance between clusters for numerator:**

