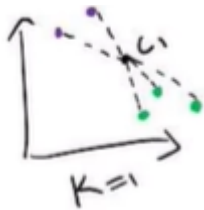


## WSS Plot

There are total of 5 observations.



For  $K=1$ ,  $C_1$  is the centroid. And WSS is calculated when  $K=1$  as shown in the graph for x-axis which is different  $K$  values that the k-means model can take and Y-axis is the total WSS (within sum of square) variations for each  $k$ -value.

So let us assume we are computing this value in this plot and marking it for  $K=1$  (say at top =10.0)

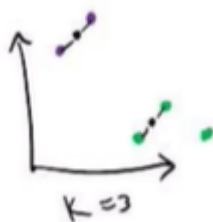
Now going to the second chart when we run the k-means model with  $K=2$



Here we have 2 centroids, now if we compute WSS for cluster 1 and cluster 2 and add them up, they are going to be significantly lesser than the WSS that we obtained from the previous cluster that is  $K=1$ .

So if we plot this in graph, we notice that from 1 when we move to 2 the total WSS has significantly dropped.

Let us go on the next chart.



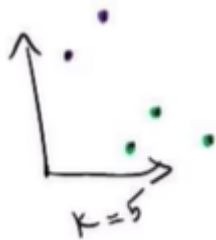
Here we have mentioned  $K=3$ , so basically we have 3 clusters here and only 1 observation in the 3<sup>rd</sup> cluster. The WSS for the 3<sup>rd</sup> cluster will be equal to zero because the single observation is the centroid itself. So if we measure the distance it is going to be zero so there is no WSS value here. So when we compute the WSS for other 2 clusters there will be a drop again in the WSS, but it is not going to be as significant as the previous drop that happened. When we moved from 1 to 2 there was a very good drop, but from 2 to 3 the drop is less.

Similarly, for the next chart



When  $K=4$  in k-means model, we now have 4 clusters namely cluster 1, cluster 2, cluster 3 and cluster 4. So again, cluster 1, 2 and 4 will have zero WSS. Only cluster 3 will have some amount of WSS which is not much different from  $K=3$  WSS value.

Finally, when we come to the 5<sup>th</sup> plot



And we measure the WSS value for  $K=5$ , so we have 5 clusters which are nothing but 5 observations. Here in this case total WSS is Zero. And Hence that is shown on the plot.

After joining all the lines in the scatter plot, we get the below graph.

From 1 to 2, there is a significant drop hence 2 is a valuable addition in k-means algorithm.

