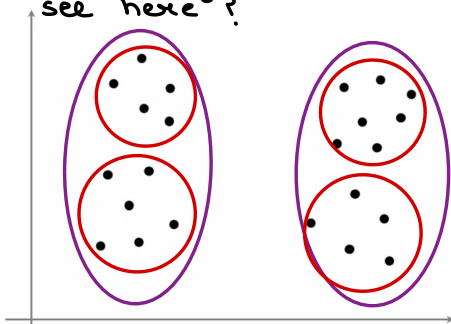


The bigger question we're targeting here is what is the right value of K ?

it is truly ambiguous ← there is none!

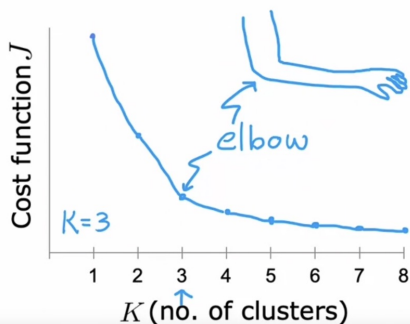
eg. How many clusters do you see here?



Some will say 2.
Some will say 4.

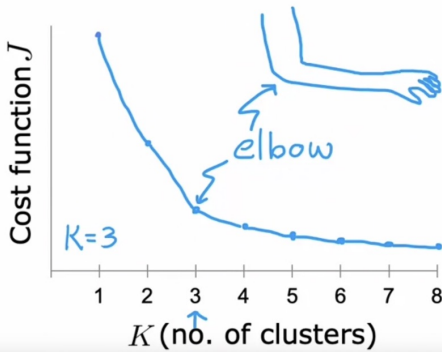
There are some methods to find the right amount of K (no. of clusters)

one such method is the elbow method.



We choose $K=3$ because at $K=3$ we notice a steep decrease in the cost function and we don't see a substantial decrease in the cost function (choose the point where you see heavy decrease)

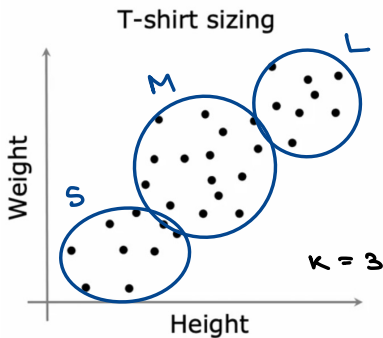
Don't choose K based on what value minimizes J . Always more the no. of clusters, more will be the minimization, but we don't want that in most cases.



8 minimizes the cost function the most.

choosing the value of K

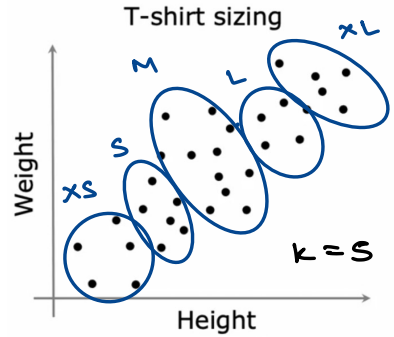
Often you want to choose K based on your use case. You might want to get clusters for some later purpose, for eg.



You may want 3 t-shirt sizes S, M and L for your business.

OR

You may want 5 T-shirt sizes for your business. Manufacturing 5 different t-shirt sizes can lead to extra costs but is often useful.



Basically you would want to evaluate cost for manufacturing 3 t-shirts or 5 t-shirts and examine the tradeoff between cost and diversity. Then you may choose the k according to your needs.

In short, value of k is always an ambiguity and there is no proper algorithm for choosing its value. You have to examine the tradeoffs of choosing different values of k and then make a decision accordingly.