Clustering is the grouping of similar things. There are various clustering methods, for example-

<br>

![image](<https://raw.githubusercontent.com/niteshjindal170988/unsupervised-learning/main/clustering/.scrap/clus0.JPG>)

<br>

Let us say that we have n different data points and K different clusters, such as . For every combination of data point and a cluster, we can compute a distance between them, for example the distance between and or distance between and and so on.

Minimizing the distance between nth data point and kth cluster can be written as:

<br>

However, every data point is not associated with every cluster. Thus, we need to associate the nth data point with the cluster it belongs to, not all the clusters. We will define a latent variable that tells us whether the nth datapoint is associated with cluster K or not. either 0 or 1. If it is 1, then the distance between the nth datapoint and kth cluster matters else not. <br>

Therefore, we can write the objective function as:

|  |  |  |
| --- | --- | --- |
| Params | Term | Definition |
| K | Hyperparameter | Number of Clusters |
|  | Parameters | Cluster Mean Vectors. |
|  | Total Number of Parameters | Each is a D-dimensional vector and hence its matrix size is . Thus, the learned parameters are |
|  | Latent Parameter | that have value of 1 when there is an association between nth datapoint and a particular cluster. For all other combinations, the value would be 0. Latent parameter are similar to parameters, but we need them so as to write the objective function in a simple way. |

Minimum number of clusters can be one (i.e. all datapoints in a single cluster) and maximum number of clusters can be equal to the number of datapoints in the data.

Hyperparameter tell us the complexity of the modelling system.

What makes a latent parameter?

We learned that the size of is and size of is . If we know , we can compute because we know that each datapoint is associated with a particular cluster only, and thus we can obtain the all the datapoints belonging to a particular cluster and take their mean to get cluster mean vector or cluster centers. So, the latent parameters are equivalent to parameters ( )

We can alternate between the and ( i.e. is given and known to us and we compute , or is given and we compute ). In the following lines, we will consider as constant and update the -

<br><br>Let us get started!<br><br>

$J(m, \lambda)=\sum\_{n=1}^{N} \sum\_{k=1}^{k} \delta\_{n, k}\left(\bar{m}\_{k}-\bar{x}\_{n}\right)^{2}$

<br>Keep as constant and update the , then take the partial derivative w.r.t ( jth cluster):

$\frac{\partial J}{\partial m\_{j}}=2 \sum\_{n=1}^{N} \delta\_{n\_{1}}\left(\bar{m}\_{j}-\bar{x}\_{n}\right)$

Clustering can be used in following ways:

* Understand the structure in data
* Generate class labels when unknown
* Summarize data points by their cluster center, or mean cluster vector.