

[◀ Back to Week 4](#)
[✕ Lessons](#)
[Prev](#)
[Next](#)

## A worked-out example for EM

When it comes to k-means, the notion of clusters is straightforward -- each point belongs to the cluster with the nearest centroid (mean). If someone gave us some data points and centroids, we can readily label each point for cluster.

The notion of clusters for EM is not as immediately intuitive. Before jumping into the minutiae of EM, it is best to develop your intuition as to what clusters really mean.

### Prelude: cluster assignment in k-means

Let us consider a toy example with three data points in 2D:

Dataset	X	Y
Data point 0	10	5
Data point 1	2	1
Data point 2	3	7

If we were running k-means, we would first take an initial set of centroids and then label all points with the closest ones. Let's take the following centroids:

Centroids	X	Y
Cluster A	3	4
Cluster B	6	3
Cluster C	4	6

(Since this is a toy example, let's ignore the fact that having three clusters over three data points is an overkill. But I digress.) Using the pairwise Euclidean distances, we label the data points with nearest clusters:

Pairwise distances	Centroid of Cluster A	Centroid of Cluster B	Centroid of Cluster C	Cluster Assignment