**Implementing K-Means**

Description

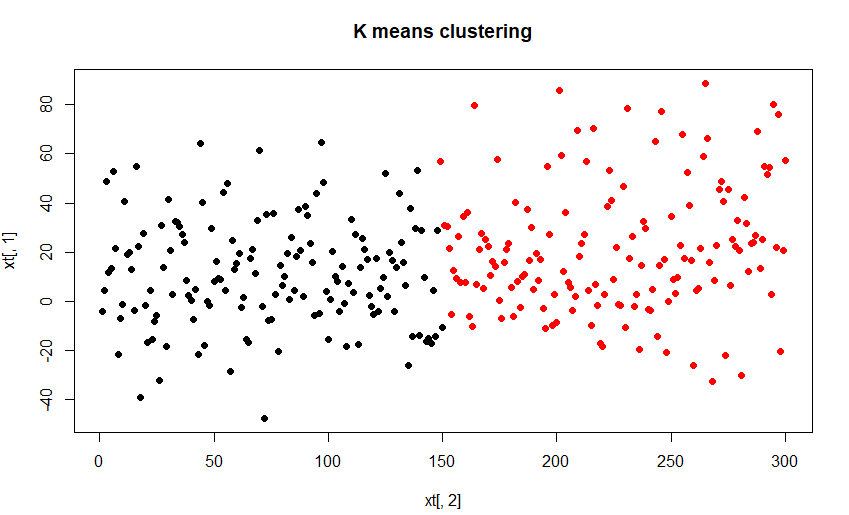
In this implementation, I am implementing the K-Means clustering in R programming. I am generating a univariate sample having samples from three gaussian distributions.

Solution

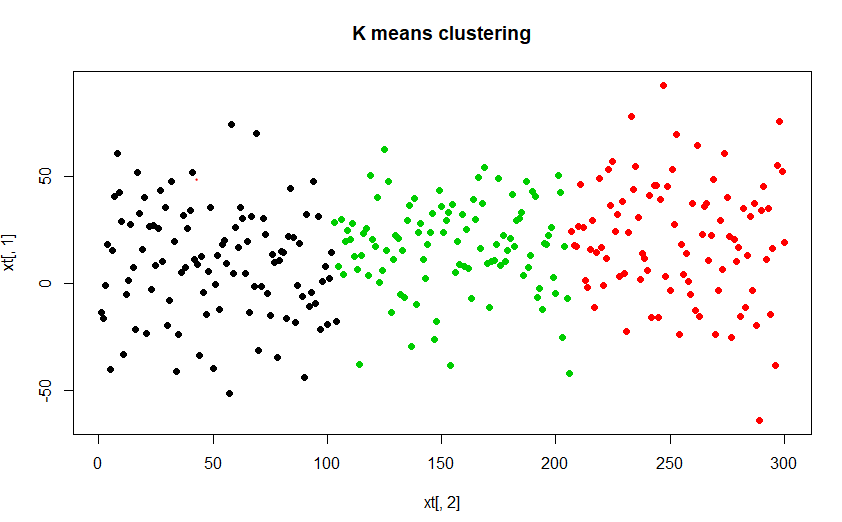
K-means clustering is a clustering algorithm which aims at clustering a sample into K partitions. I am using Euclidean distance between the points as a measure to cluster the samples. I ran the algorithm with K being 2, 3 and 4 and plotted the results. I also plotted the reconstruction error for different values of K.

Result

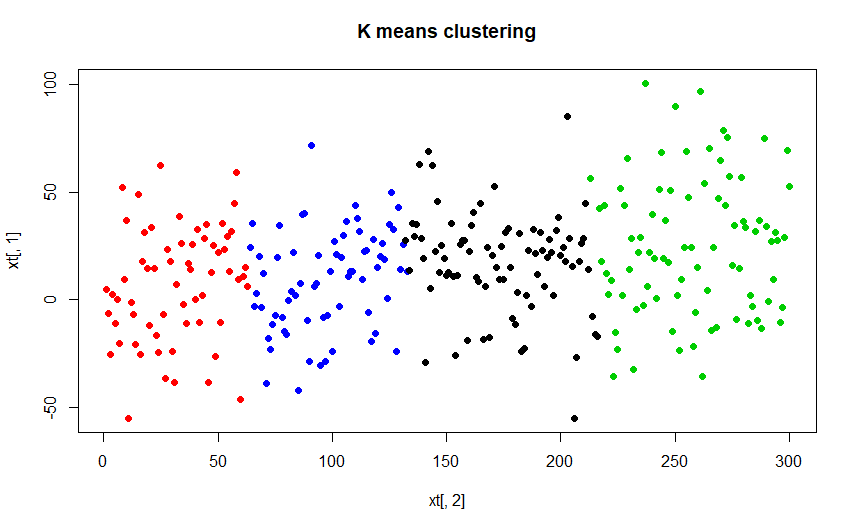
For K = 2



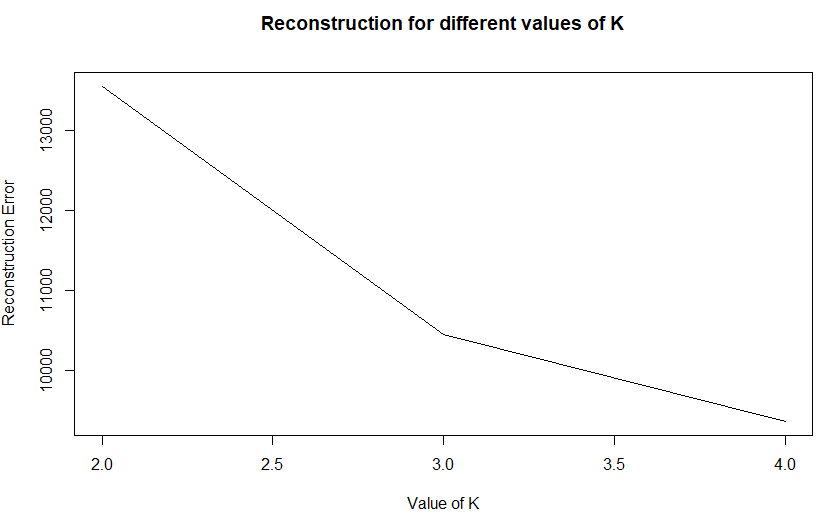
For K = 3



For K = 4



Reconstruction Error



From the reconstruction error we can see that there is a huge drop in reconstruction error when we move from k=2 to k=3, but the drop reduces from k=3 to k=4. We can see an elbow at k=3 and can pick k=3 as the number of centres in our clustering algorithm.