|  |  |  |
| --- | --- | --- |
| **2 K-means clustering | Machine Learning for BiostatisticsK-Means** | Unsupervised  Clustering | No  Non-parametric |

from sklearn.cluster import KMeans

kmeans = KMeans(n\_clusters = 3).fit(X)

K-means is a centroid-based algorithm or a distance-based algorithm, where we calculate the distances to assign a point to a cluster. In K-Means, each cluster is associated with a centroid.

Clustering of Kmeans

A diagram of blue and orange dots

Description automatically generated

Variance to determine most efficient biomarkers

Cluster 1:

13947 - RPL41

13981 - RPS18

13956 - RPLP1

16712 - TMSB4X

13941 - RPL37A

16709 - TMSB10

4964 - EEF1A1

13995 - RPS29

194 - ACTB

13953 - RPL8

Cluster 2:

13947 - RPL41

13956 - RPLP1

13981 - RPS18

13941 - RPL37A

4964 - EEF1A1

16712 - TMSB4X

13995 - RPS29

13953 - RPL8

13957 - RPLP2

6191 - FTH1