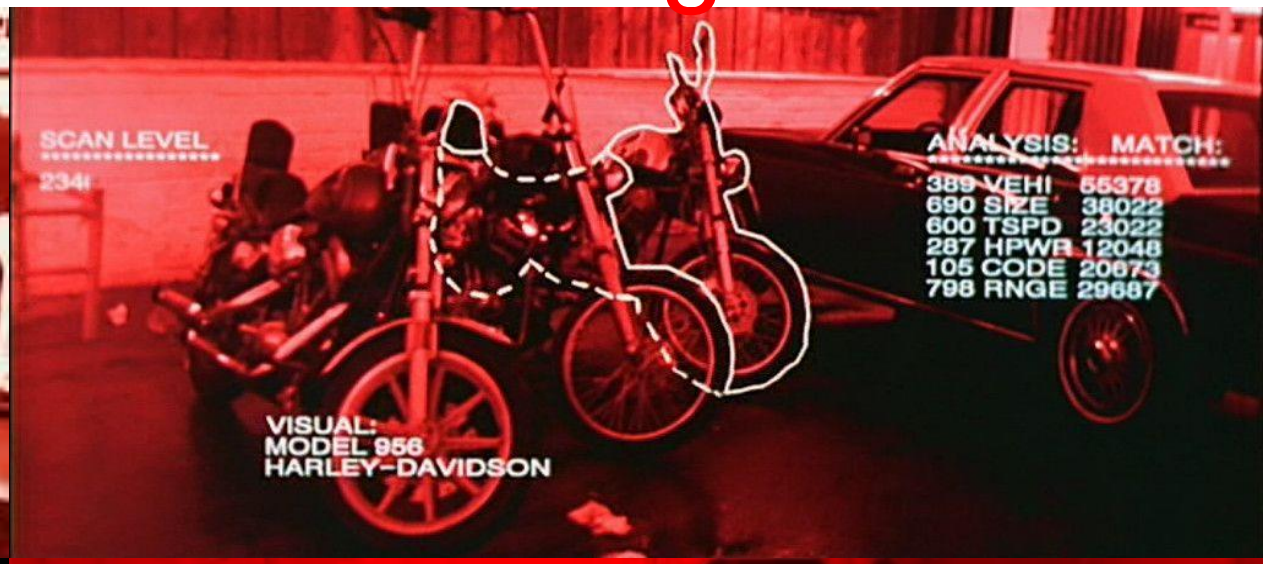
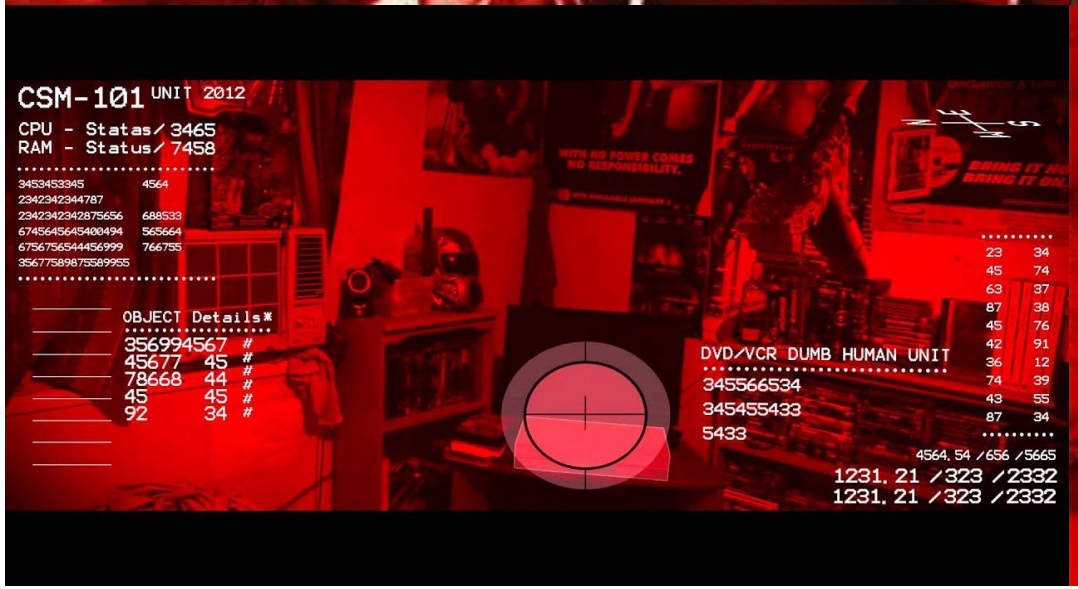
UnSupervised Learning





Un-Supervised Machine Learning

- We put unlabeled data and let machine understand the characteristics and classify it

Clustering:

Used for exploratory data analysis to find hidden patterns or grouping in data

- E,g: K-means Algorithm

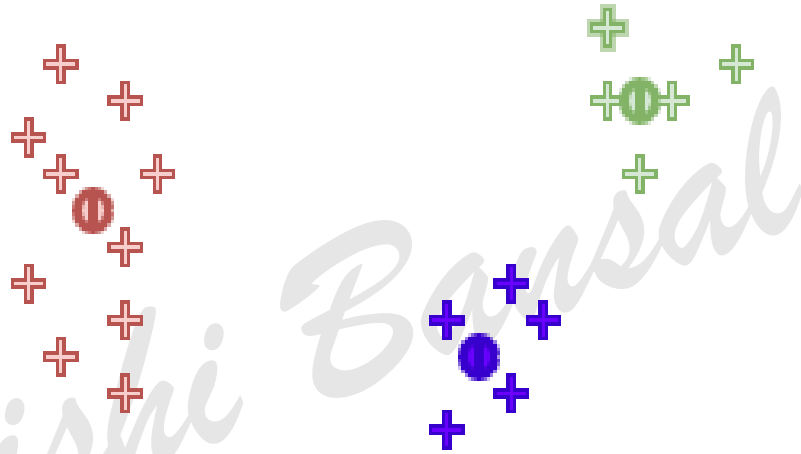
K-Means

Algorithm:

1. Initialize k centroids.
2. Select at random K points, the centroids(not necessary from the dataset)
3. Assign each data to the nearest centroid, this step will create clusters.
4. Compute and place the new centroid of each cluster.
5. Reassign each data point to the new closest centroid. If any new reassignment, Repeat steps 4 otherwise go to Finish

Animated Implementation of the Algorithm - http://tech.nitoyon.com/en/blog/2013/11/07/k-means/

Random Initialization Trap

- 

Random Initialization Trap

- 

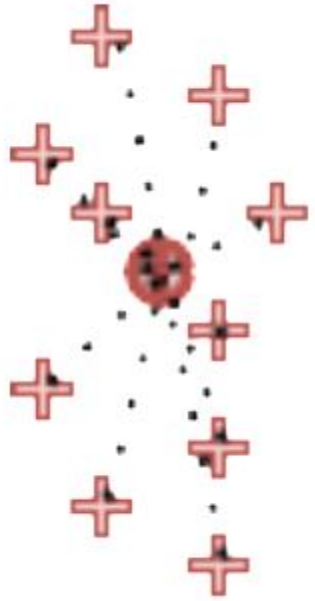
Random Initialization Trap

- Solution -> K-Means++

- K-Means++ -> smarter initialization of centroids, rest is same

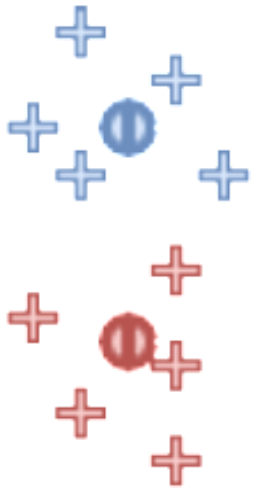
Choosing Right Number of Clusters

- WCSS: Within-Cluster Sum of Square
- Euclidean distance between a given point and centroid to which it is assigned.
- Iterate this process for all the points in the cluster
- Sum all the values and divide by no. of points



Choosing Right Number of Clusters

- Total WCSS decreases as no. of clusters increases
- Total WCSS is minimum when No. of clusters is equal to no. of data points



Elbow Method

- Elbow Method to find the optimal number of clusters