



## K-Means++ Initialization Algorithm:

Step 1: Choose first centroid at random among data points

Step 2: For each of the remaining data points compute the distance ( $D$ ) to the nearest out of already selected centroids

Step 3: Choose next centroid among remaining data points using weighted random selection – weighted by  $D^2$

Step 4: Repeat Steps 2 and 3 until all  $k$  centroids have been selected

Step 5: Proceed with standard k-means clustering

