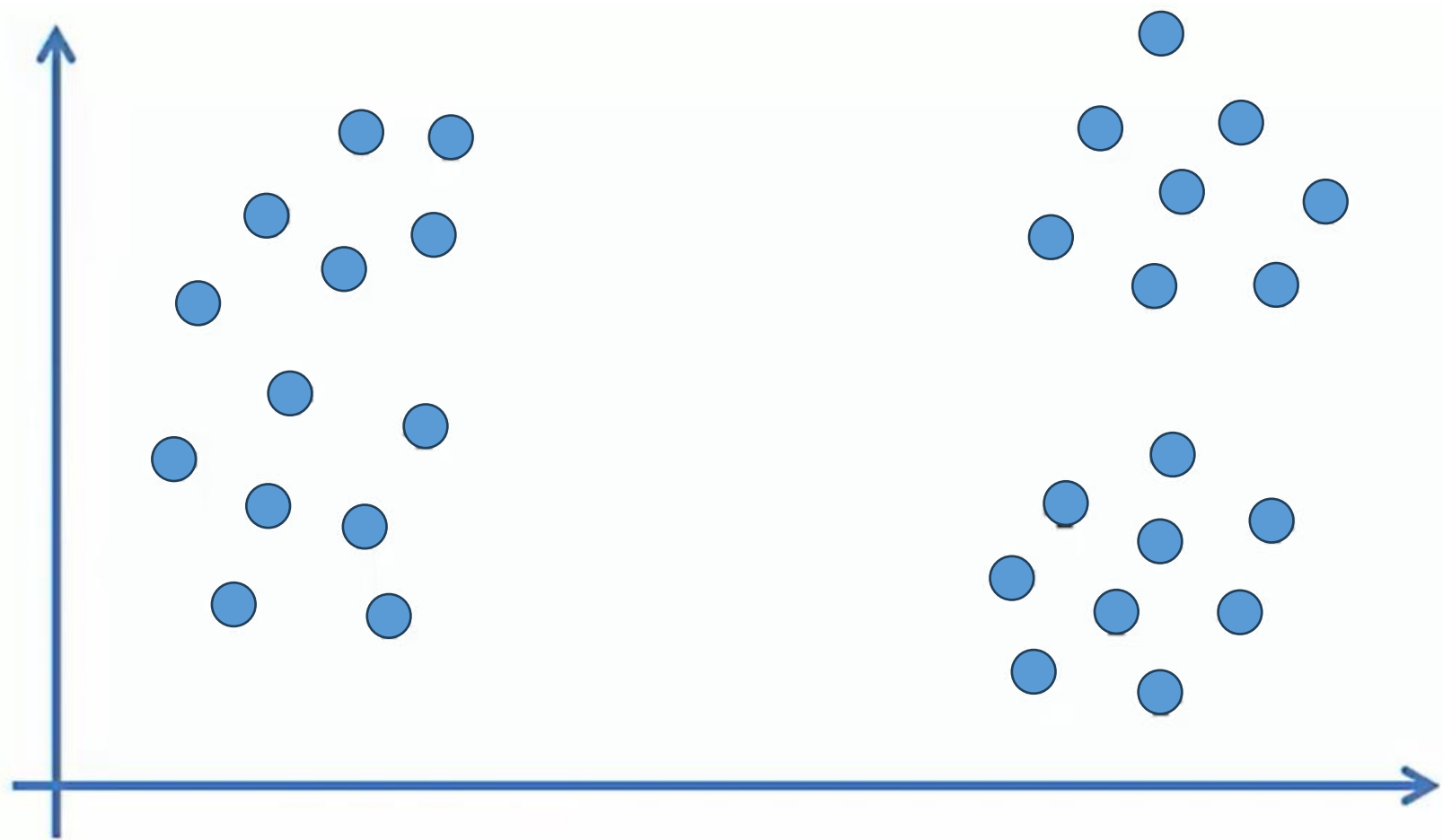


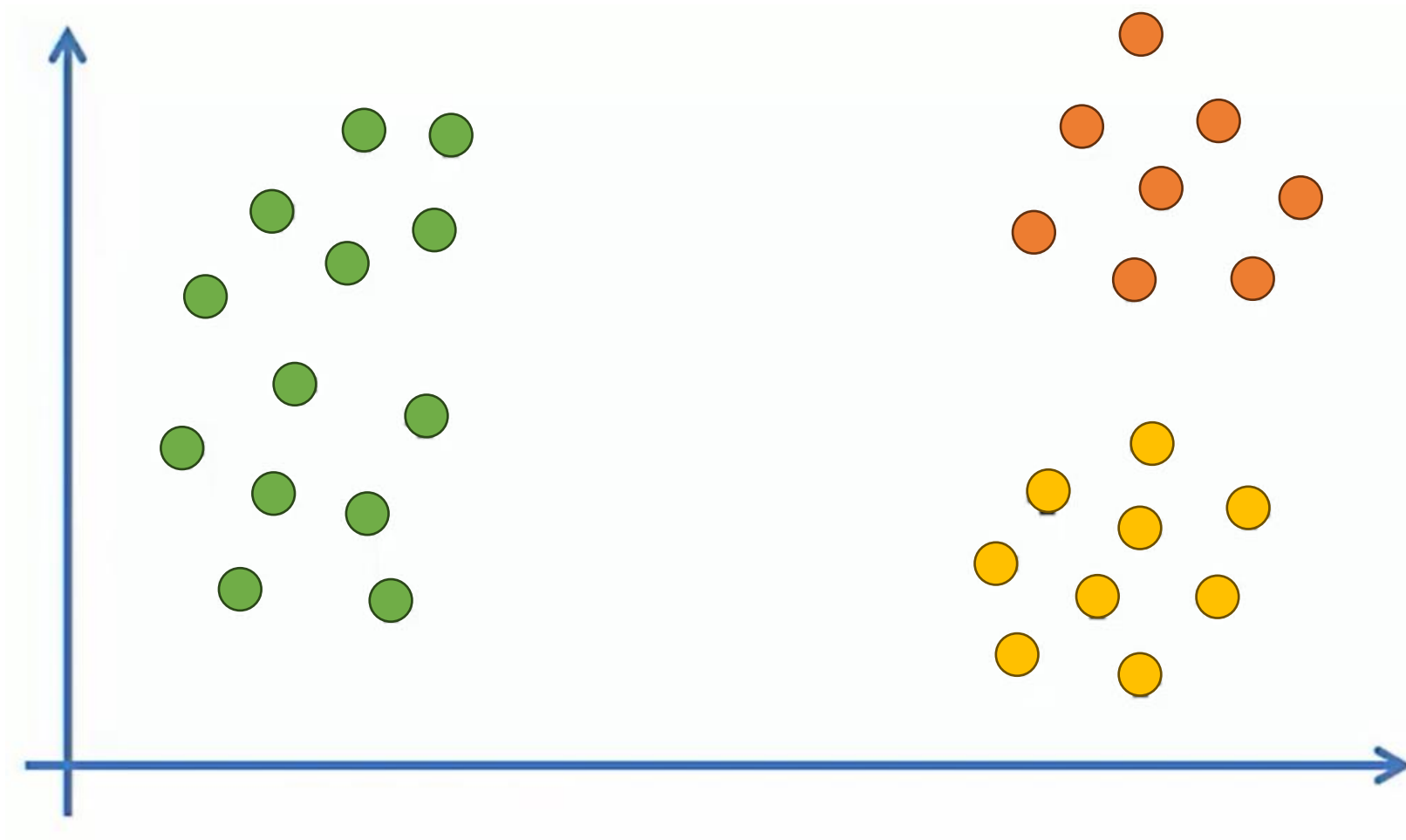
# Random Initialization Trap in K-means Clustering

Dr. Muhammad Wasim

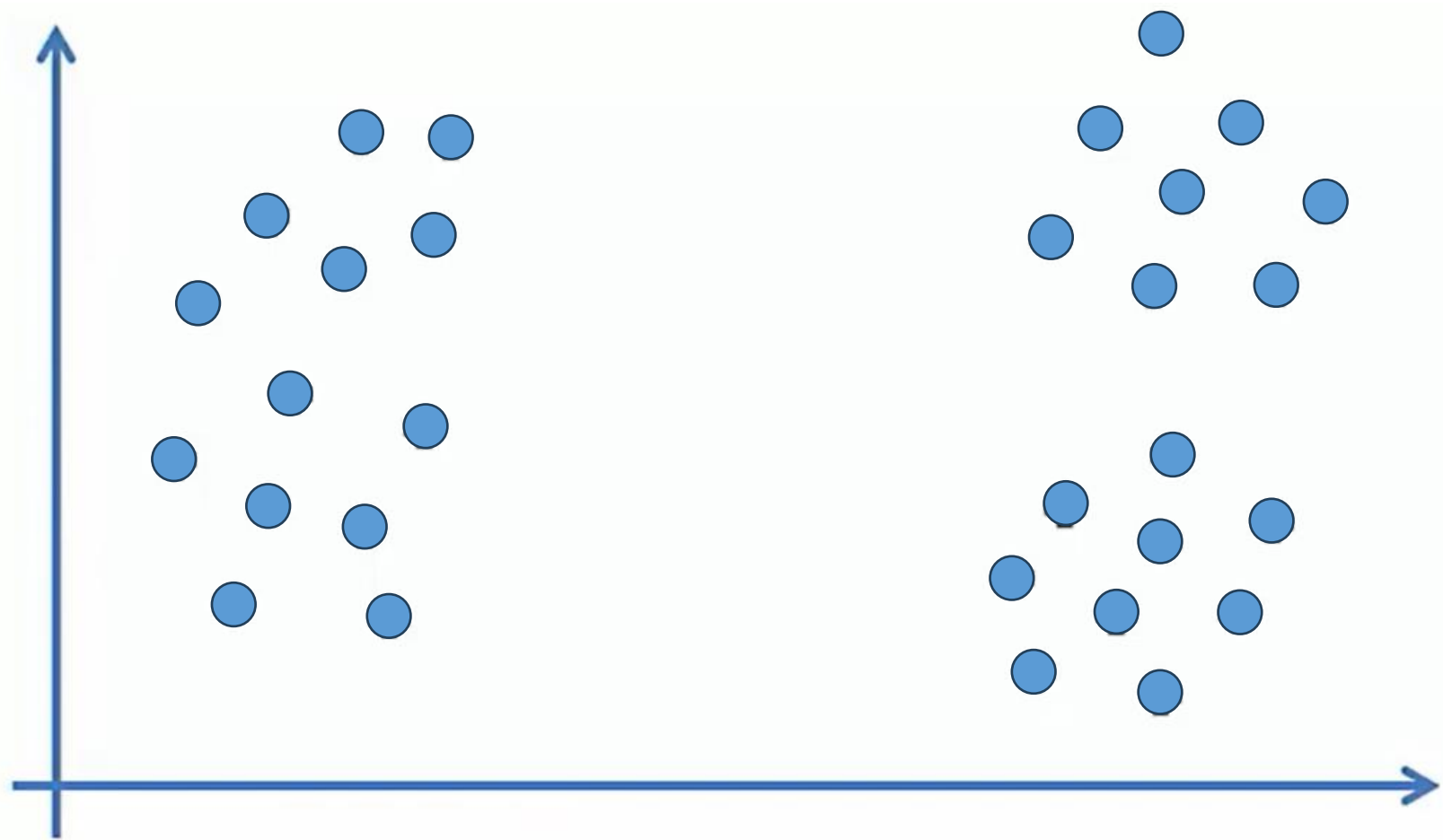
# Random Initialization Trap



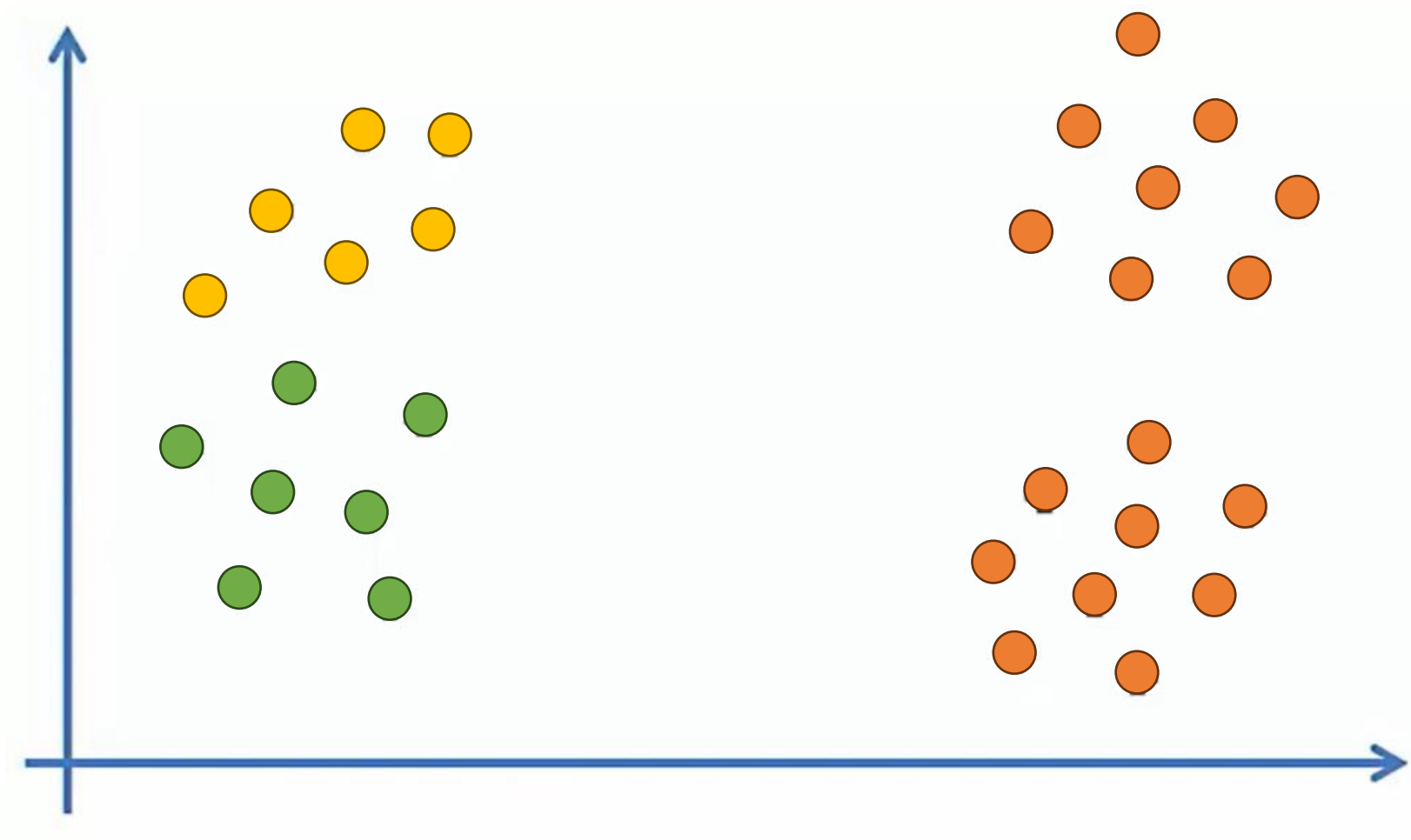
# Random Initialization Trap



# Random Initialization Trap



# Random Initialization Trap



# Solution: K-means++

- Choose random points multiple times
- For every random point, run the K-means clustering algorithm and evaluate the cluster using some measure such as **within cluster sum of squares (WCSS)**.
- For cluster  $S_k$  with center  $\mu_k$ , define the squared error measure to quantify the quality of clusters:

$$\sum_{x_n \in S_k} ||x_n - \mu_k||^2$$

- The k-means error function just sums this cluster error over all clusters known as **WCSS**:

$$\sum_{k=1}^K \sum_{x_n \in S_k} ||x_n - \mu_k||^2$$

# WCSS Intuition

$$\sum_{k=1}^K \sum_{x_n \in S_k} ||x_n - \mu_k||^2$$

