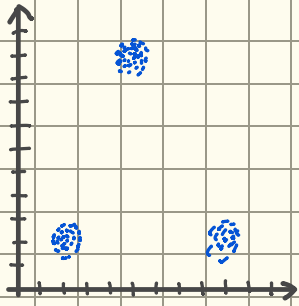


# KMEANS

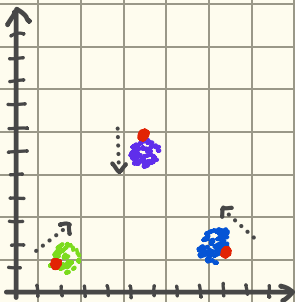


①  $K=3$ , `kmeans++`

Problem: We want to identify the groups or "clusters" within the dataset.

K-Means:

1. Pick  $K$  centroids by:
  - a. pick random point
  - b. find point further from points picked
  - c. loop a-b until you have picked  $K$  points
2. create clusters by associating point with closest centroid
3. Move centroid to average of cluster
4. if centroid moved more than tolerance continue



①  $K=3$ , `kmeans++`

Placed red dots according to Step 1

Created clusters (1, 2, 3)

Movement of centroids indicated by .....>