

K-Means

Algorithm

- 1) Randomly choose k data points as initial centroids
- 2) Assign each data point to the closest centroid
- 3) Recompute centroids
- 4) Do this (2-4) until convergence criterion is met

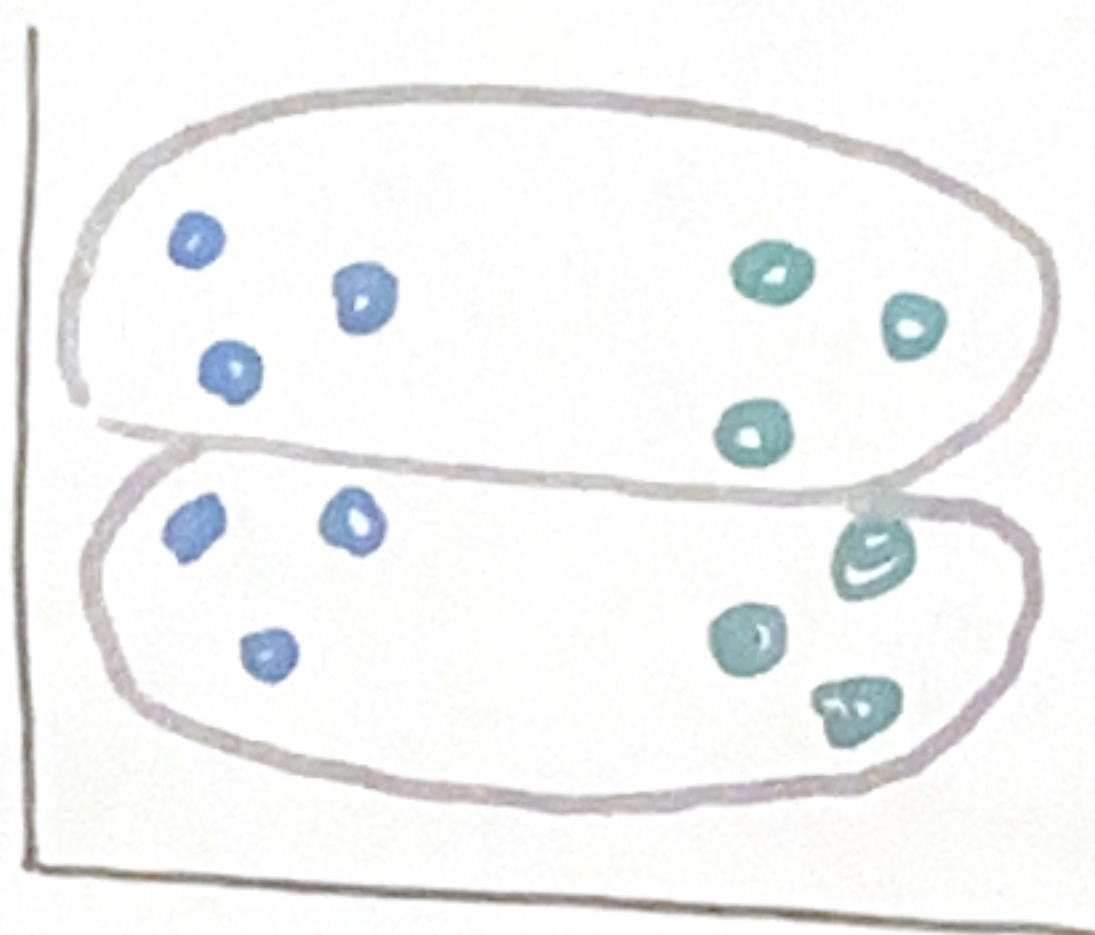
* Time complexity: $O(tkn)$

#iterations \leftarrow #clusters } t & k is so small that this is considered linear

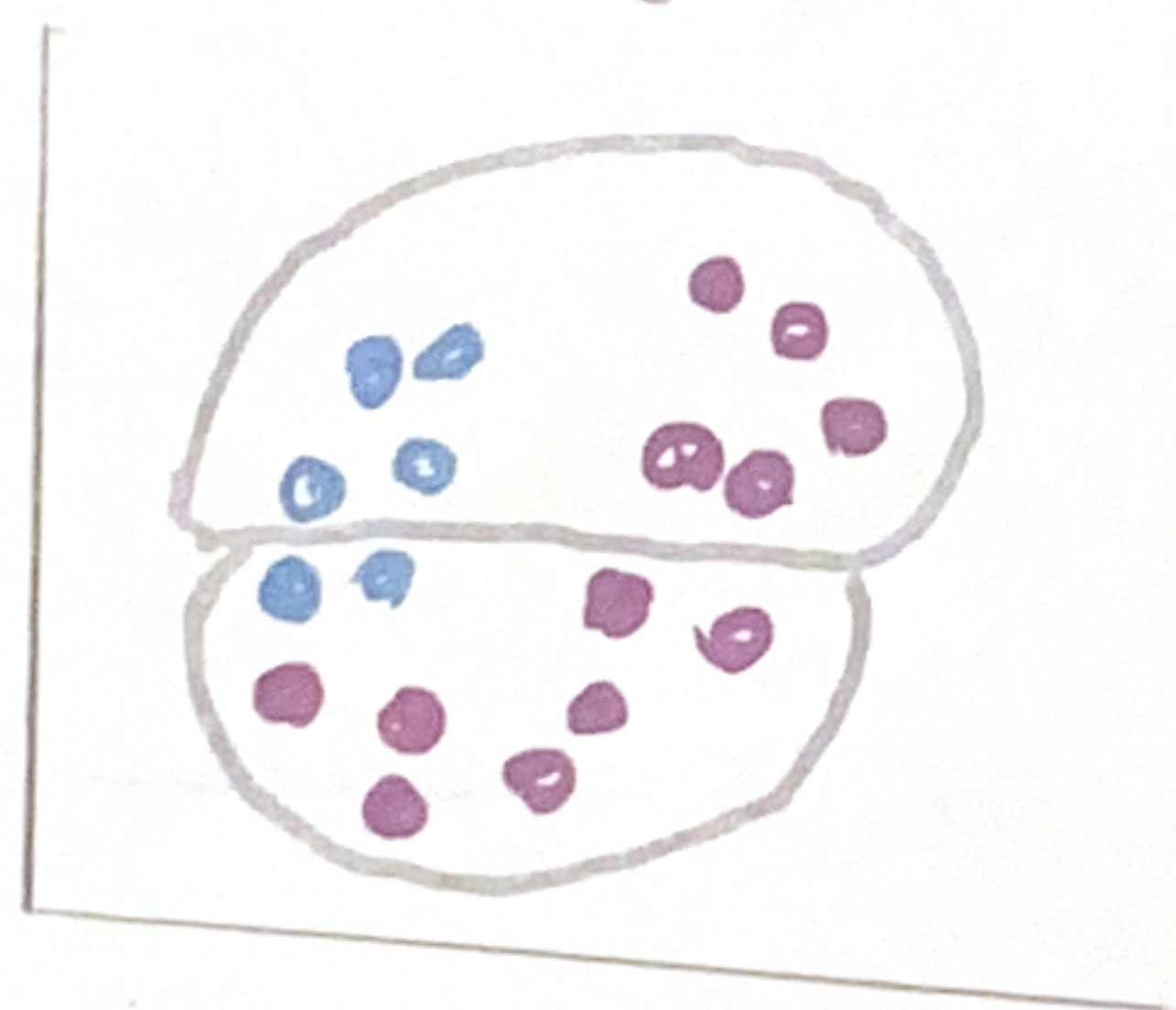
* Algorithm is sensitive to outliers



Sensitive to outliers



Sensitive to initial seeds



→ Performs well on hyperspheres but not on clusters with no hyperspheres
Use hierarchical clustering!