Applied Machine Learning

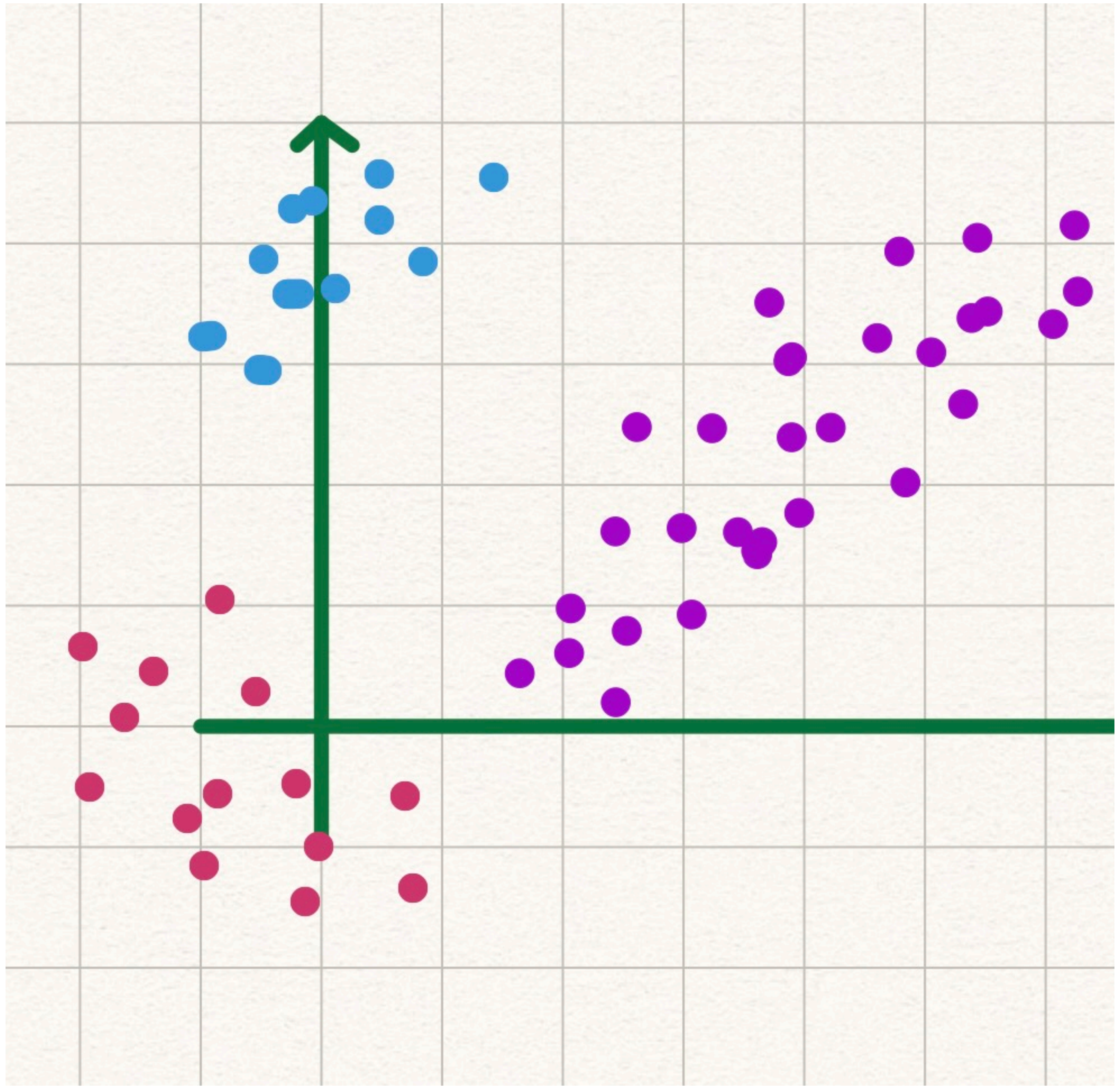
k-Means Clustering - Variants

k-Means Clustering - Variants

- k-Means clustering algorithm
- Soft assignment of clusters
- Hierarchical k-Means
- k-Medoids

k-Means Clustering

1. Initialization: choose k data items as cluster centers $\mathbf{c}_j$
2. While (cluster centers have significant changes)
    1. For each data item $\mathbf{x}_i$
        - **closest_center_from($\mathbf{x}_i$).assign($\mathbf{x}_i$)**
    2. For each empty cluster center $\mathbf{c}_j$
        - **$\mathbf{c}_j$.assign(select_far_item_from($\mathbf{c}_j$))**
    - For each cluster center $\mathbf{c}_j$
        - **$\mathbf{c}_j$.center = $\mathbf{c}_j$.mean()**



Soft Assignment of Clusters

- Items may be at a similar distance to more than one cluster center
    - Hard assignment: item i assigned to cluster center j through $\delta_{i,j}$
    - Soft assignment: item i assigned to cluster center j with weight $w_{i,j}$
        - $\sum_j w_{i,j} = 1 \qquad w_{i,j} > 0$
        - $w_{i,j} = \frac{s_{i,j}}{\sum_{l=1}^{k} s_{i,l}}$

- Affinity between point $\mathbf{x}_i$ and center $\mathbf{c}_j$
    - $s_{i,j} = e^{\frac{-\|\mathbf{x}_i - \mathbf{c}_j\|^2}{2\sigma^2}}$
    - σ: scaling parameter
    - the longer the distance between $\mathbf{x}_i$ and $\mathbf{c}_j$ in units of σ, the smaller the affinity

Hard Assignment vs Soft Assignment

- Hard assignment
- $$\Phi(\delta, c) = \sum_{i,j} \delta_{i,j} \left[(\mathbf{x}_i - \mathbf{c}_j)^\top (\mathbf{x}_i - \mathbf{c}_j) \right]$$
  - $$\delta_{i,j} = \begin{cases} 1 & \mathbf{x_i} \text{ belongs to cluster } j \\ 0 & \text{otherwise} \end{cases}$$
  - $$\sum_j \delta_{i,j} = 1 \qquad \sum_i \delta_{i,j} > 0$$

- Soft assignment
- $$\Phi(\delta, c) = \sum_{i,j} w_{i,j} \left[(\mathbf{x}_i - \mathbf{c}_j)^\top (\mathbf{x}_i - \mathbf{c}_j) \right]$$
  - $$w_{i,j} = \frac{s_{i,j}}{\sum_{l=1}^{k} s_{i,l}}$$
  - $$\sum_j w_{i,j} = 1 \qquad w_{i,j} > 0$$
- $\mathbf{c}_j$.center = $\dfrac{\sum_i w_{i,j} \mathbf{x}_i}{\sum_i w_{i,j}}$

Hierarchical k-Means

- Each k-Means iteration needs distances from each of the k clusters to all items
- Hierarchical k-means:
  - subsample dataset at random to identify cluster centers using small k
  - allocate every item in the dataset to clusters
  - in each of the k clusters:
    - cluster recursively with k-Means, or with hierarchical k-Means

k-Medoids

- Option when items cannot be averaged to compute centers
- Cluster centers can be data items

1. Initialization: choose k data items at random as cluster centers $\mathbf{c}_j$
2. While (cluster centers have significant changes)
    1. For each data item $\mathbf{x}_i$
        - closest_center_from($\mathbf{x}_i$).assign($\mathbf{x}_i$)
    - For each cluster center $\mathbf{c}_j$
        - $\mathbf{c}_j$.center = $\mathbf{c}_j$.choose_best_medoid()

k-Means Clustering - Variants

- k-Means clustering algorithm
- Soft assignment of clusters
- Hierarchical k-Means
- k-Medoids

Applied Machine Learning

k-Means Clustering - Variants