資料應用與機器學習

Classification and Clustering

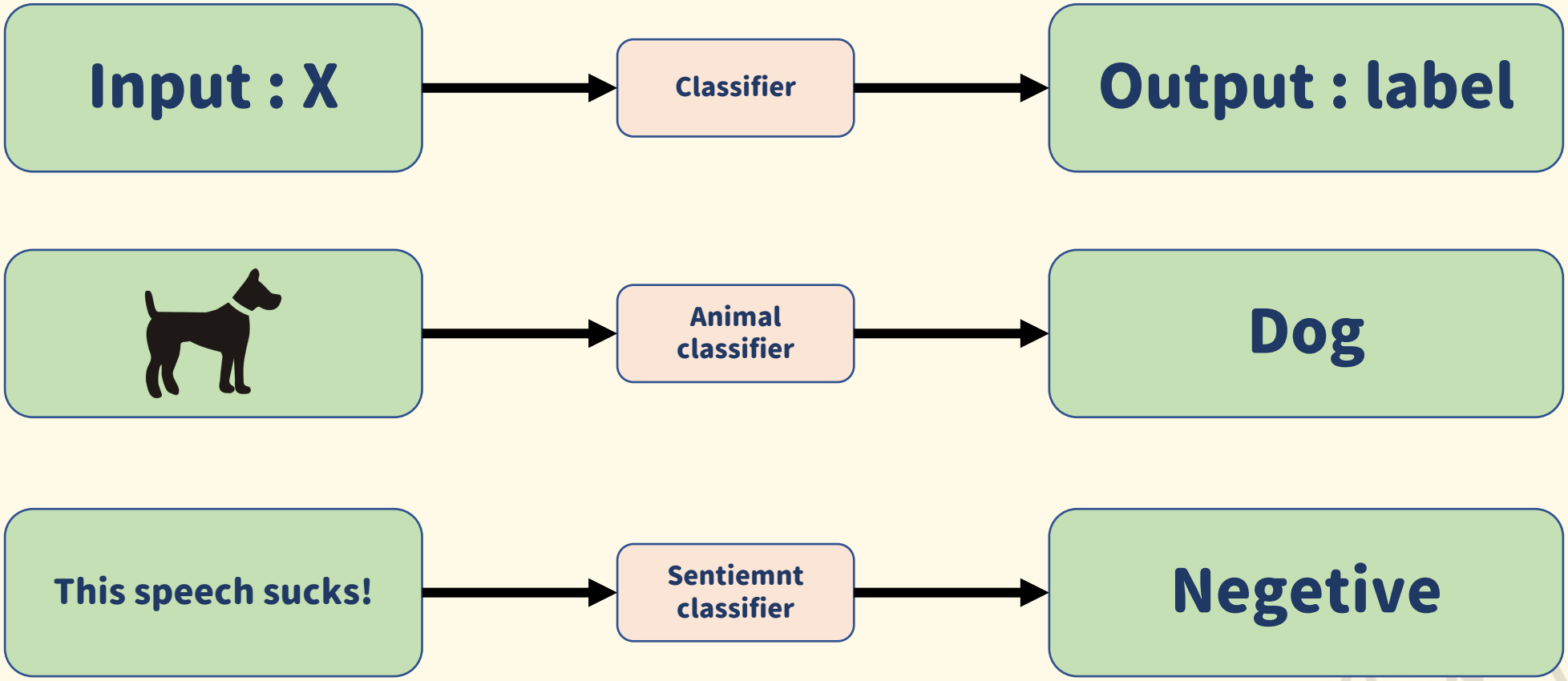
OutLine

- Classification
- Clustering
- Practical use
  - Classification(Air Data, pyCIOT)
  - Clustering(Air Data, https://history.colife.org.tw/)
  - Comprehensive application

Slido - 2259802

Classification

What is Classification?

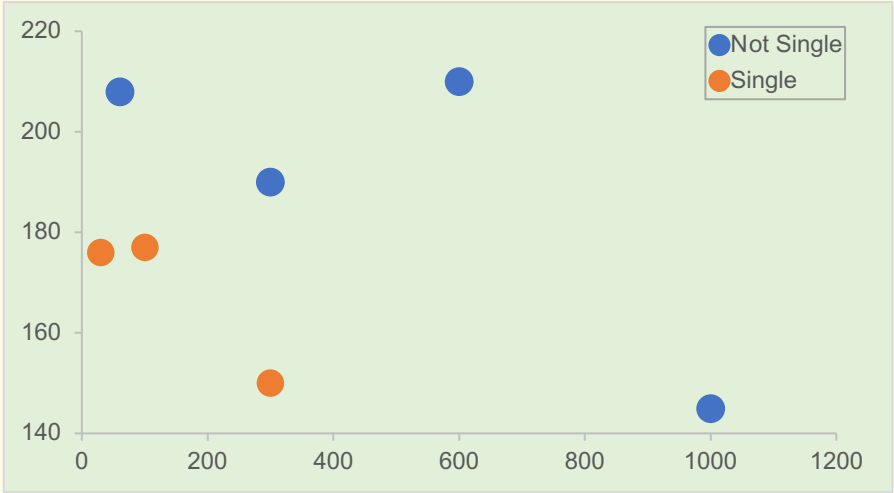


How to form classifier?

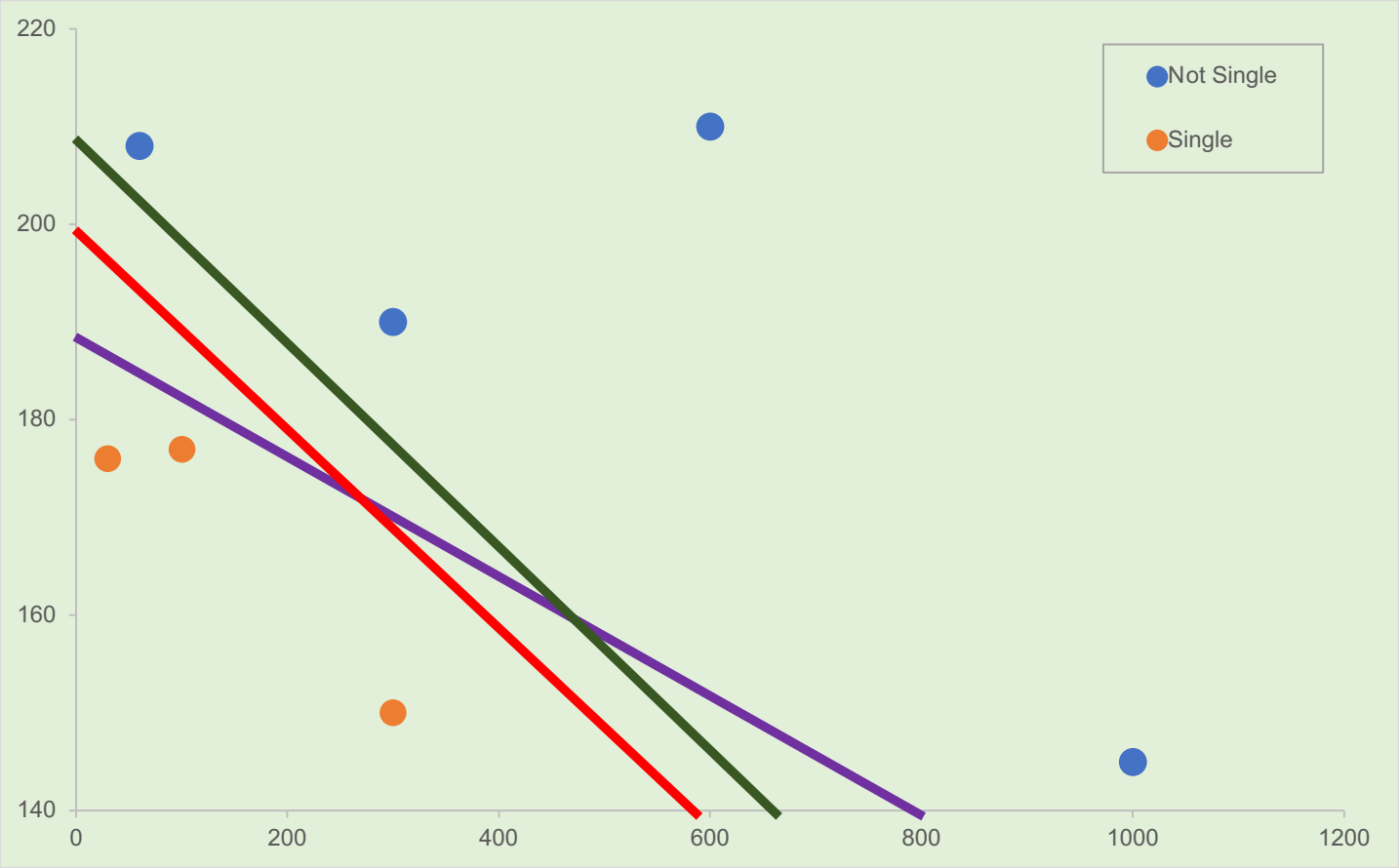
- SVM
- KNN
- Decision Tree
- And more…

SVM

Name	Salary (k)	Height (cm)	Single or not?
Bill	600	210	No
Jobs	300	190	No
Diess	60	208	No
Ring	300	150	Yes
Allen	50	177	Yes
John	30	176	Yes
Elon	1000	145	No



SVM



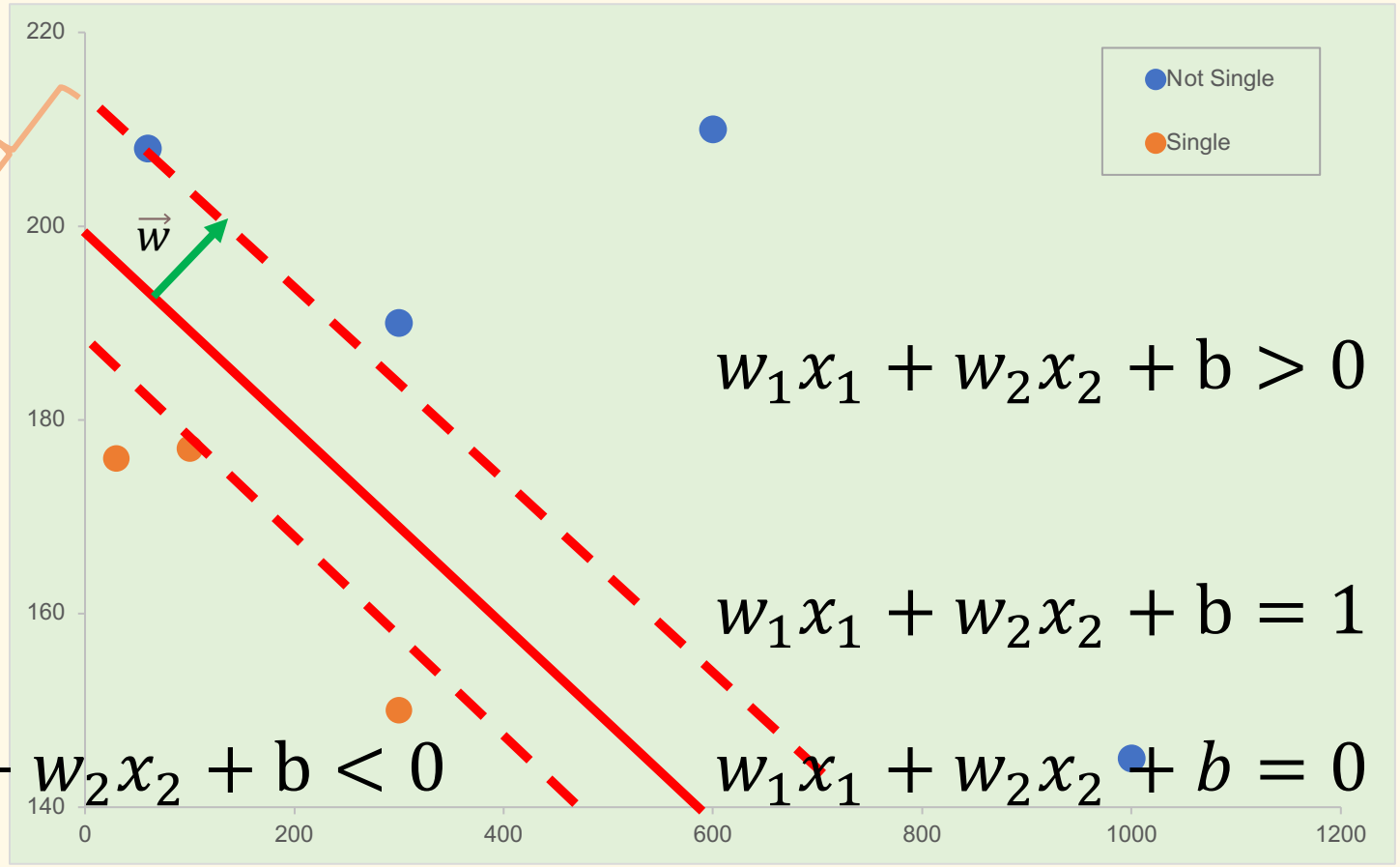
$$w_1 x_1 + w_2 x_2 + \cdots = k$$

$$w_1 x_1 + w_2 x_2 + \cdots + b = 0$$

$$w_1 x_1 + w_2 x_2 + \mathrm{b} = 0$$

Which one is the best one?

SVM



Margins

$$\frac{2}{\|\vec{w}\|}$$

$$w_1x_1 + w_2x_2 + \mathrm{b} > 0$$

$$w_1x_1 + w_2x_2 + \mathrm{b} = 1$$

$$w_1x_1 + w_2x_2 + b = 0$$

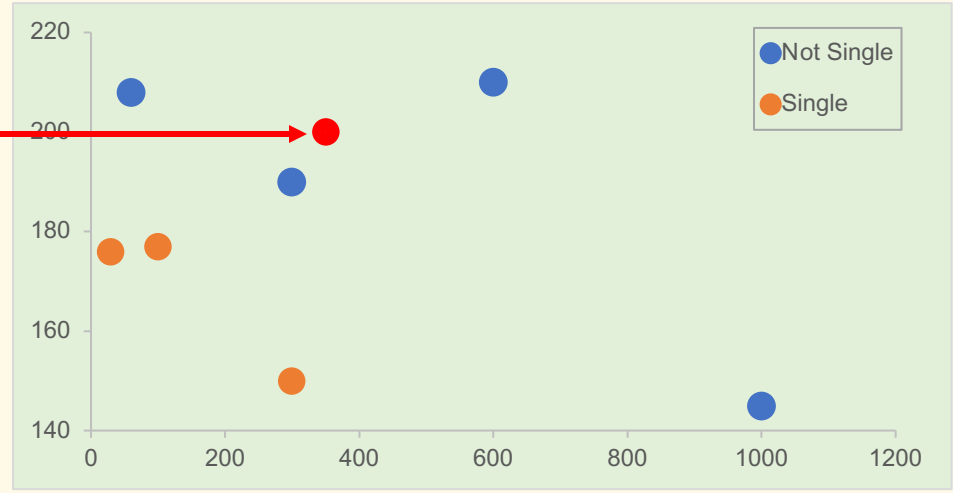
$$w_1x_1 + w_2x_2 + \mathrm{b} < 0$$

$$w_1x_1 + w_2x_2 + \mathrm{b} = -1$$

Make "Margins" as large as possible!!

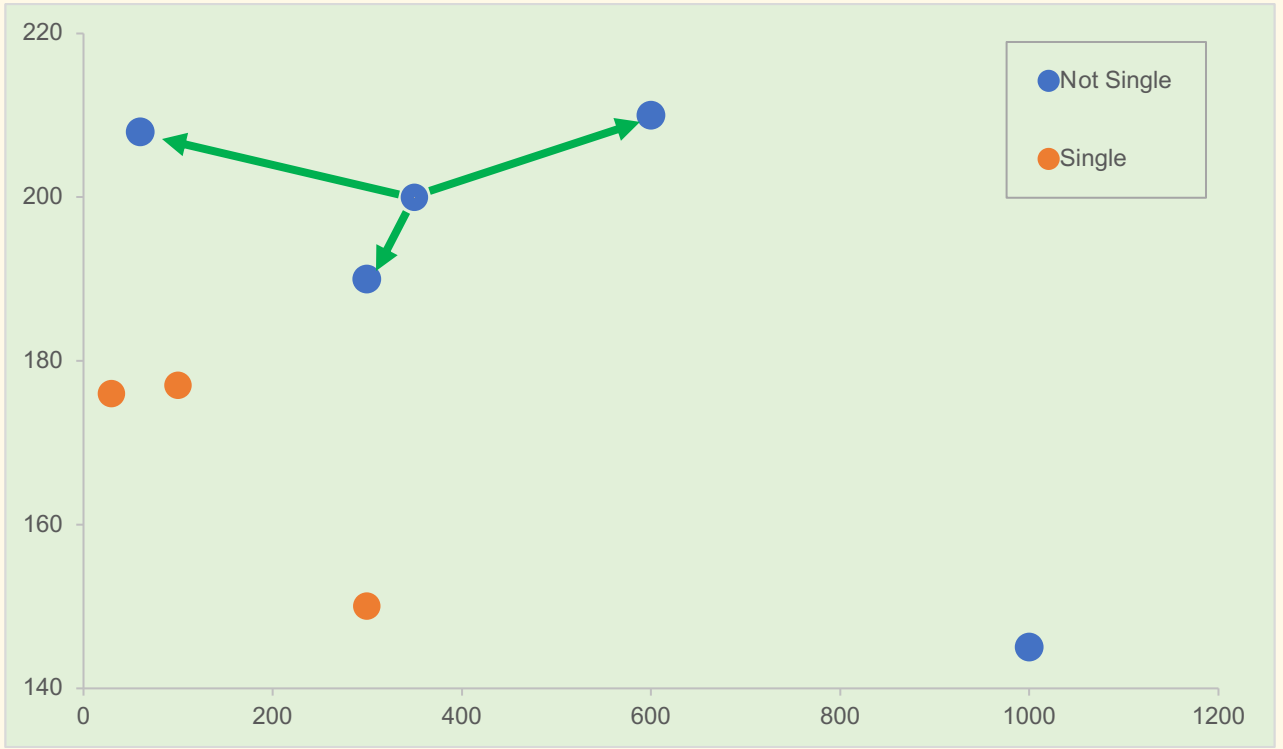
K-nearest Neighbors

Name	Salary (k)	Height (cm)	Single or not?
Bill	600	210	No
Jobs	300	190	No
Diess	60	208	No
Elon	1000	145	No
Allen	100	177	Yes
John	30	176	Yes
Ring	300	150	Yes
Alex	350	200	?



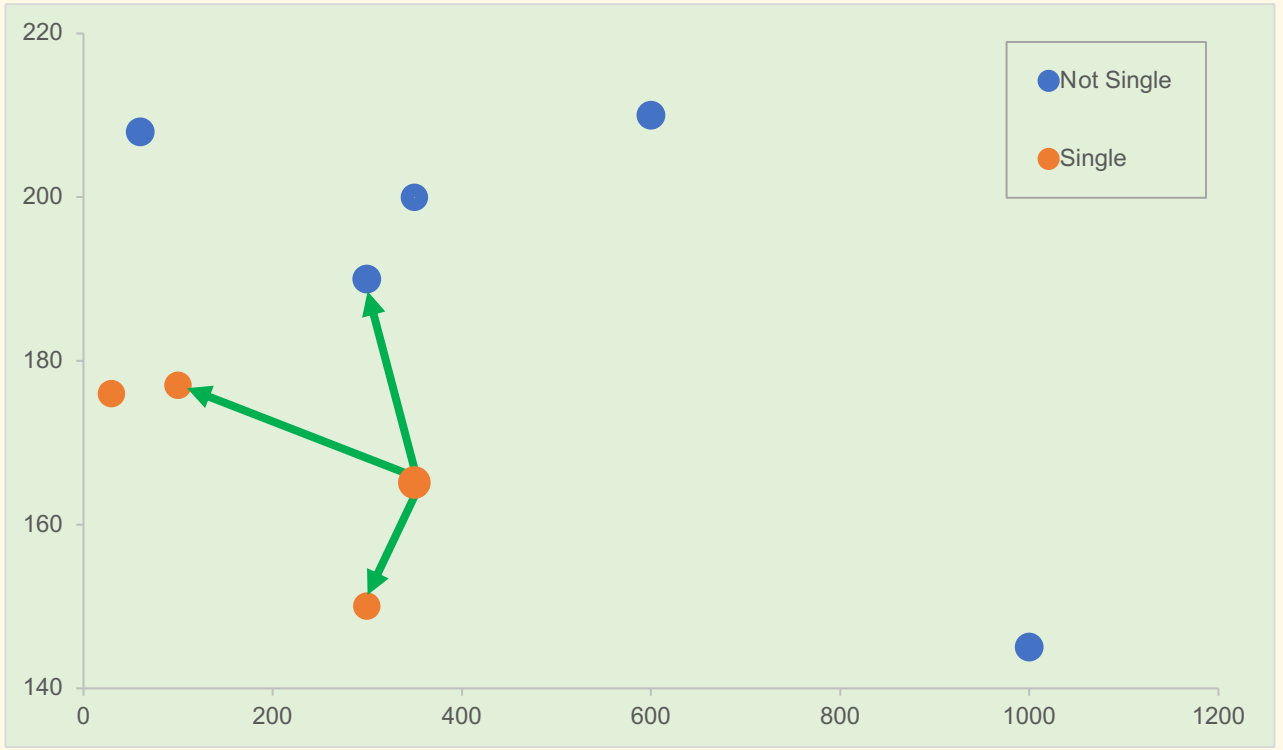
K-nearest Neighbors

K = 3



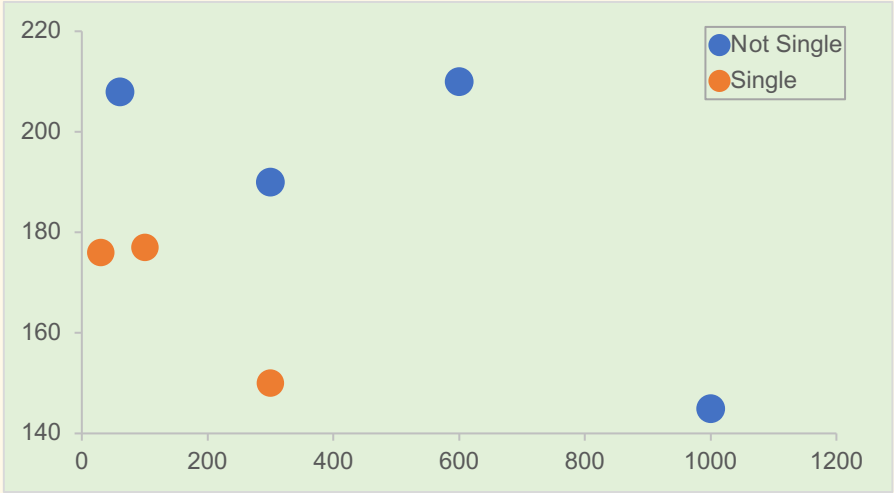
K-nearest Neighbors

K = 3

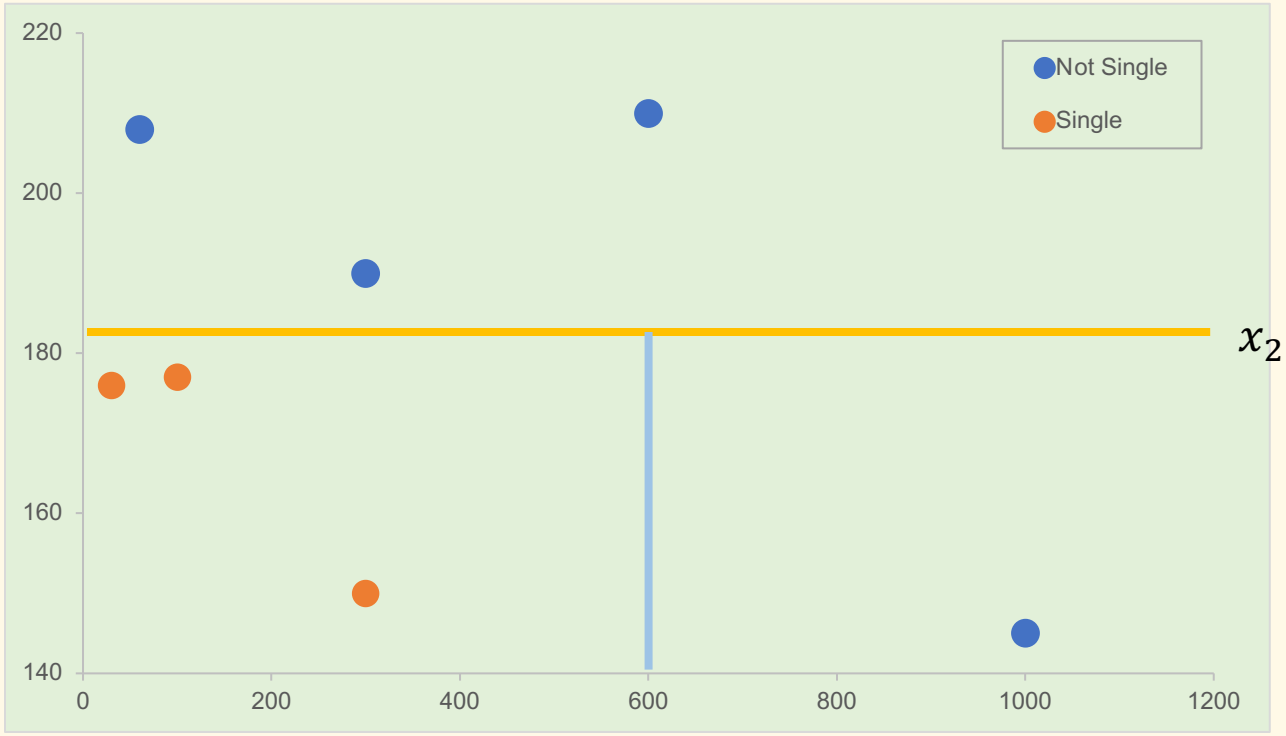


Decision Tree

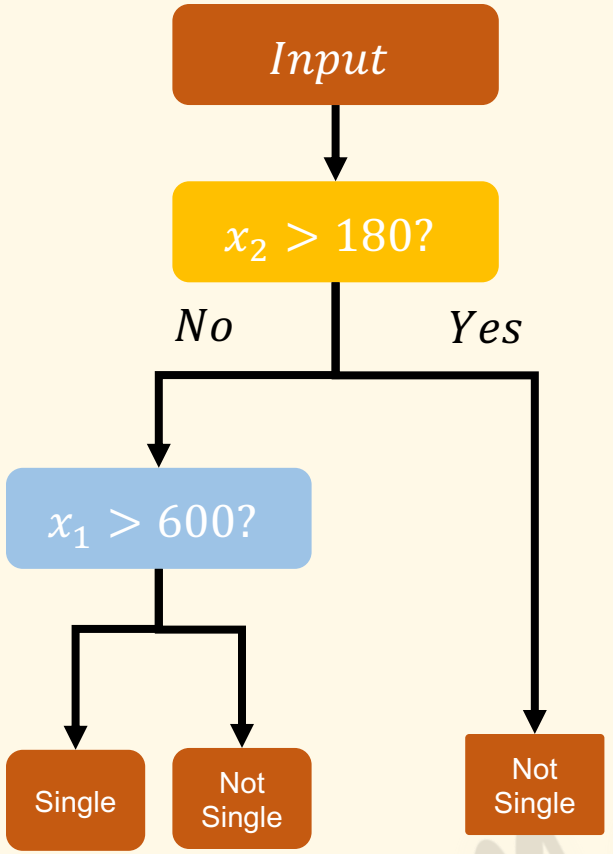
Name	Salary (k)	Height (cm)	Single or not?
Bill	600	210	No
Jobs	300	190	No
Diess	60	208	No
Ring	300	150	Yes
Allen	50	177	Yes
John	30	176	Yes
Elon	1000	145	No



Decision Tree

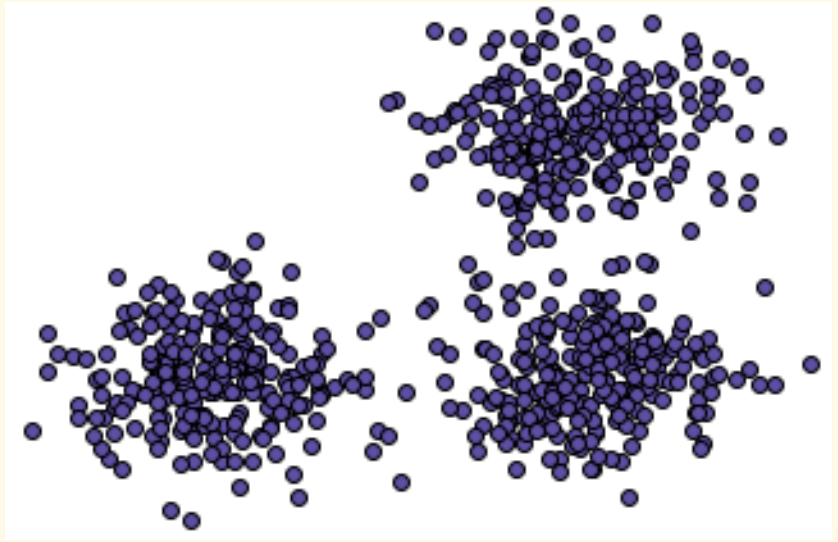


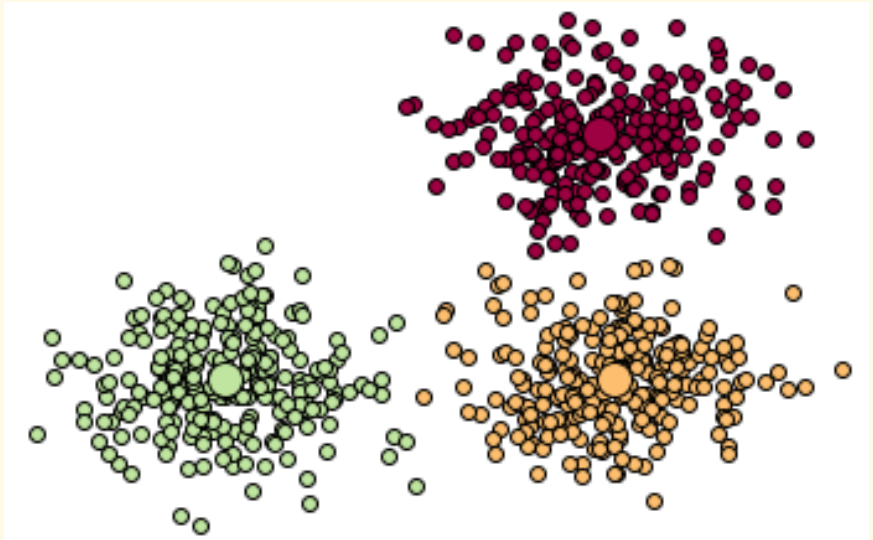
$x_2 = 180$



Clustering

What is Clustering?





Why do we need Clustering?

- Help people find out potential knowledge and make a good decision.

VIP

5000 up?

10000 up?

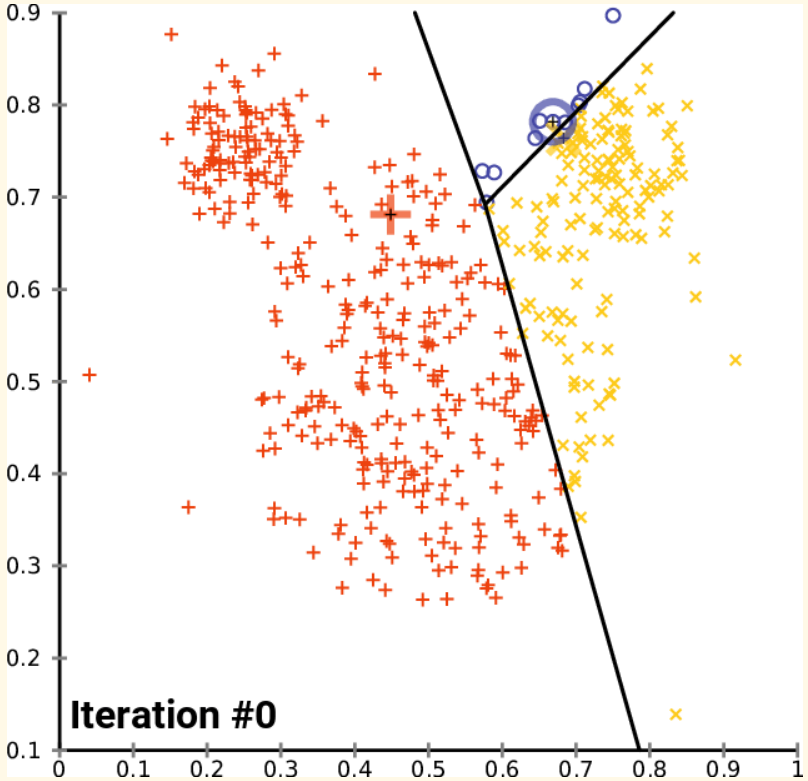
15000 up?

How to do clustering?

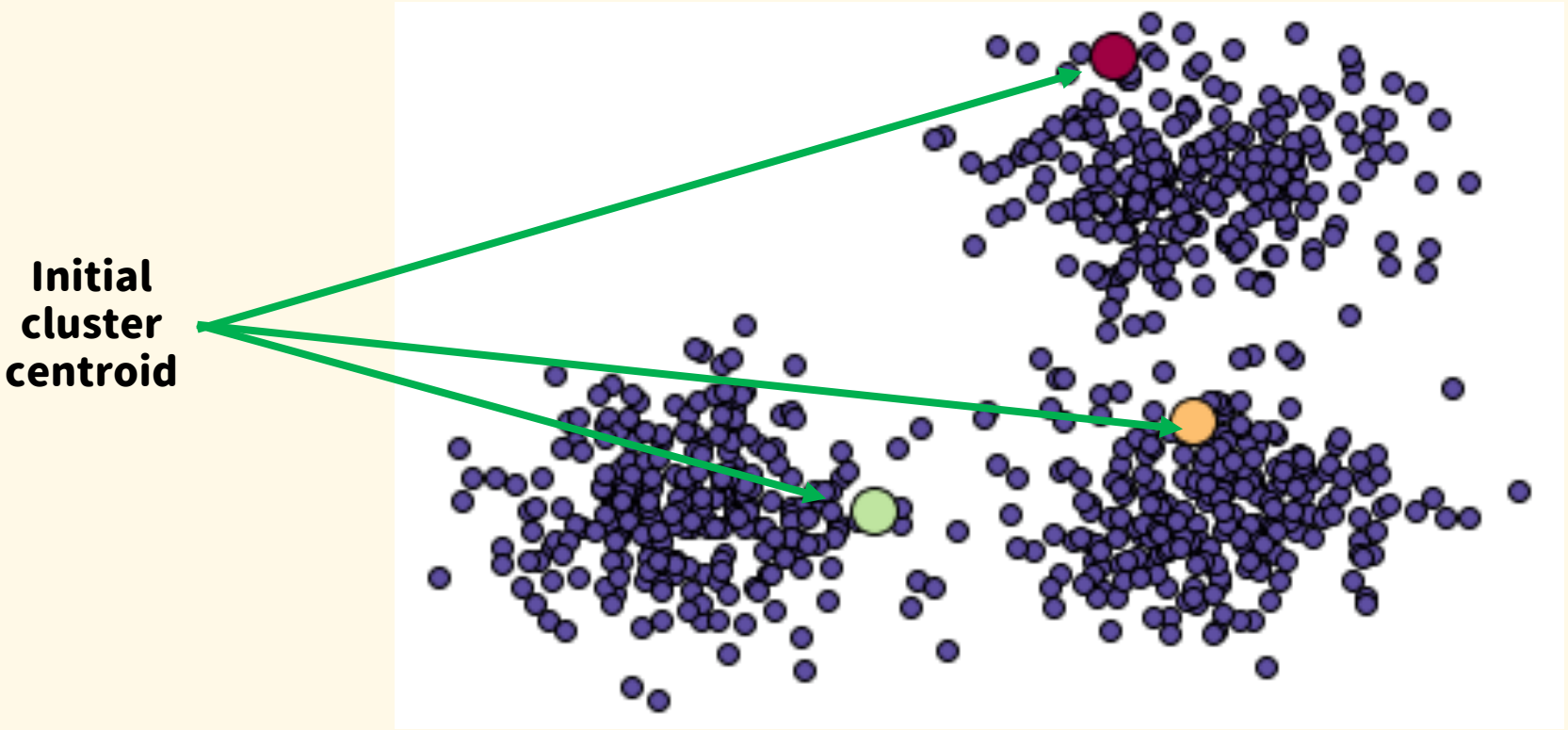
- K-means
- Hierarchical Clustering
- DBSCAN

K-Mean

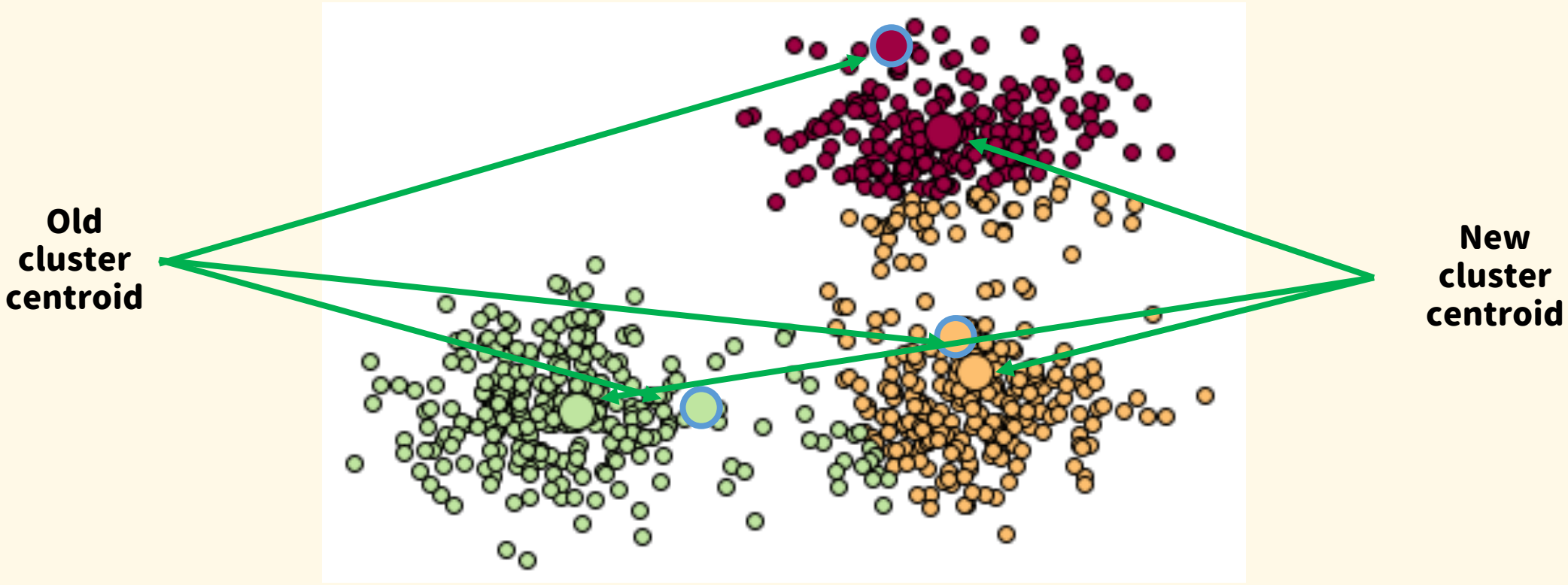
- Minimizes the variances between the data points and the cluster's centroid.



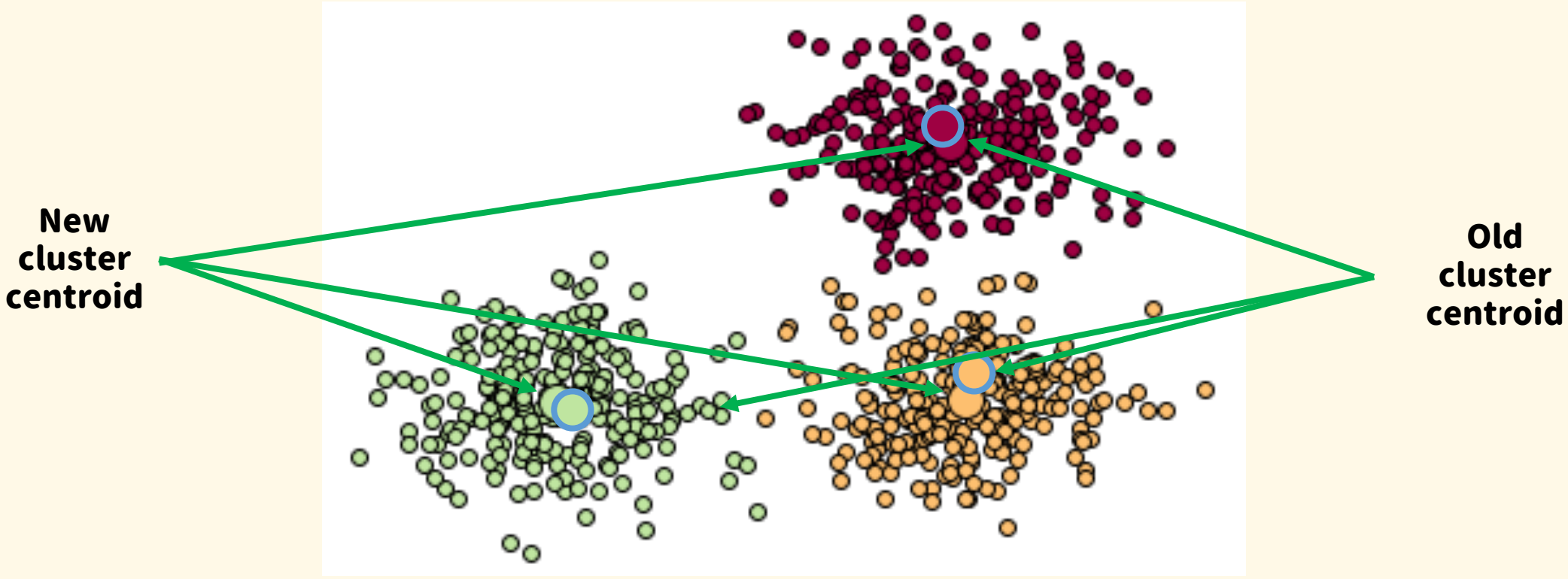
K-Mean



K-Mean

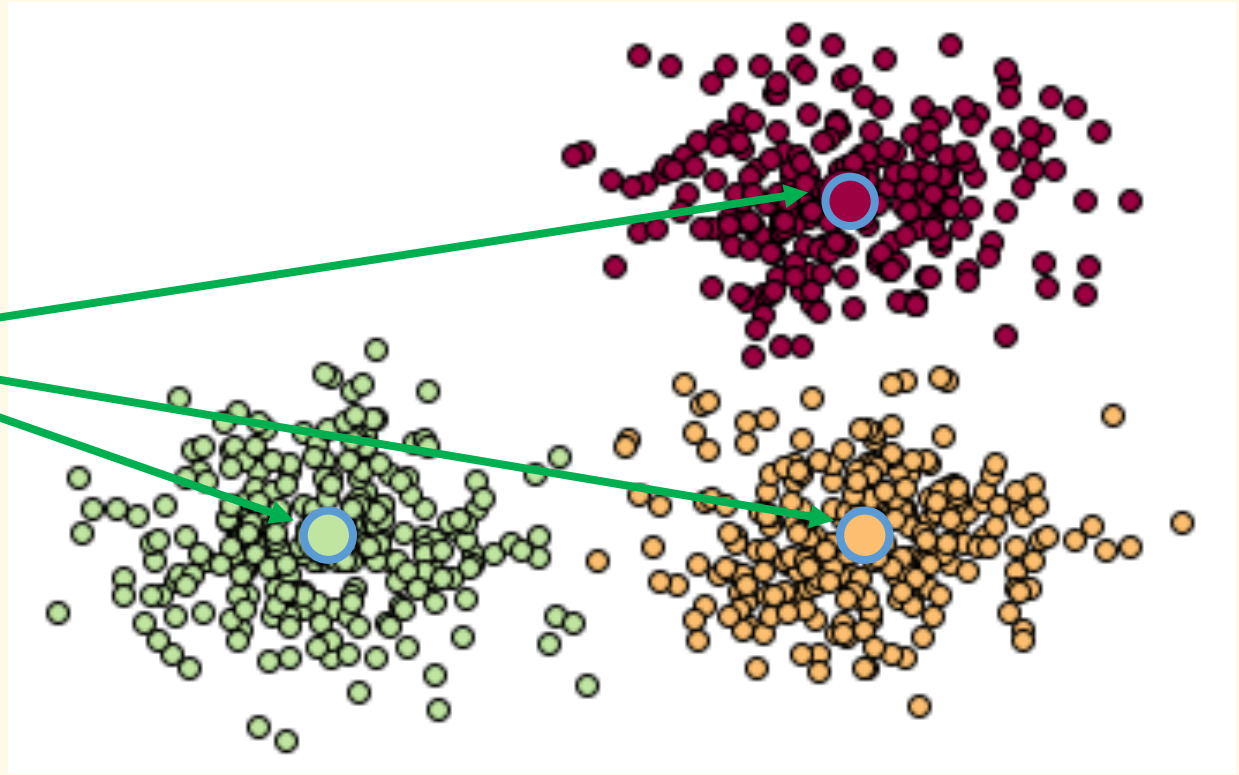


K-Mean



K-Mean

Cluster centroids didn't change!!



Converge ! End of algorithm.

Practical use

https://reurl.cc/mZrEOj (Google Colaboratory requires google account)

Practical use

https://reurl.cc/mZrEOj (Google Colaboratory requires google account)