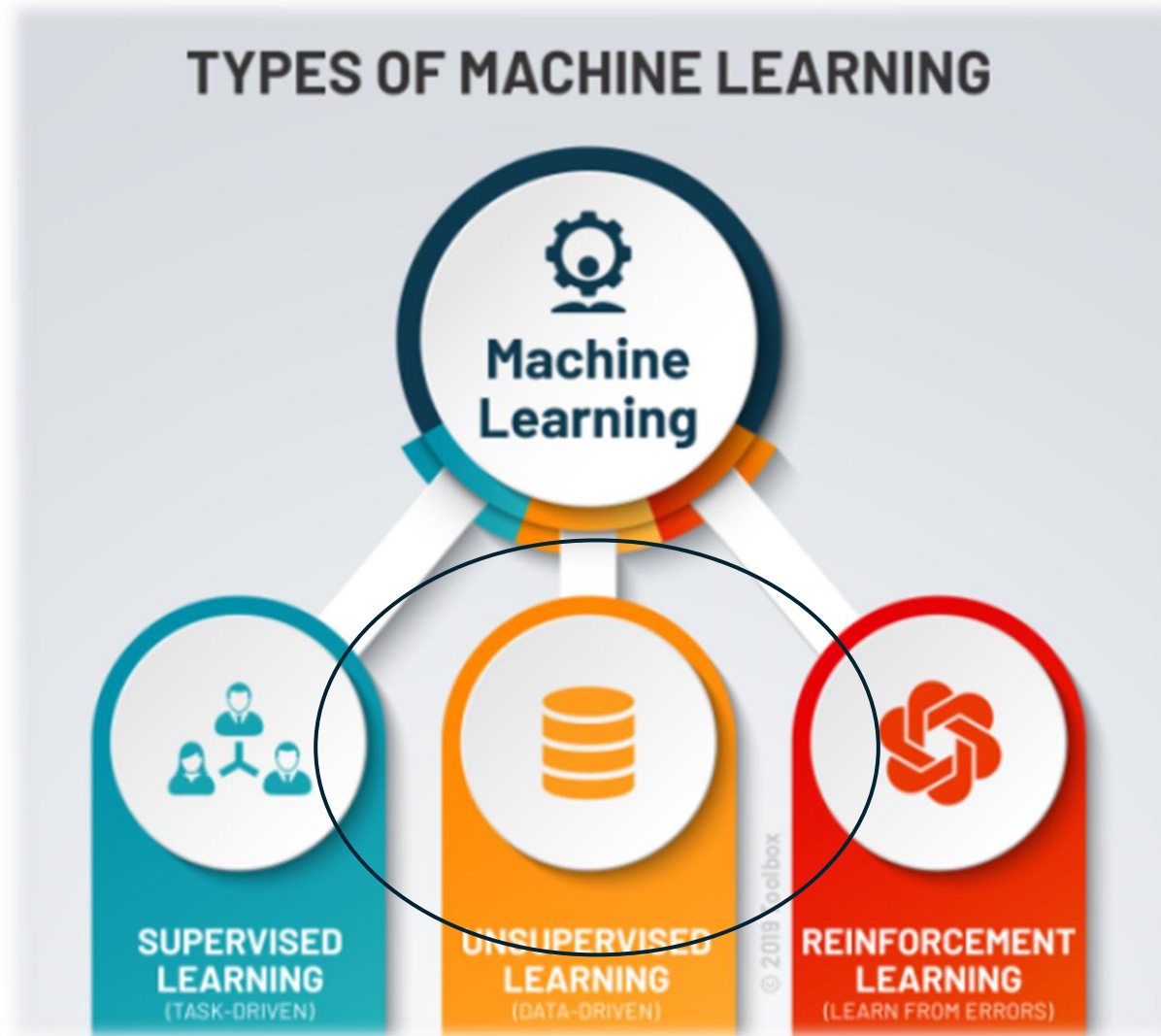


Machine Learning

Types of ML Models



Types of ML Models

| Aspect | Supervised Learning | Unsupervised Learning | Reinforcement Learning |
|------------|--|--|---|
| Goal | Learn a mapping from inputs to known outputs using labeled data | Discover hidden patterns or structure in unlabeled data | Learn by interacting with an environment and maximizing rewards |
| Input Data | Labeled data (input + correct output) | Unlabeled data | No direct labels; receives rewards or penalties through interaction |
| Output | Class label (classification) or numeric value (regression) | Clusters, groups, or low-dimensional structure | A policy (sequence of actions) or value function |
| Examples | Email spam detection, housing price prediction | Customer segmentation, anomaly detection | Game playing agents, robotics, autonomous driving |

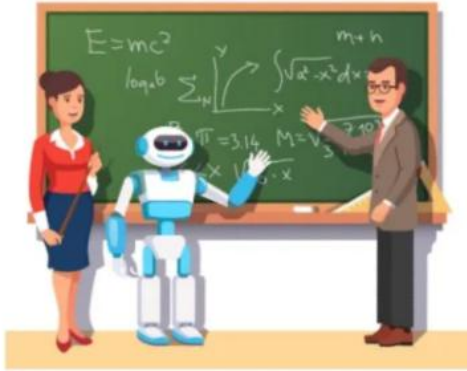
What is Clustering?

Clustering = grouping unlabeled data

What is Clustering?

Supervised Learning

(e.g. Regression, Classification)



Input data



Annotations

These are apples



Model



Prediction

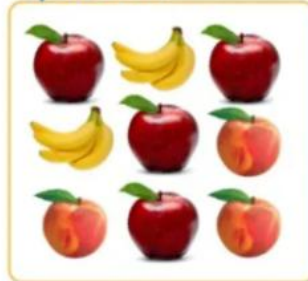
Its an apple!

Unsupervised Learning

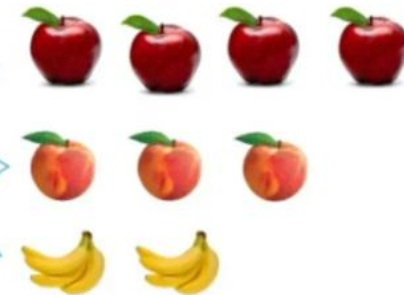
(e.g. Clustering)



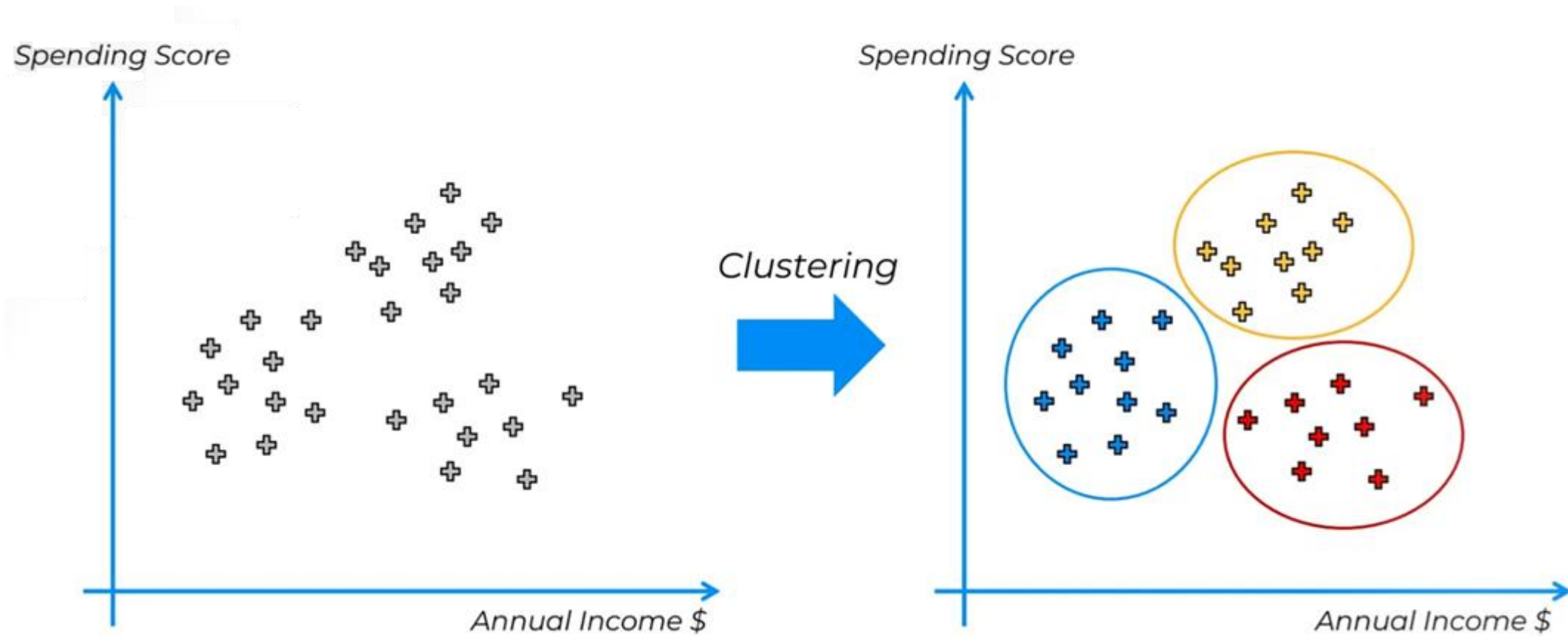
Input data



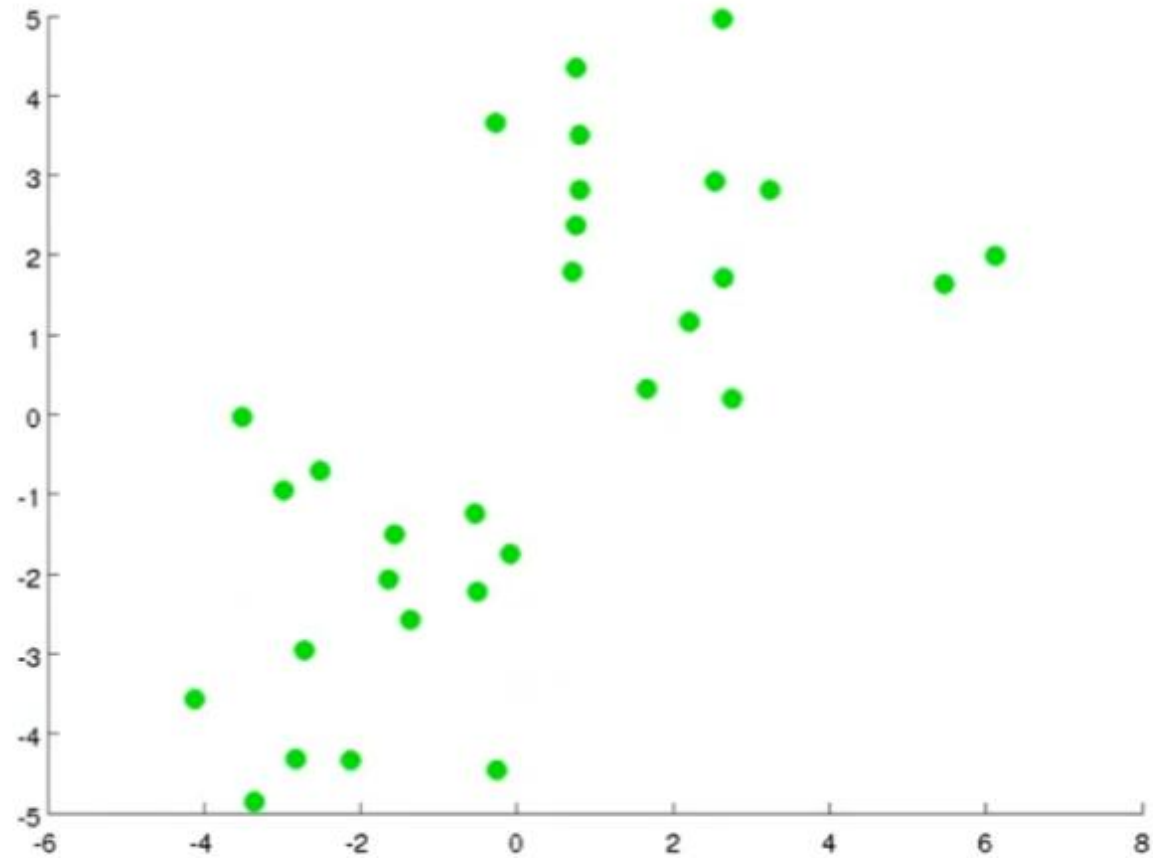
Model



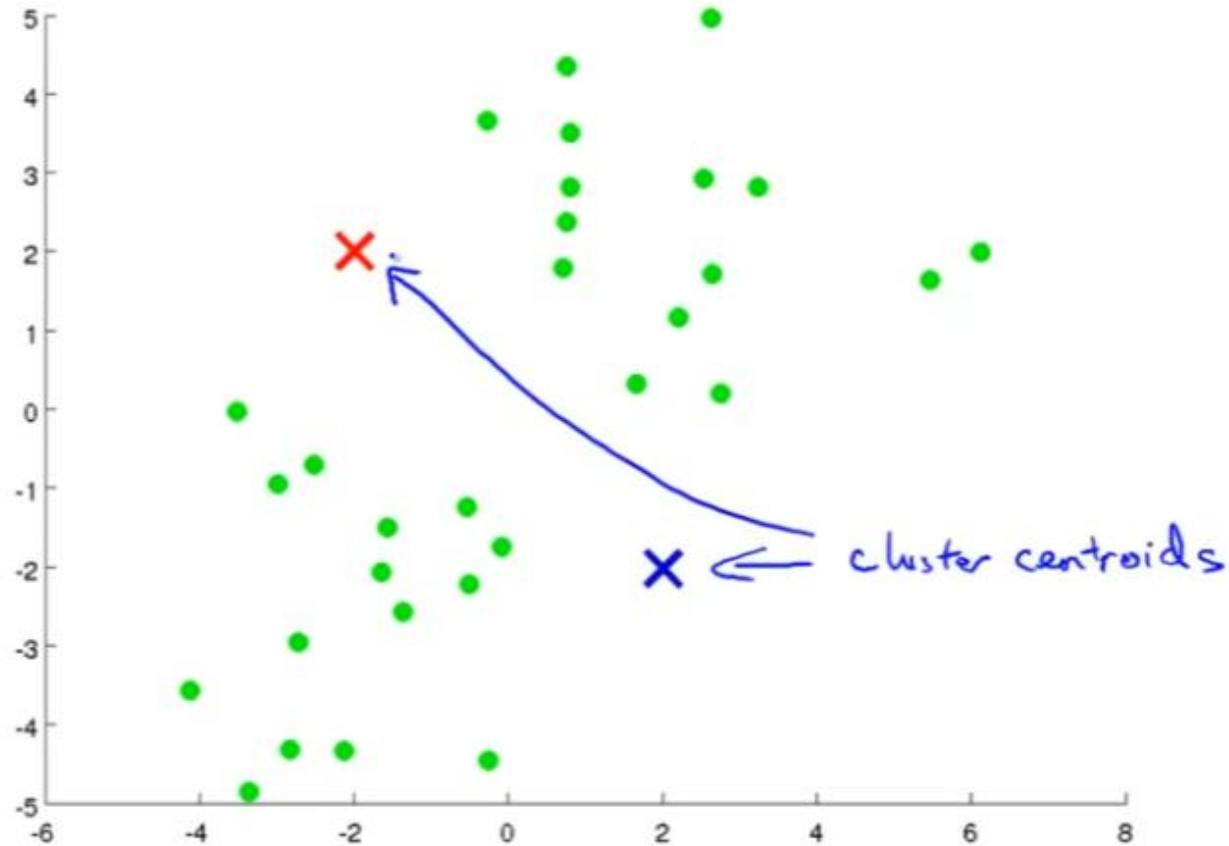
What is Clustering?



K-means Algorithm



K-means Algorithm

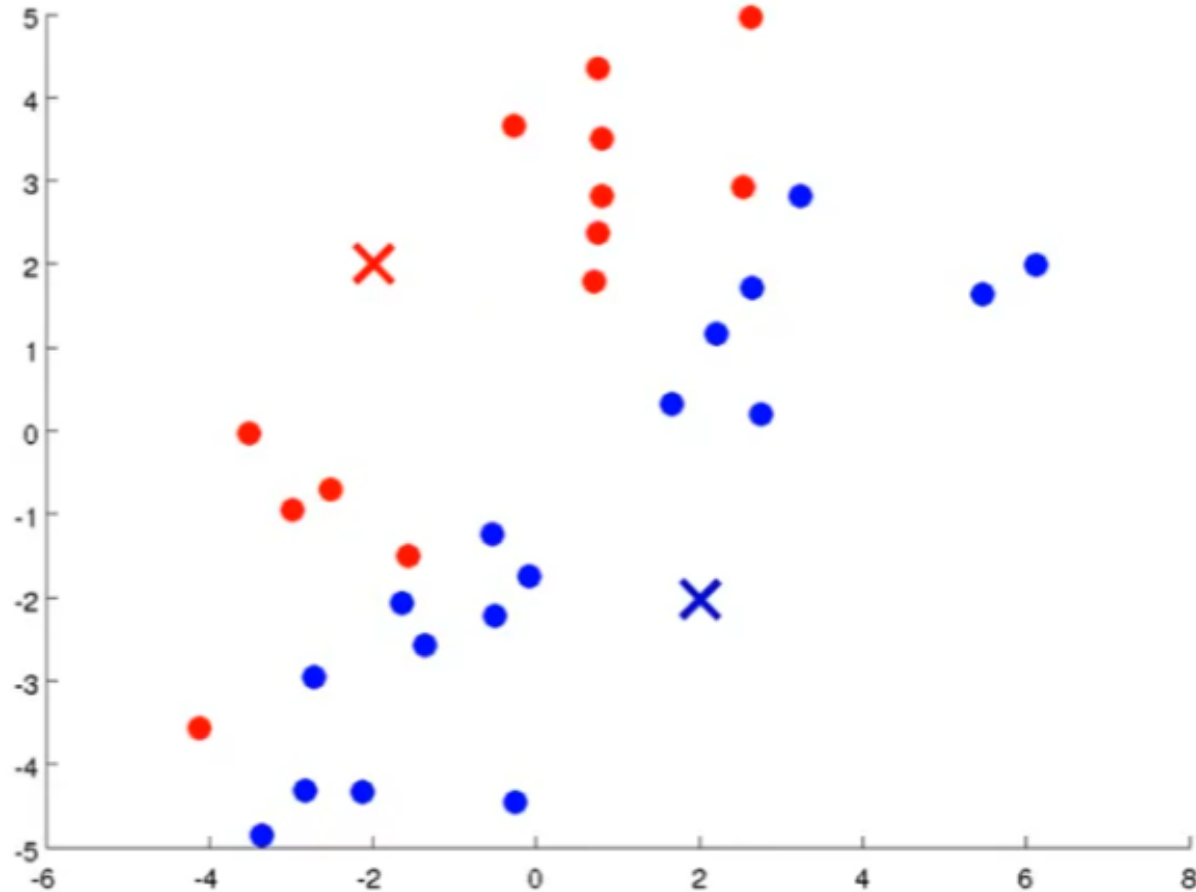


איתחול מרכזים

1. הקצאת האשכולות

2. העברת מרכזים

K-means Algorithm

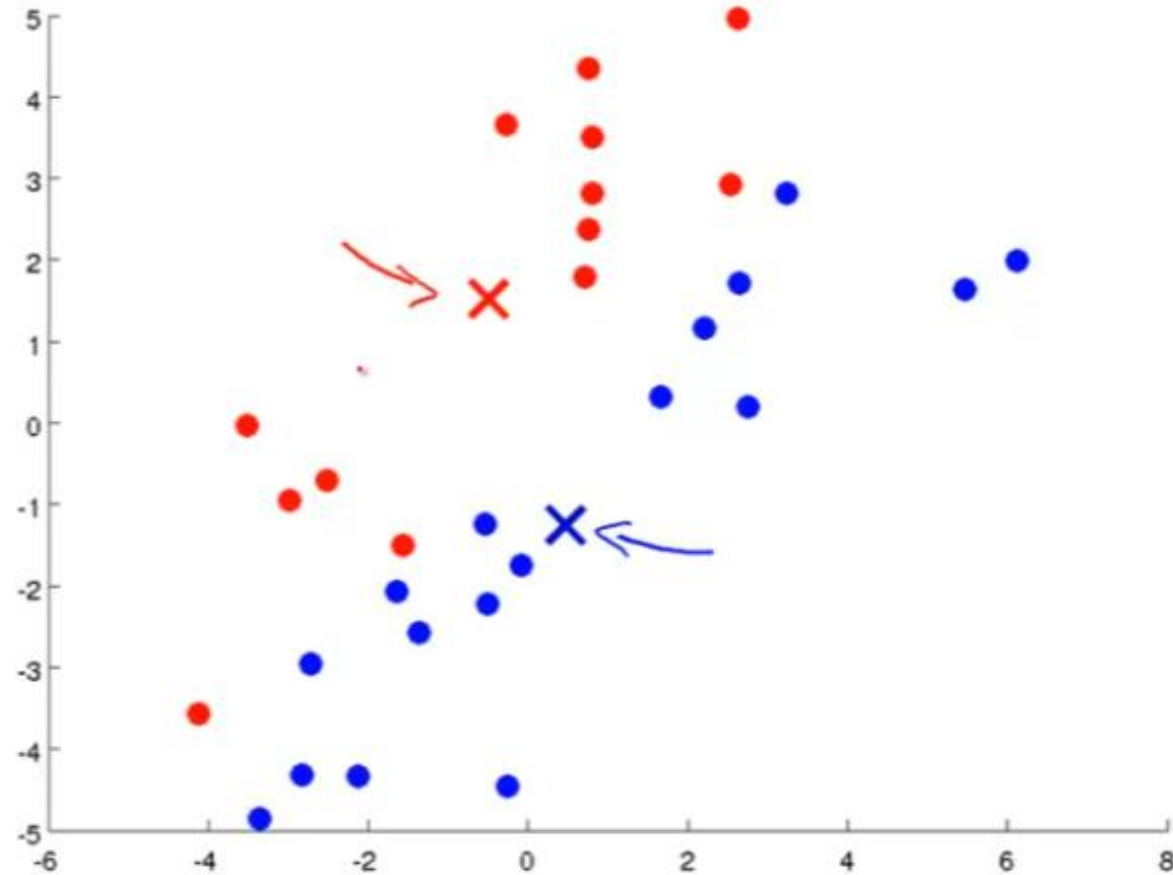


איתחול מרכזים

1. הקצאת האשכולות

2. העברת מרכזים

K-means Algorithm

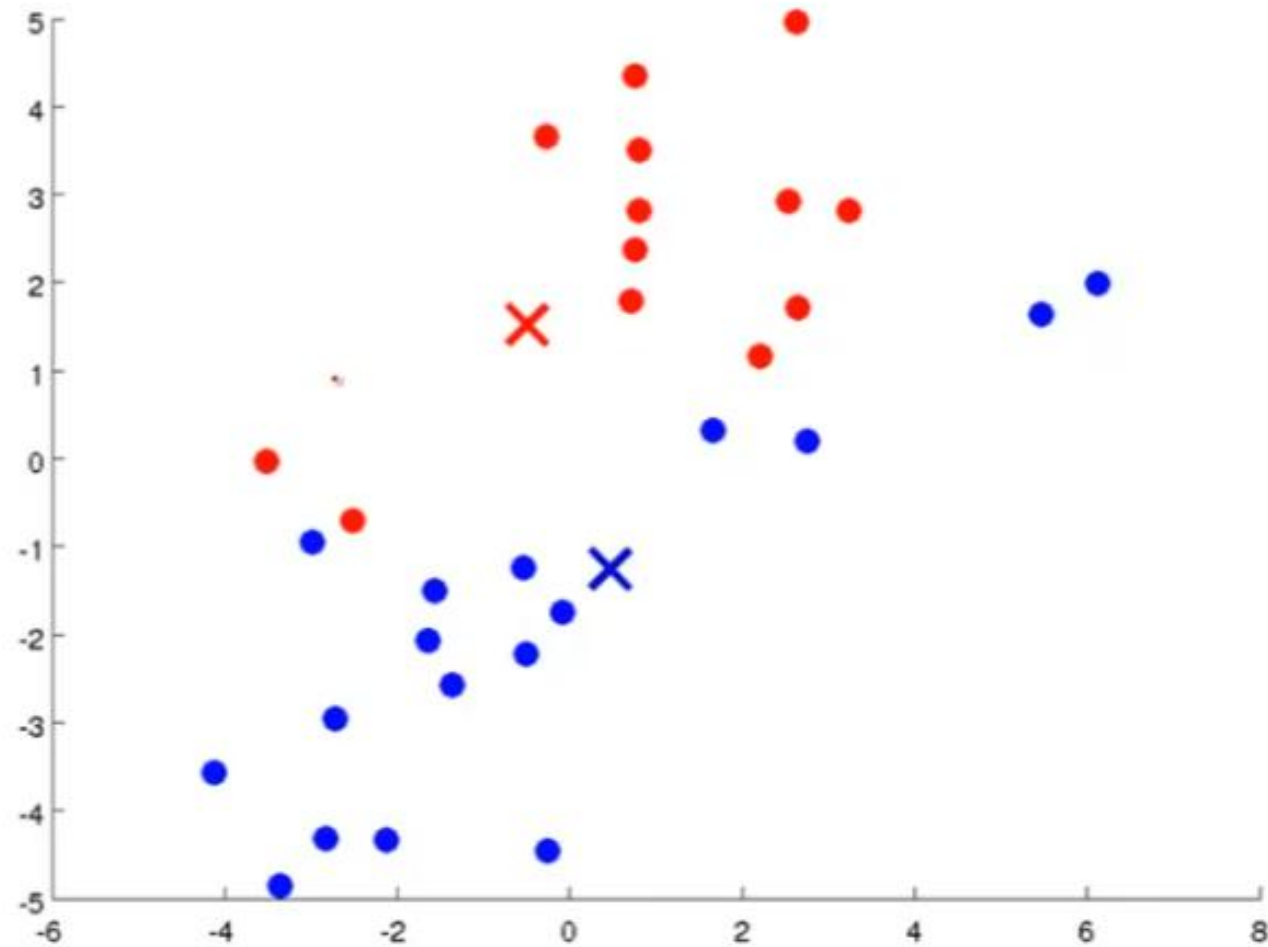


איתחול מרכזים

1. הקצאת האשכולות

2. העברת מרכזים

K-means Algorithm

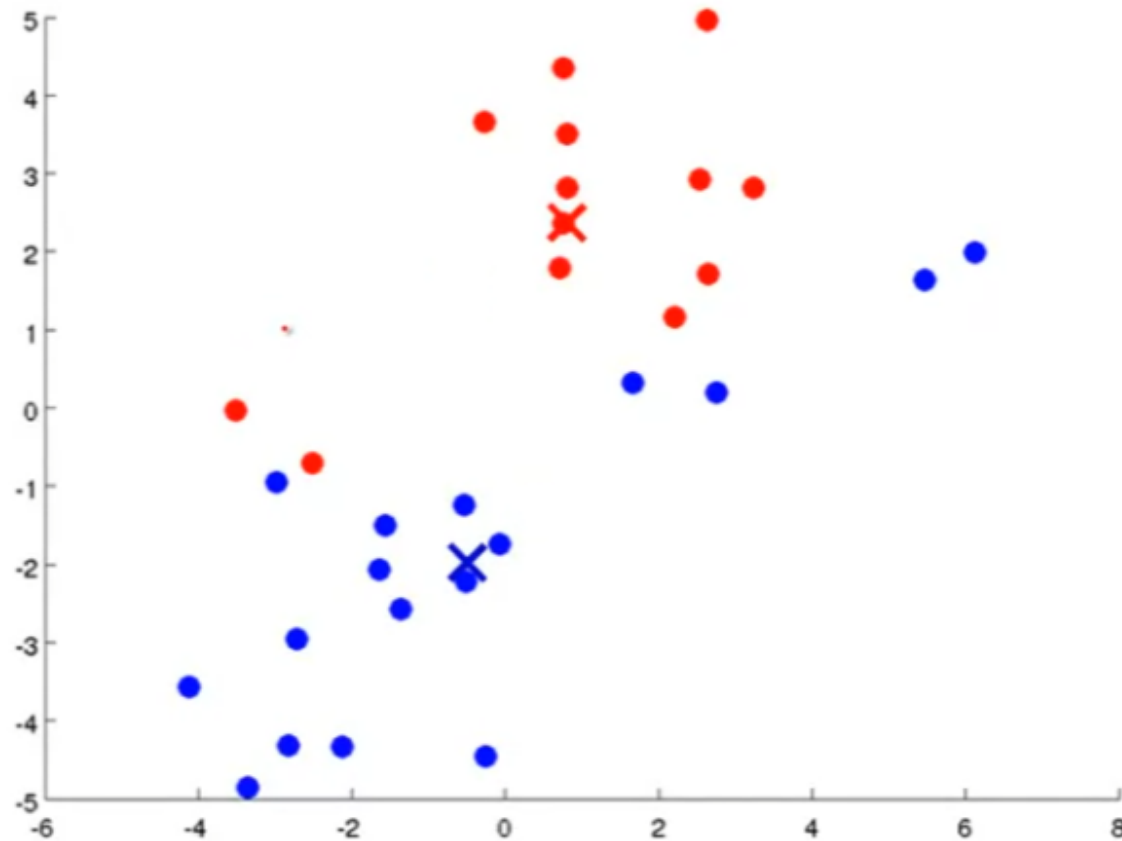


איתחול מרכזים

1. הקצאת האשכולות

2. העברת מרכזים

K-means Algorithm

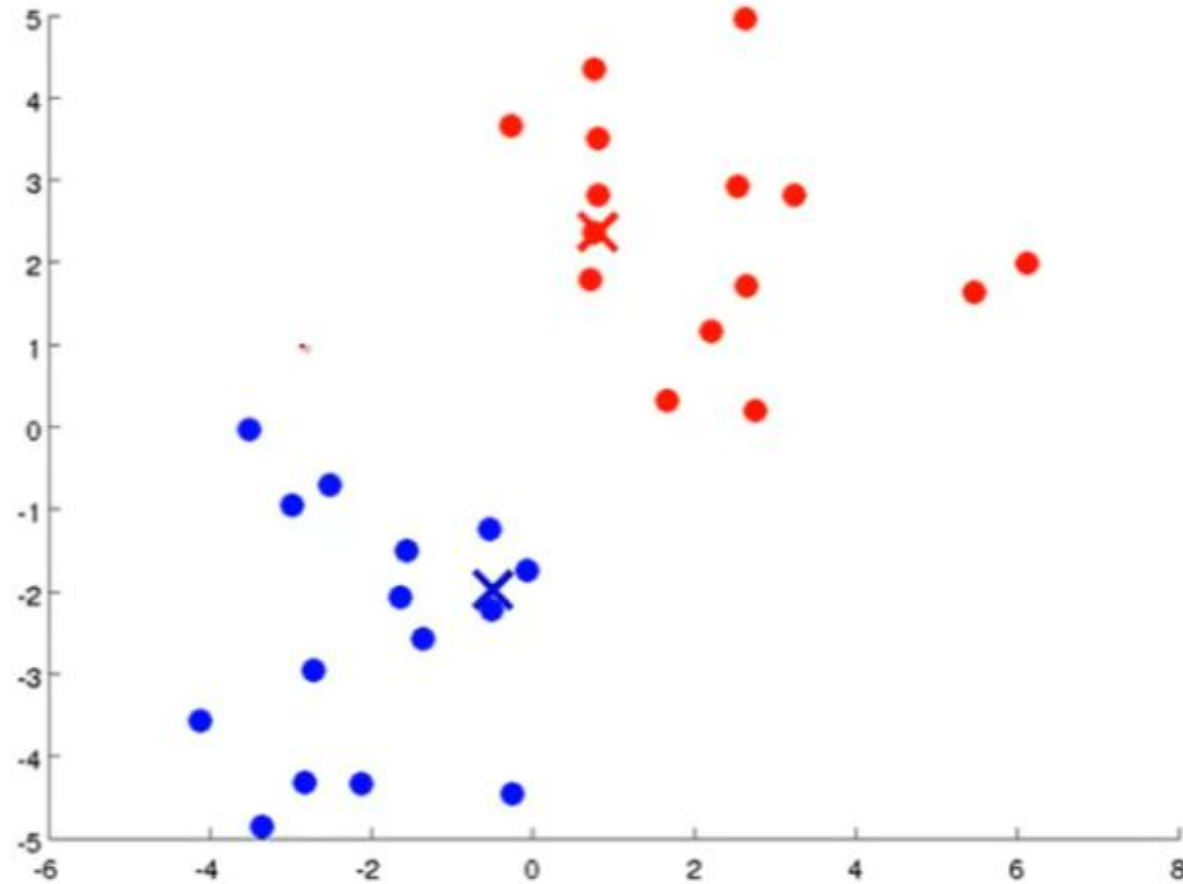


איתחול מרכזים

1. הקצאת האשכולות

2. העברת מרכזים

K-means Algorithm

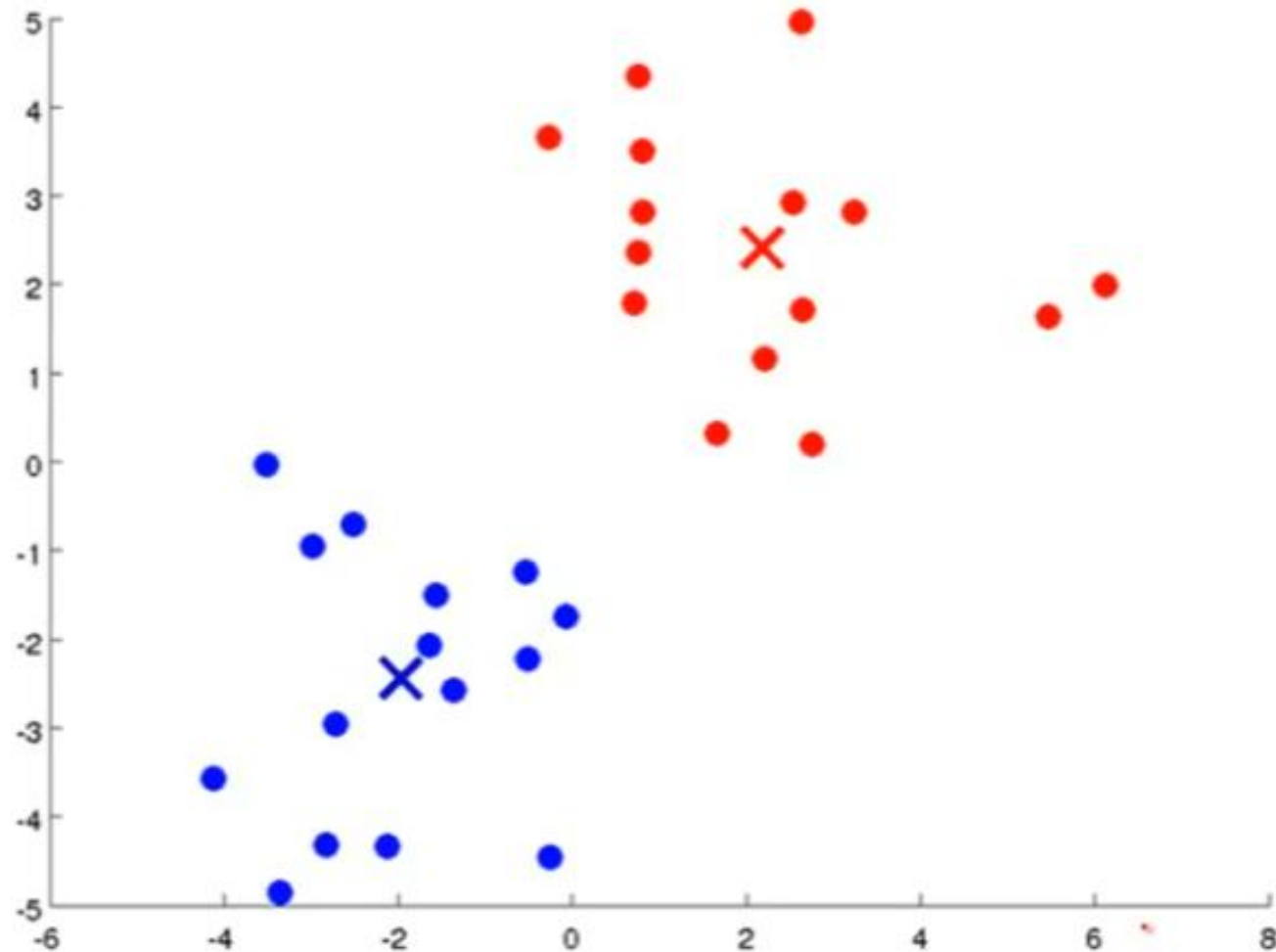


איתחול מרכזים

1. הקצאת האשכולות

2. העברת מרכזים

K-means Algorithm



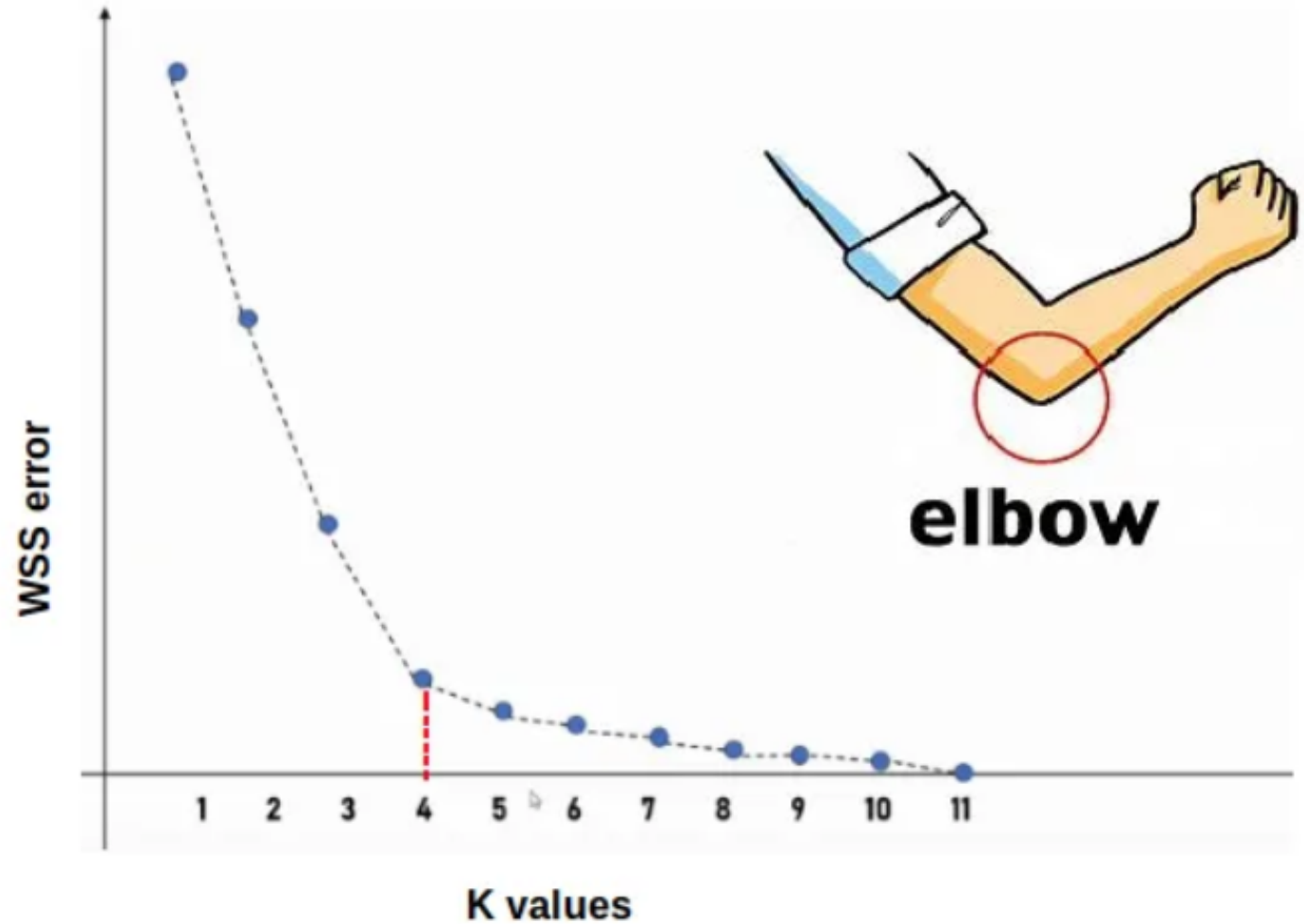
איתחול מרכזים

1. הקצאת האשכולות

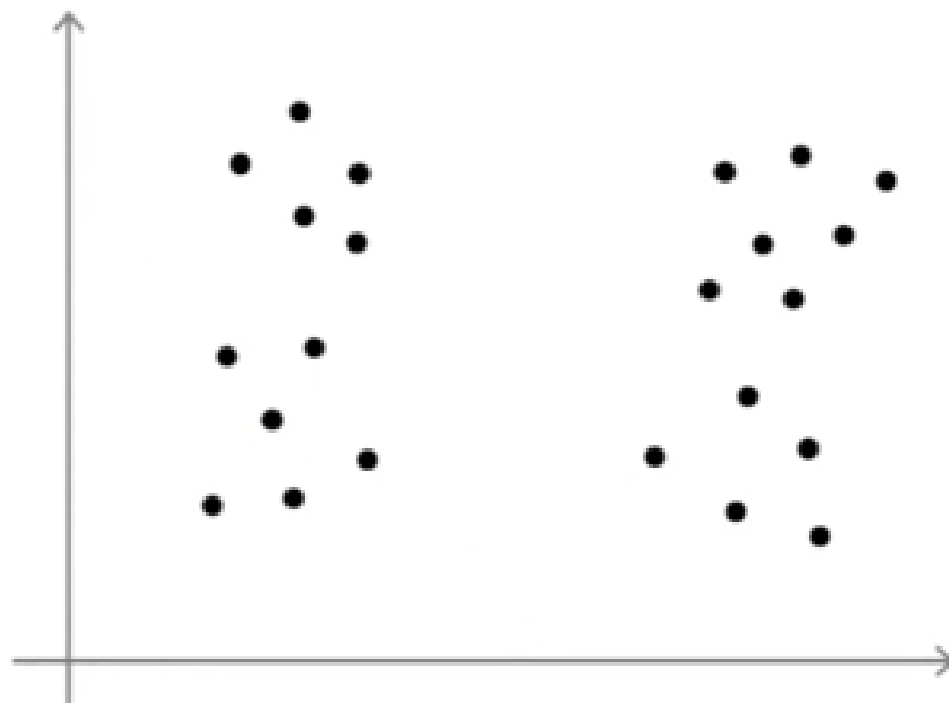
2. העברת מרכזים

How to
choose K
(number of
clusters)?

Elbow method



What is the right value of K?



WCSS = Within-Cluster Sum of Squares

? מה זה WCSS?

WCSS = Within-Cluster Sum of Squares

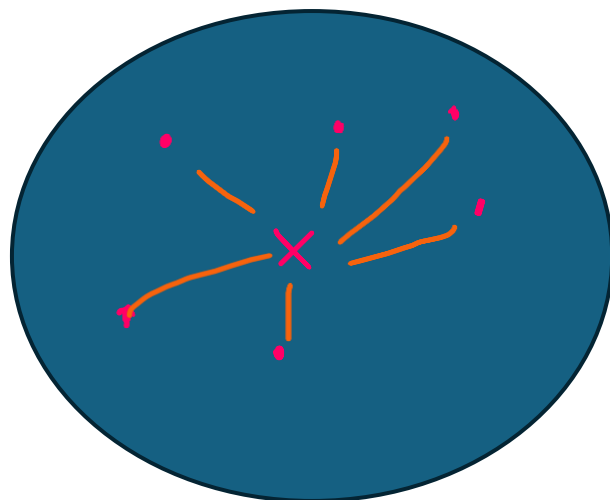
זו דרך למדוד כמה הנקודות דומות למרכז הקלאסטר שלהן.
כלומר: כמה "צפוף" הקלאסטר.

בפשטות: 🧠

תאר לך שיש לך קבוצה של נקודות שכל אחת שייכת לקלאסטר.
לכל נקודה מחשבים המרחק שלה מהמרכז של הקלאסטר, ואז מרבעים את המרחק הזה (כמו בסטטיסטיקה).
בסוף – מחברים את כל המרחקים המרובעים האלה לכל הקלאסטרים יחד – זה ה-WCSS.

למה זה חשוב? ■

- WCSS קטן = הנקודות בקלאסטרים קרובות למרכז → הקלאסטרים טובים.
- WCSS גדול = הנקודות מפוזרות מדי → אולי הקלאסטרים לא מתאימים.

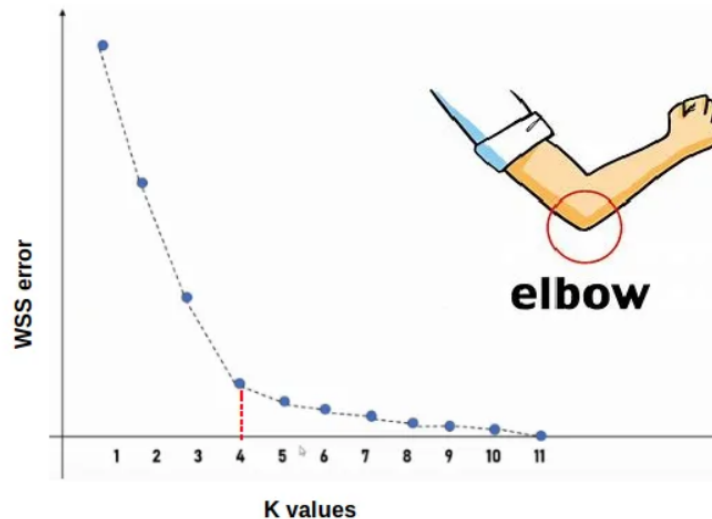


Elbow Method

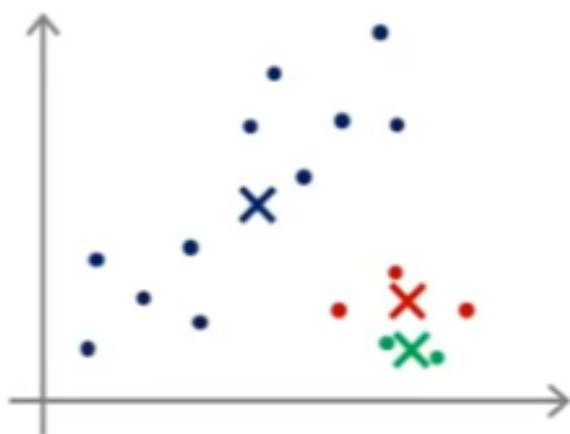
איך זה עובד?

- מחשבים את ה-WCSS (Within-Cluster Sum of Squares) עבור כל k אפשרי (למשל, מ-1 עד 10).
- מציירים גרף של k מול WCSS.
- מחפשים את ה-"נקודת הכר" – הנקודה שבה הקיטון ב-WCSS מתחיל להיות פחות משמעותי (שבירה ברורה בעקומה).

Elbow method



Kmeans++



◆ מה הבעיה ב-KMeans רגיל?

- המרכזים ההתחלתיים נבחרים באקראי.
- עלול להוביל לקלאסטרים גרועים או להתכנסות איטית.
- רגיש לנקודות קצה או מבנה הנתונים.

k-means++ מציע דרך חכמה לבחור את המרכזים ההתחלתיים כך שהאלגוריתם יגיע מהר יותר ולפתרון טוב יותר.
הבחירה מתבצעת כך:

1. בוחרים מרכז ראשון באקראי מבין כל הנקודות.
2. לכל נקודה אחרת מחשבים את המרחק הריבועי למרכז הקרוב ביותר שנבחר עד כה.
3. בוחרים את המרכז הבא בהסתברות פרופורציונלית למרחק הזה – כלומר, נקודות רחוקות ממרכזים קיימים נבחרות בהסתברות גבוהה יותר.
4. חוזרים על זה עד שיש k מרכזים.
5. מפעילים את אלגוריתם k-means הרגיל על סמך מרכזים אלה.

k_means_clustering

