- **Marketing is crucial for the growth and sustainability of any business.**
- **Marketers can help build the company's brand, engage customers, grow revenue, and increase sales.**

- One of the key pain points for marketers is to know their customers and identify their needs.
- By understanding the customer, marketers can launch a targeted marketing campaign that is tailored for specific needs.
- If data about the customers is available, data science can be applied to perform market segmentation.

- In this case study, you have been hired as a consultant to a bank in New York City.
- The bank has extensive data on their customers for the past 6 months.
- The marketing team at the bank wants to launch a targeted ad marketing campaign by dividing their customers into at least 3 distinctive groups.

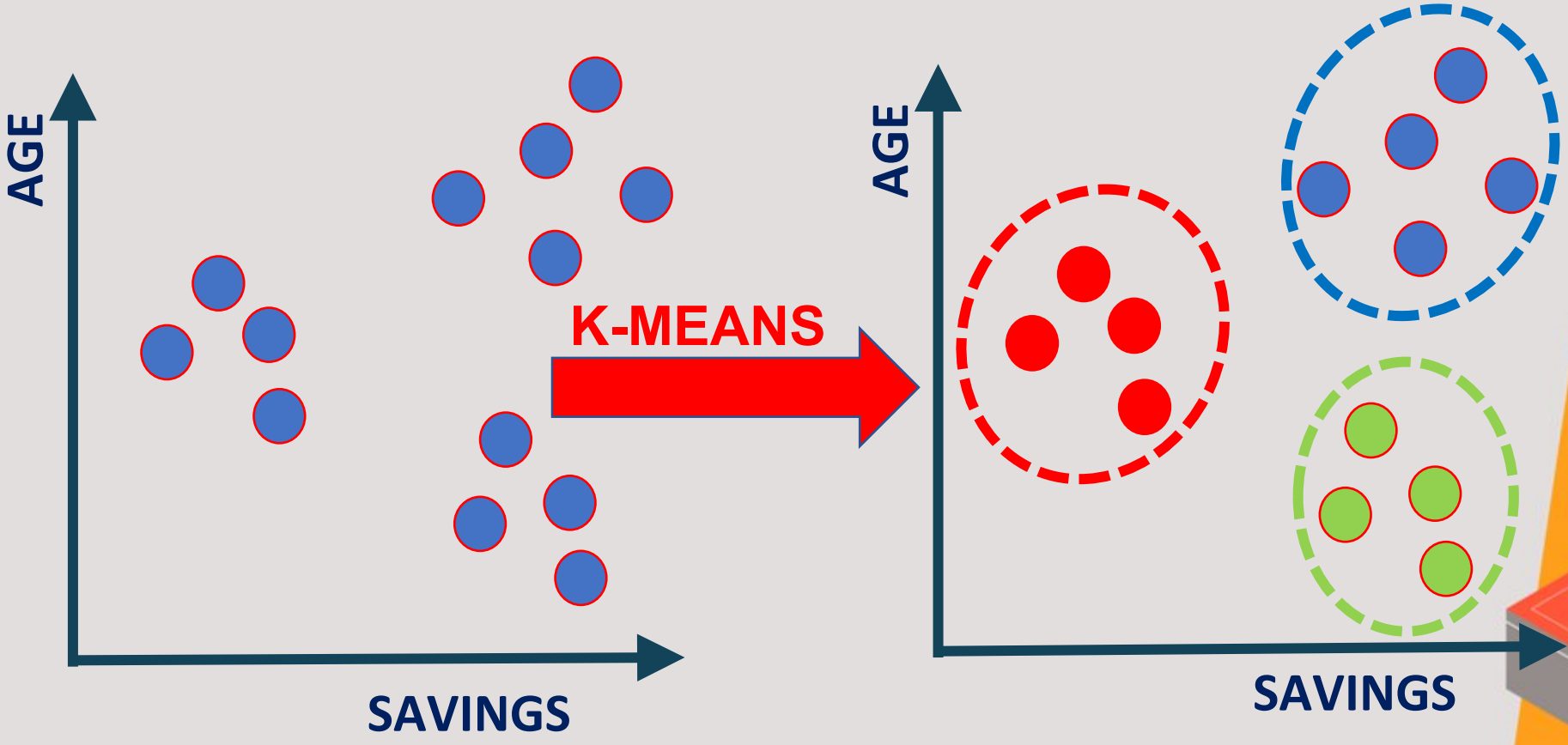
- # CUSTID: Identification of Credit Card holder
- # BALANCE: Balance amount left in customer's account to make purchases
- # BALANCE_FREQUENCY: How frequently the Balance is updated, score between 0 and 1 (1 = frequently updated, 0 = not frequently updated)
- # PURCHASES: Amount of purchases made from account
- # ONEOFFPURCHASES: Maximum purchase amount done in one-go
- # INSTALLMENTS_PURCHASES: Amount of purchase done in installment
- # CASH_ADVANCE: Cash in advance given by the user
- # PURCHASES_FREQUENCY: How frequently the Purchases are being made, score between 0 and 1 (1 = frequently purchased, 0 = not frequently purchased)

- # PURCHASES_FREQUENCY: How frequently the Purchases are being made, score between 0 and 1 (1 = frequently purchased, 0 = not frequently purchased)
- # ONEOFF_PURCHASES_FREQUENCY: How frequently Purchases are happening in one-go (1 = frequently purchased, 0 = not frequently purchased)
- # PURCHASES_INSTALLMENTS_FREQUENCY: How frequently purchases in installments are being done (1 = frequently done, 0 = not frequently done)
- # CASH_ADVANCE_FREQUENCY: How frequently the cash in advance being paid

- # CASH_ADVANCE_TRX: Number of Transactions made with "Cash in Advance"
- # PURCHASES_TRX: Number of purchase transactions made
- # CREDIT_LIMIT: Limit of Credit Card for user
- # PAYMENTS: Amount of Payment done by user
- # MINIMUM_PAYMENTS: Minimum amount of payments made by user
- # PRC_FULL_PAYMENT: Percent of full payment paid by user
- # TENURE: Tenure of credit card service for user

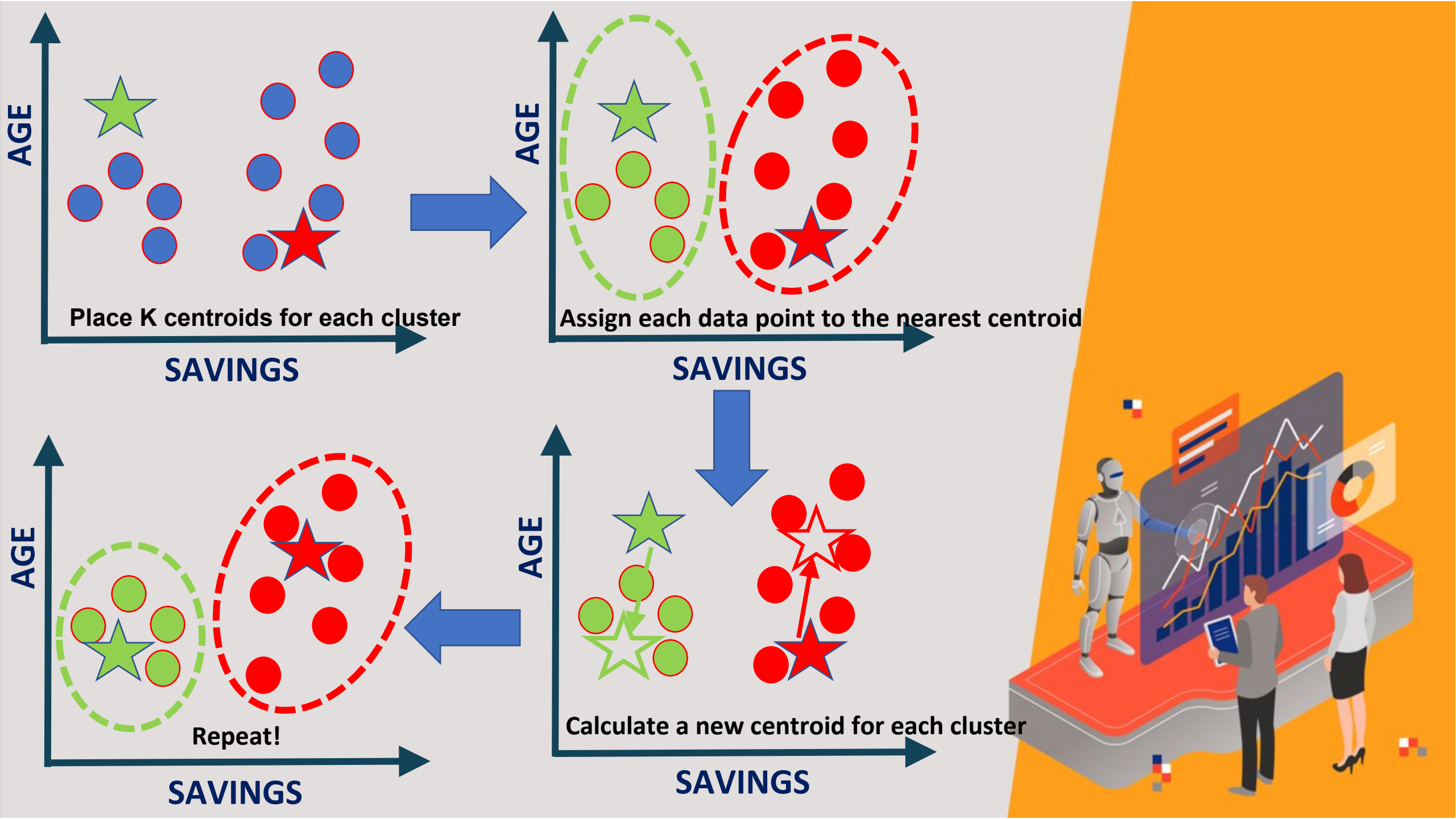
K-MEANS INTUITION

- K-means is an unsupervised learning algorithm (clustering).
- K-means works by grouping some data points together (clustering) in an unsupervised fashion.
- The algorithm groups observations with similar attribute values together by measuring the Euclidian distance between points.



K-MEANS ALGORITHM STEPS

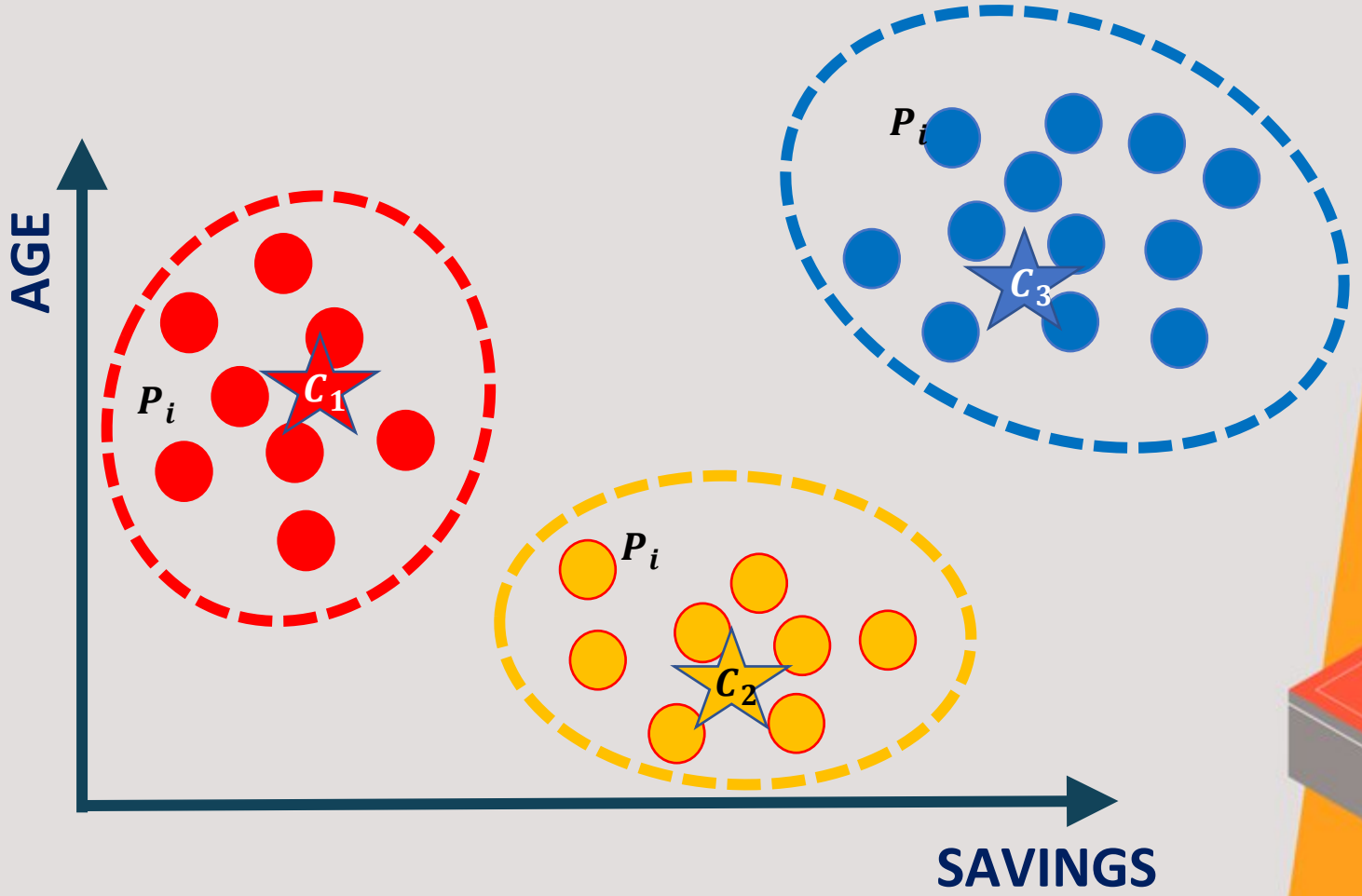
1. Choose number of clusters “K”
2. Select random K points that are going to be the centroids for each cluster
3. Assign each data point to the nearest centroid, doing so will enable us to create “K” number of clusters
4. Calculate a new centroid for each cluster
5. Reassign each data point to the new closest centroid
6. Go to step 4 and repeat.


AGE
SAVINGS
Place K centroids for each cluster
AGE
SAVINGS
Assign each data point to the nearest centroid
AGE
SAVINGS
Calculate a new centroid for each cluster
AGE
SAVINGS
Repeat!

HOW TO SELECT THE OPTIMAL NUMBER OF CLUSTERS (K)? “ELBOW METHOD”

$$Within\ Cluster\ Sum\ of\ Squares\ (WCSS)$$

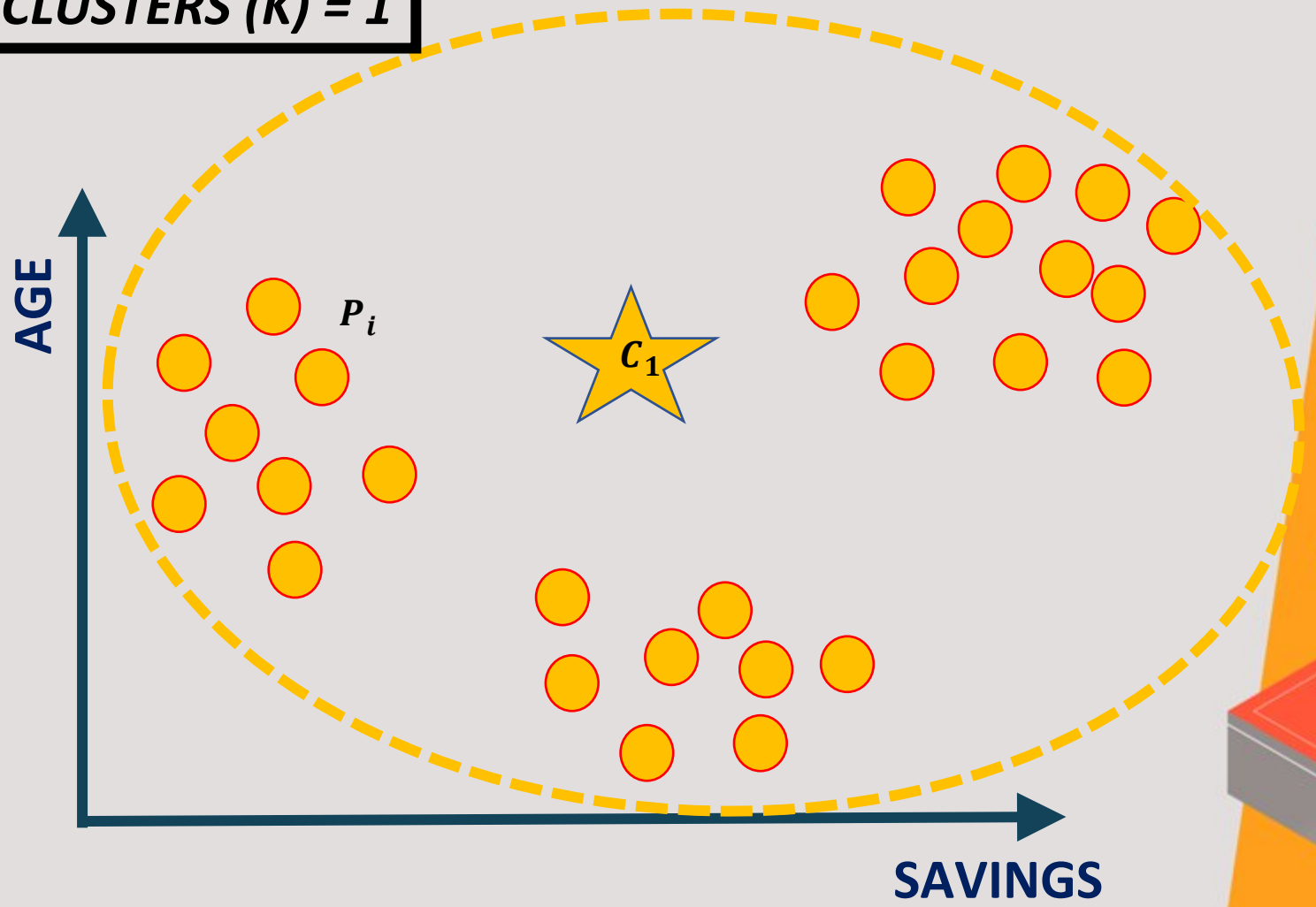
$$= \sum_{P_i\ in\ Cluster\ 1} distance\,(P_i, C_1)^2 + \sum_{P_i\ in\ Cluster\ 2} distance\,(P_i, C_2)^2 + \sum_{P_i\ in\ Cluster\ 3} distance\,(P_i, C_3)^2$$



HOW TO SELECT THE OPTIMAL NUMBER OF CLUSTERS (K)? “ELBOW METHOD”

$$Within\ Cluster\ Sum\ of\ Squares\ (WCSS) = \sum_{P_i\ in\ Cluster\ 1} distance(P_i, C_1)^2$$

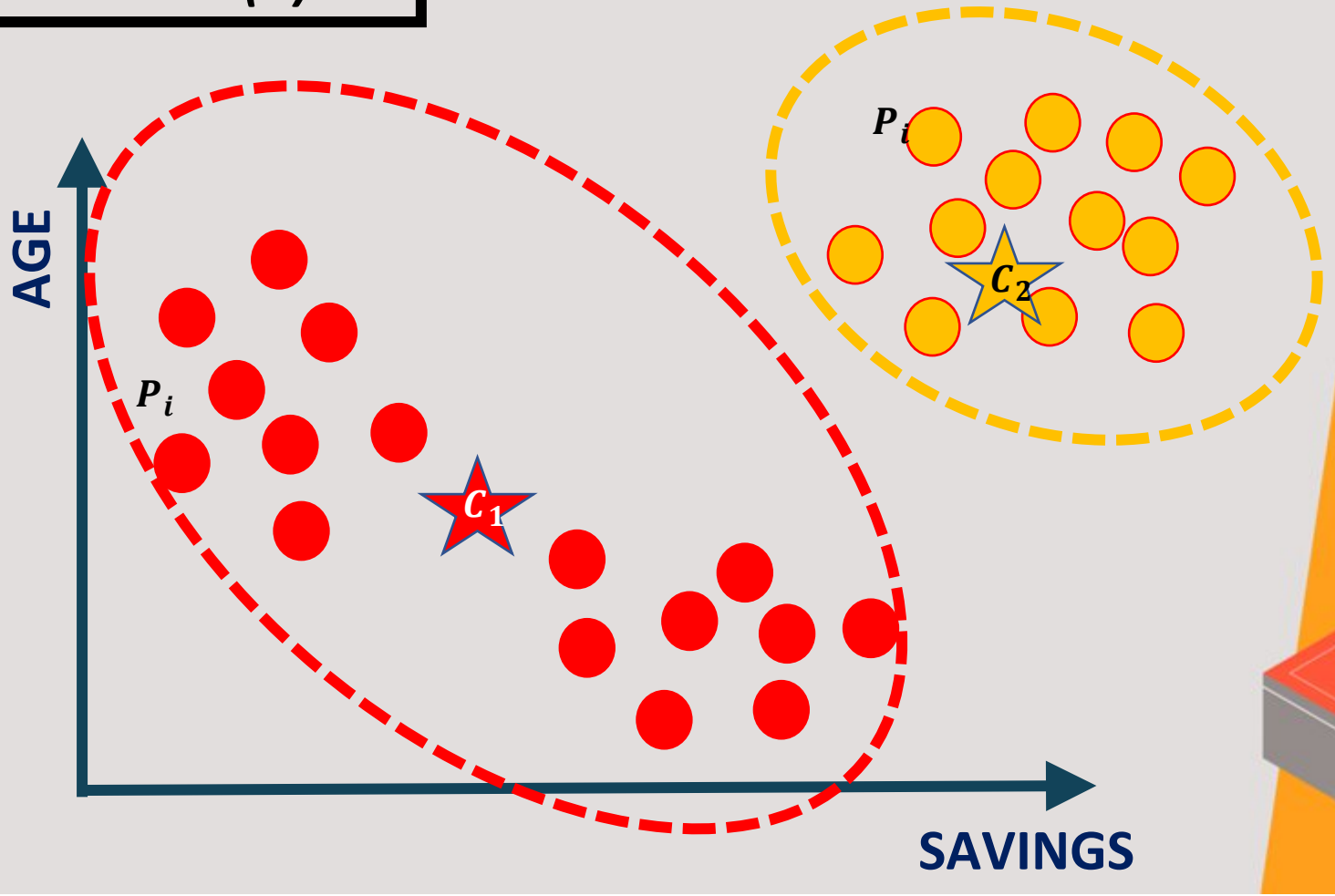
NUMBER OF CLUSTERS (K) = 1



HOW TO SELECT THE OPTIMAL NUMBER OF CLUSTERS (K)? "ELBOW METHOD"

$$Within\ Cluster\ Sum\ of\ Squares\ (WCSS) = \sum_{P_i\ in\ Cluster\ 1} distance\,(P_i, C_1)^2 + \sum_{P_i\ in\ Cluster\ 2} distance\,(P_i, C_2)^2$$

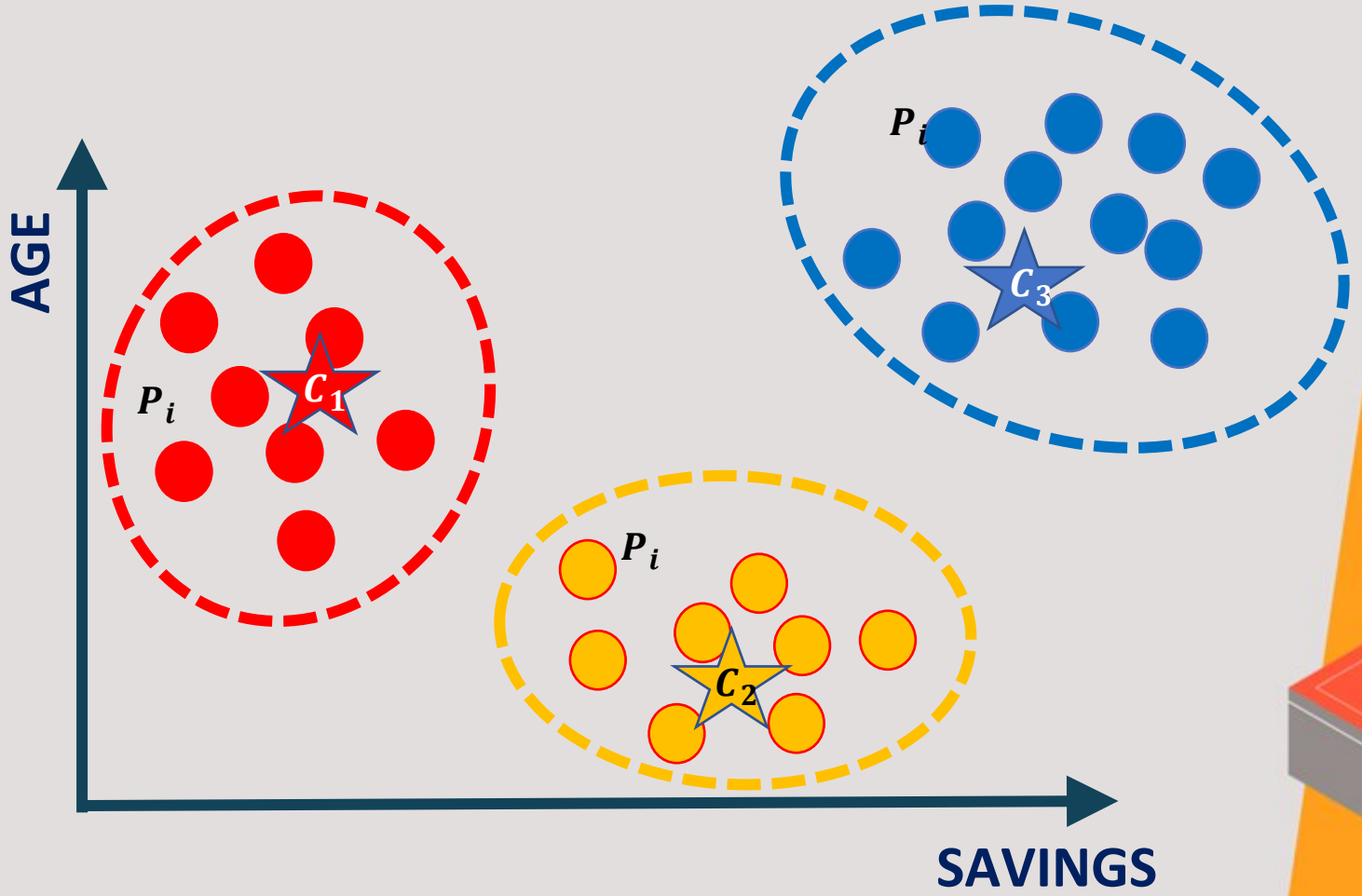
NUMBER OF CLUSTERS (K) = 2



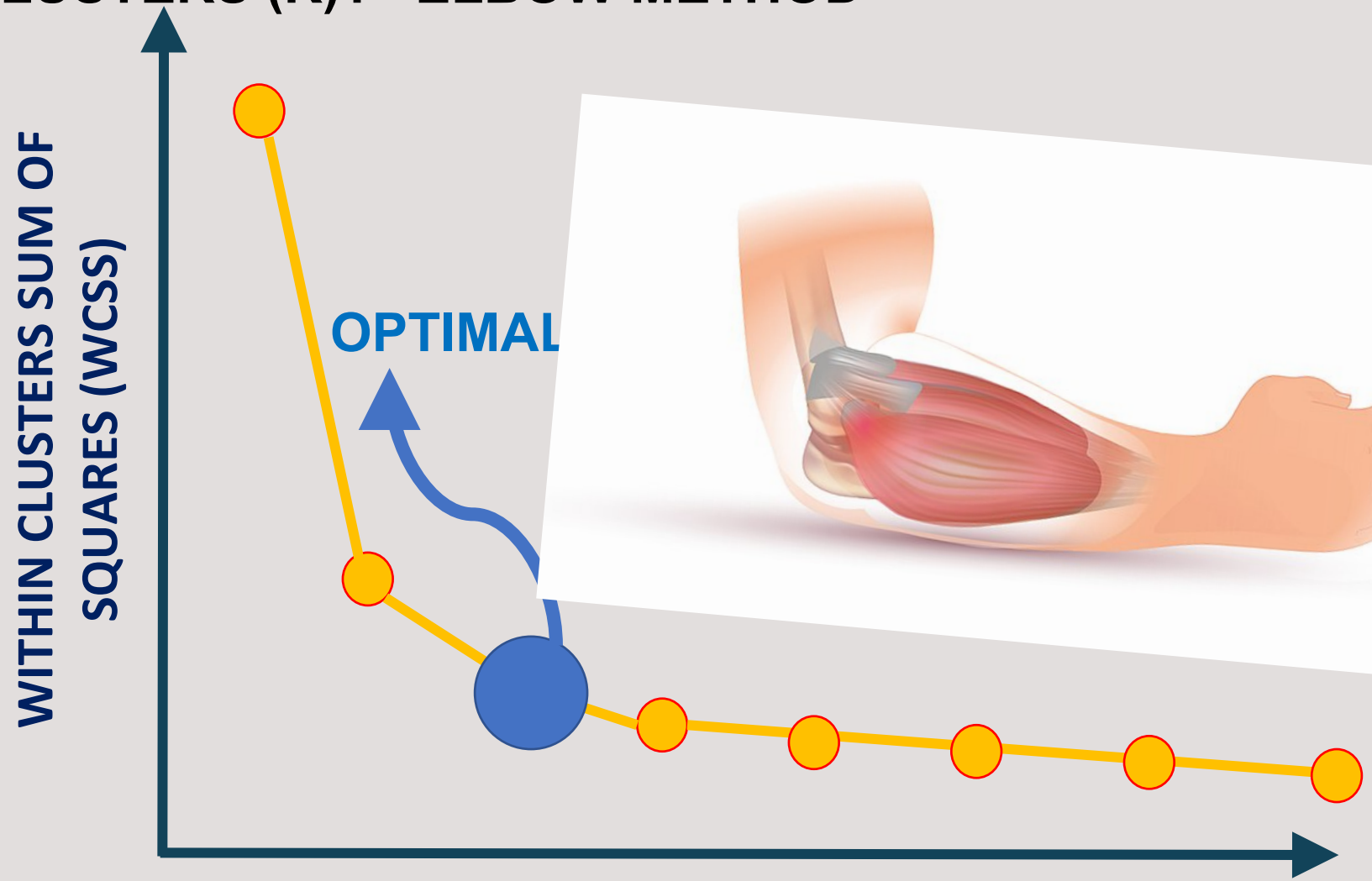
HOW TO SELECT THE OPTIMAL NUMBER OF CLUSTERS (K)? “ELBOW METHOD”

$$Within\ Cluster\ Sum\ of\ Squares\ (WCSS)$$

$$= \sum_{P_i\ in\ Cluster\ 1} distance\,(P_i, C_1)^2 + \sum_{P_i\ in\ Cluster\ 2} distance\,(P_i, C_2)^2 + \sum_{P_i\ in\ Cluster\ 3} distance\,(P_i, C_3)^2$$



HOW TO SELECT THE OPTIMAL NUMBER OF CLUSTERS (K)? “ELBOW METHOD”



Source: https://commons.wikimedia.org/wiki/File:Tennis_Elbow_Illustration.jpg

AUTOENCODERS INTUITION

- Auto encoders are a type of Artificial Neural Networks that are used to perform a task of data encoding (representation learning).
- Auto encoders use the same input data for the input and output, Sounds crazy right!?

AUTOENCODERS UNSUPERVISED TRAINING

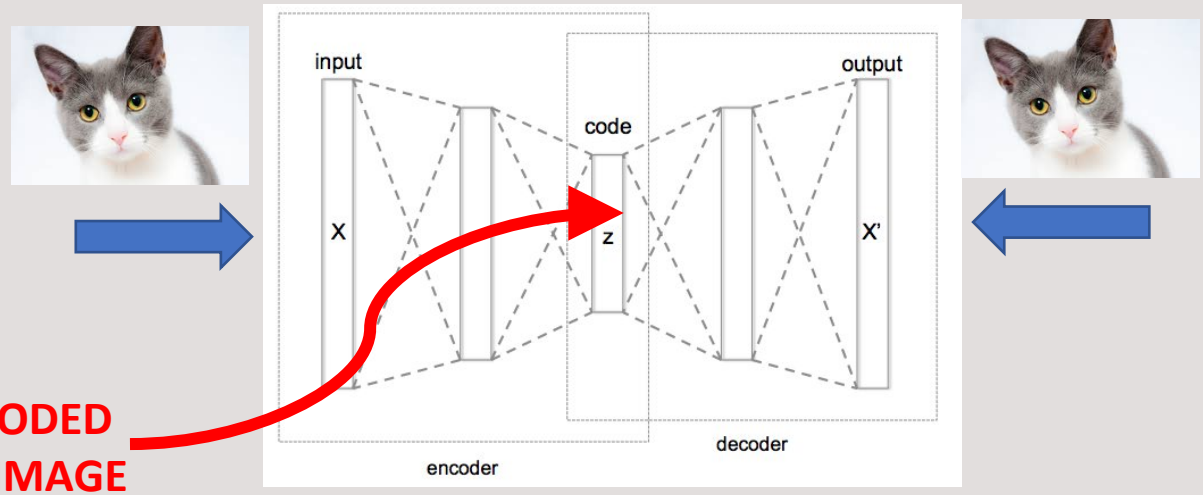


Photo Credit: https://commons.wikimedia.org/wiki/File:Autoencoder_structure.png
Photo Credit: https://commons.wikimedia.org/wiki/File:Artificial_neural_network_image_recognition.png
Photo Credit: https://www.pexels.com/photo/grey-and-white-short-fur-cat-104827/

THE CODE LAYER

- Auto encoders work by adding a bottleneck in the network.
- This bottleneck forces the network to create a compressed (encoded) version of the original input
- Auto encoders work well if correlations exists between input data (performs poorly if the all input data is independent)

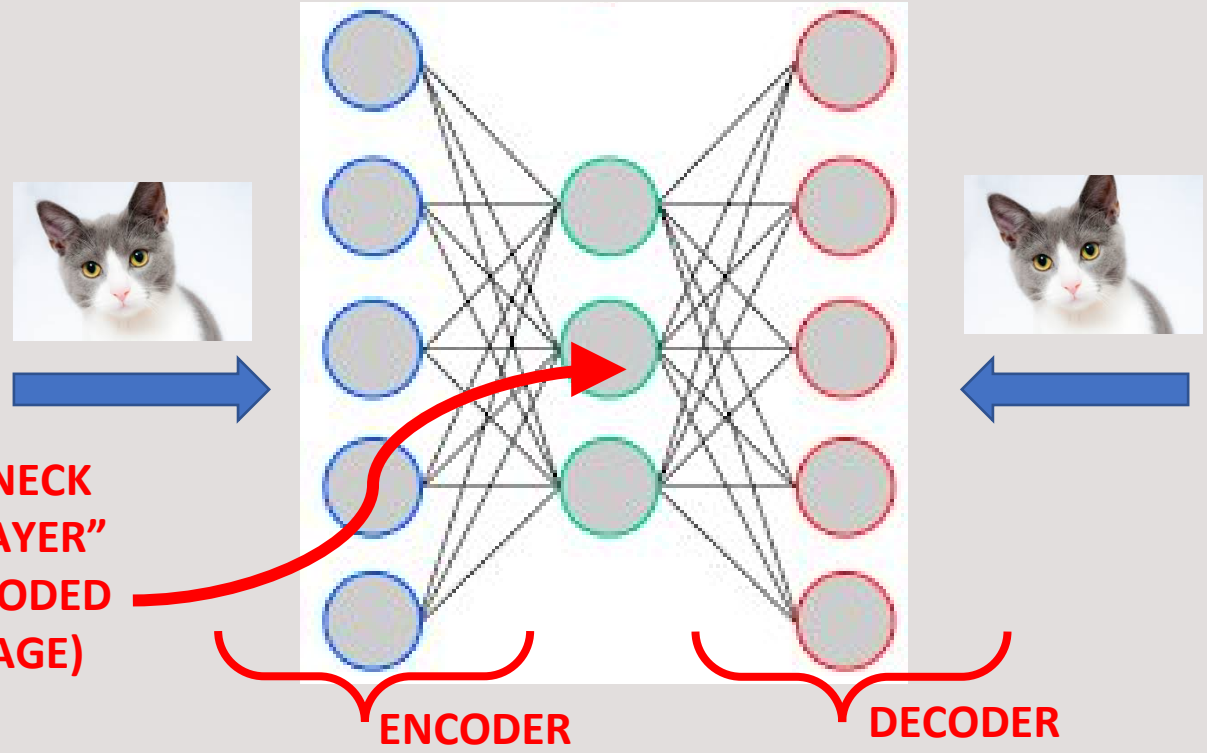
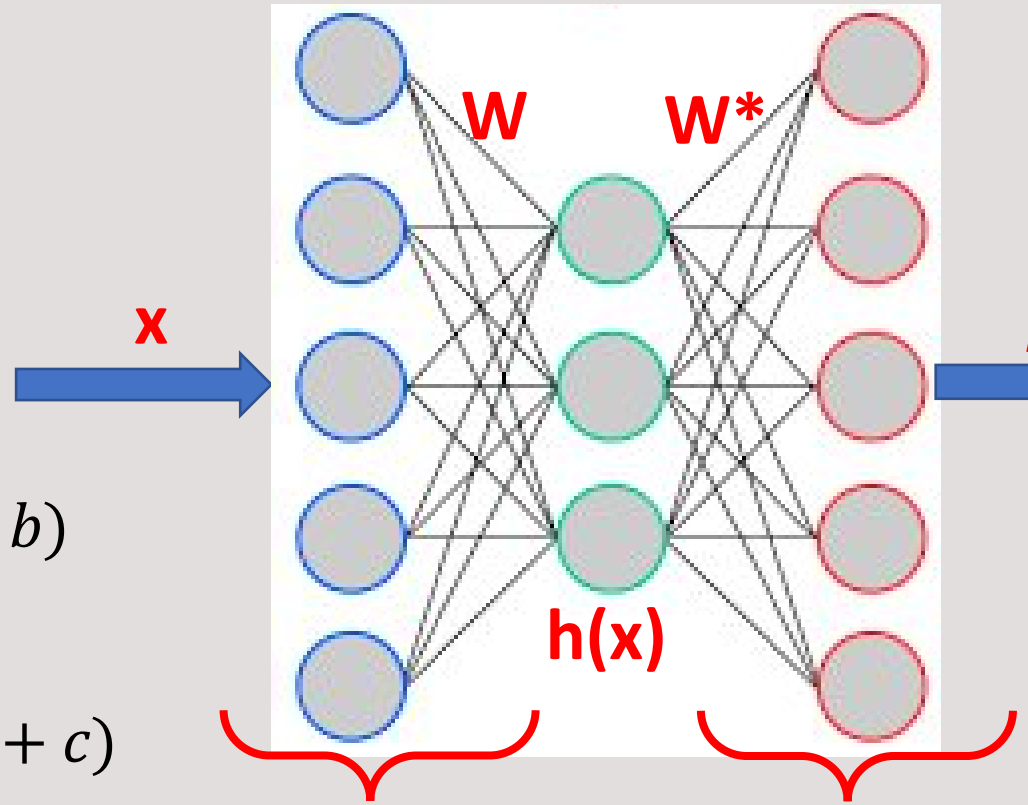


Photo Credit: https://commons.wikimedia.org/wiki/File:Autoencoder_structure.png
Photo Credit: https://commons.wikimedia.org/wiki/File:Artificial_neural_network_image_recognition.png
Photo Credit: https://www.pexels.com/photo/grey-and-white-short-fur-cat-104827/

AUTOENCODER MATH



ENCODER:

$$h(x) = sigmoid(W*x + b)$$

DECODER:

$$\hat{x} = sigmoid(W^**h(x) + c)$$

TIED WEIGHTS:

Weights from input to hidden layer will be equal to the weights from hidden layer to output

$$W^* = W^T$$

Photo Credit: https://commons.wikimedia.org/wiki/File:Autoencoder_structure.png
Photo Credit: https://commons.wikimedia.org/wiki/File:Artificial_neural_network_image_recognition.png
Photo Credit: https://www.pexels.com/photo/grey-and-white-short-fur-cat-104827/

PRINCIPAL COMPONENT ANALYSIS: OVERVIEW

- PCA is an unsupervised machine learning algorithm.
- PCA performs dimensionality reductions while attempting at keeping the original information unchanged.
- PCA works by trying to find a new set of features called components.
- Components are composites of the uncorrelated given input features.

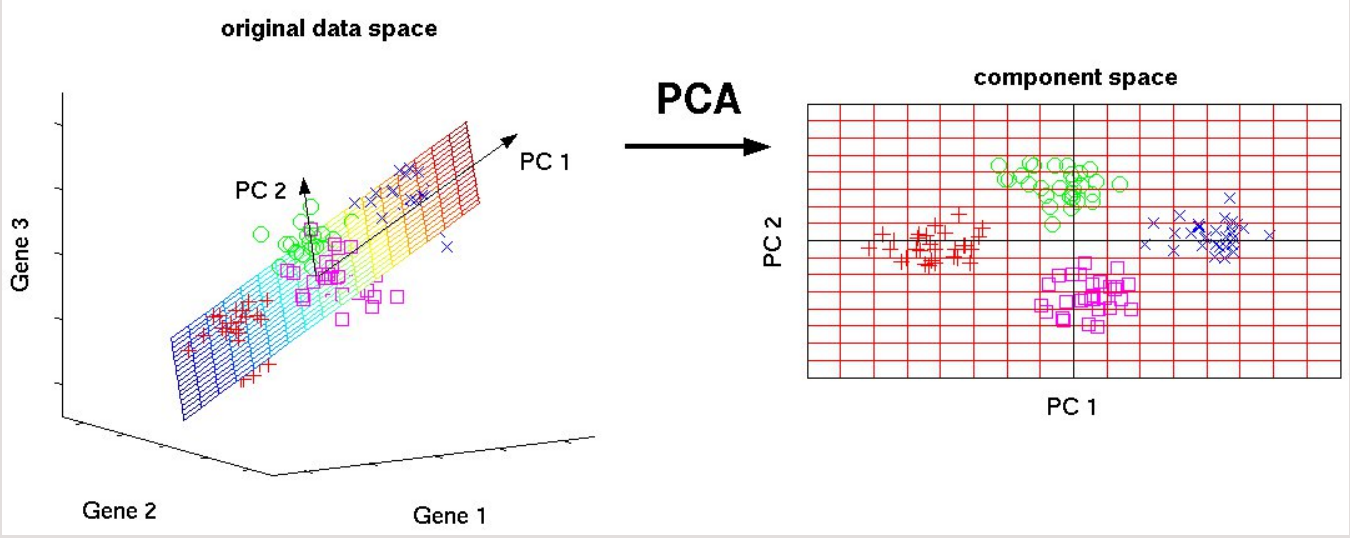


Photo Credit: http://phdthesis-bioinformatics-maxplanckinstitute-molecularplantphys.matthias-scholz.de/