

- Marketing is crucial for the growth and sustainability of retail 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 and AI/ML can be applied to perform market segmentation.
- In this case study, you have been hired as an expert data scientist for a retail analytics company in Seattle, U.S..
- The company has extensive data on their customers for the duration of 2.5 years.
- You have been tasked to create targeted ad marketing campaign by dividing their customers into at least 3 distinctive groups.



Source: <https://www.needpix.com/photo/896541/analytics-data-analytics-graph-chart-analysis-business-data-statistics-analyzing>



- **# ORDERNUMER:** Identification of order placed
- **# QUANTITYORDERED:** Number of items ordered
- **# PRICEEACH:** Price of each item
- **# SALES:** Total amount of sales
- **# ORDERDATE:** Date in which order is placed
- **# STATUS:** Status of the order
- **# QTR\_ID:** Quarter in which order is placed
- **# MONTH\_ID:** Month in which order is placed
- **# YEAR\_ID :** Year in which order is placed
- **# PRODUCTLINE:** Product category
- **# CUSTOMERNAME:** Name of the customer
- **# PHONE:** Phone number

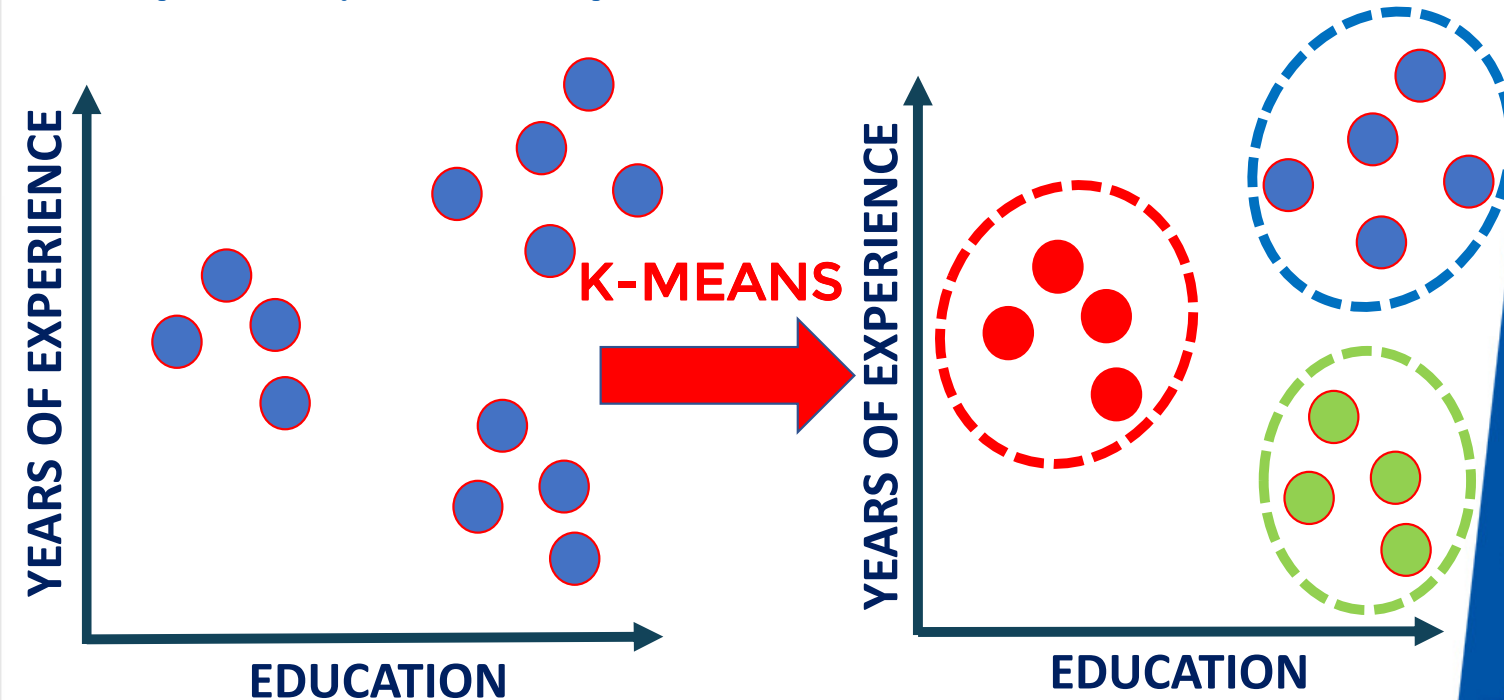


- **# ADDRESSLINE1:** Address to be shipped
- **# ADDRESSLINE2:** Address to be shipped
- **# CITY:** City in which customer resides
- **# STATE:** State in which customer resides
- **# POSTALCODE:** Postal code in which customer resides
- **# COUNTRY:** Country in which customer resides
- **# TERRITORY:** Territory in which customer resides
- **# DEALSIZE:** Size of the order
- **# CONTACTFIRST NAME:** Contact person's first name
- **# CONTACTLAST NAME:** Contact person's last name



## 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.

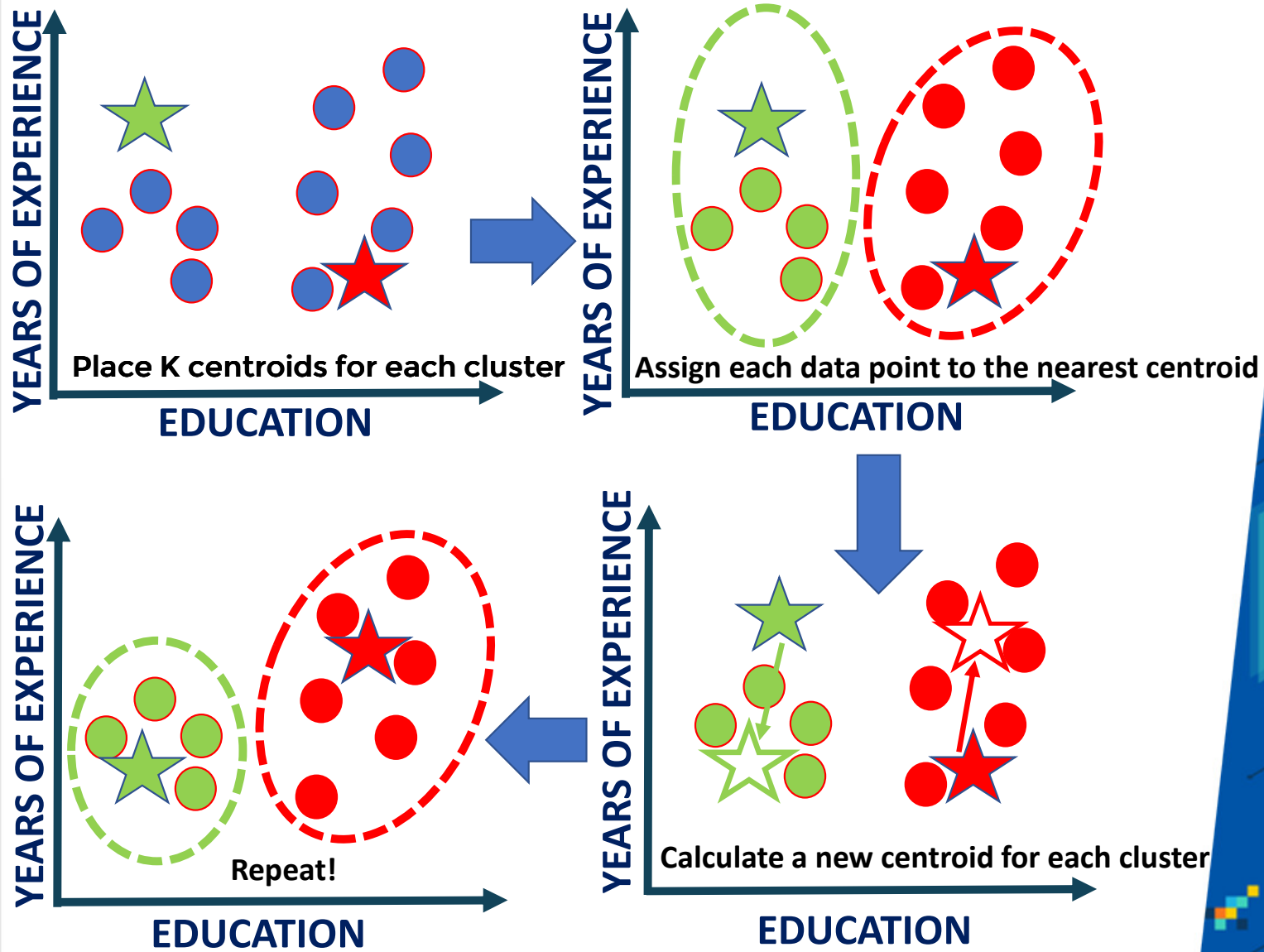


## K-MEANS ALGORITHM STEPS

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.



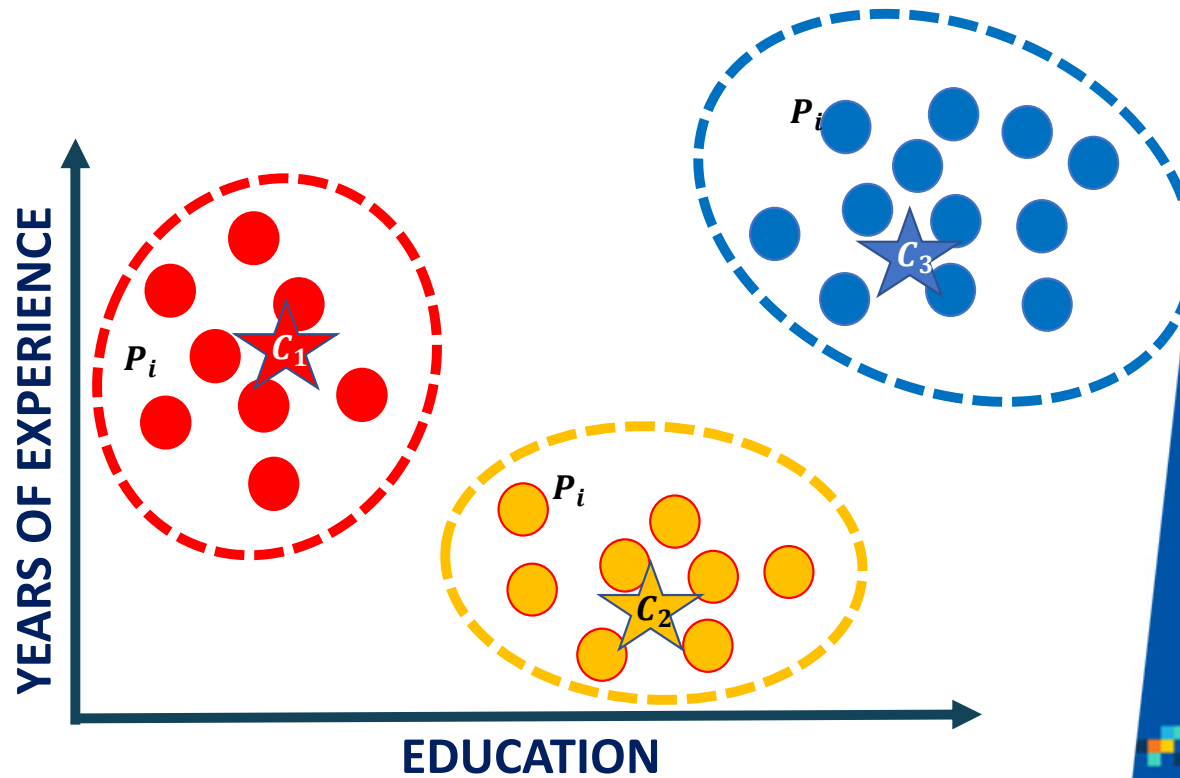




## HOW TO SELECT THE OPTIMAL NUMBER OF CLUSTERS (K)? “ELBOW METHOD”

*Within Cluster Sum of Squares (WCSS)*

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

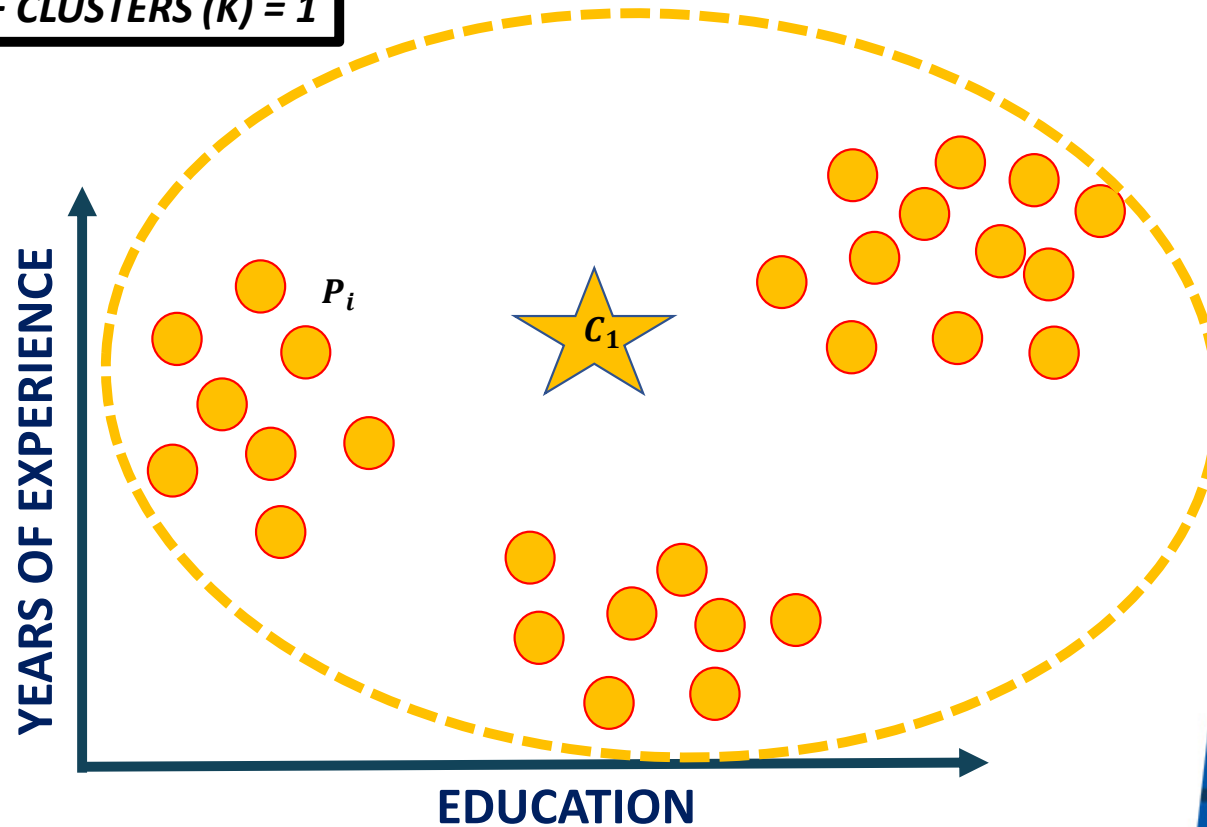




## HOW TO SELECT THE OPTIMAL NUMBER OF CLUSTERS (K)? “ELBOW METHOD”

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

**NUMBER OF CLUSTERS (K) = 1**

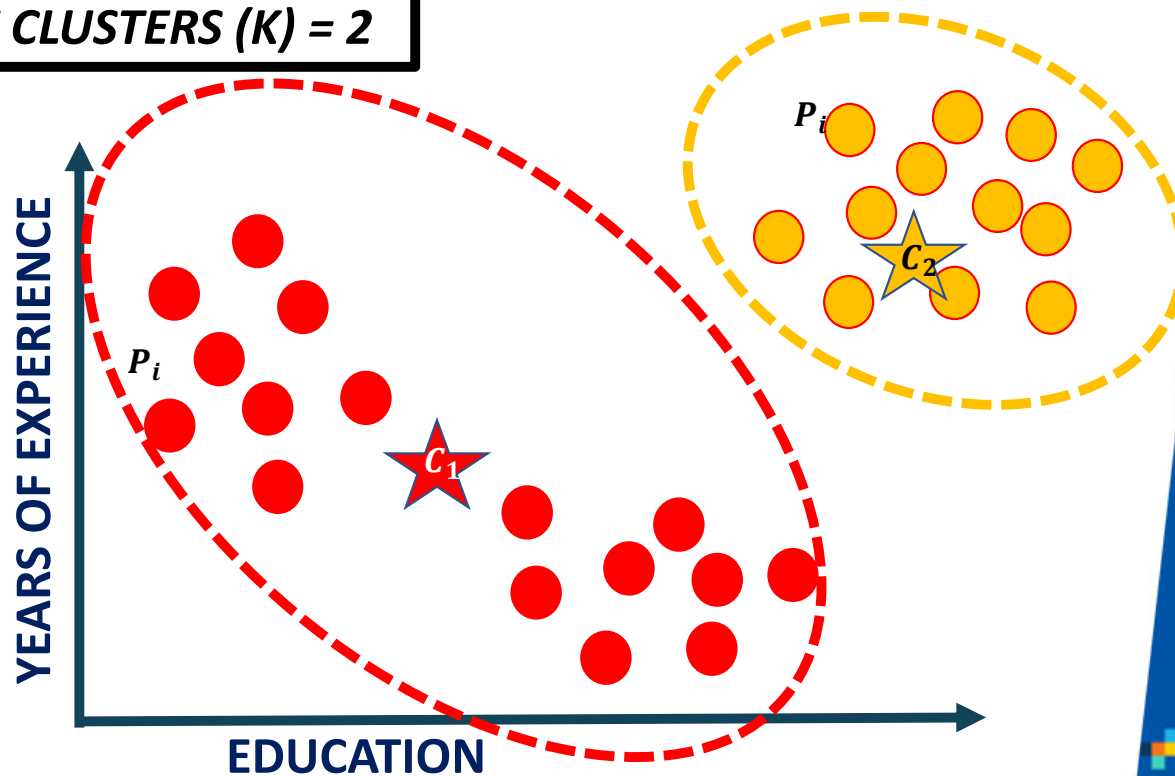


## HOW TO SELECT THE OPTIMAL NUMBER OF CLUSTERS (K)? “ELBOW METHOD”

*Within Cluster Sum of Squares (WCSS)*

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

**NUMBER OF CLUSTERS (K) = 2**

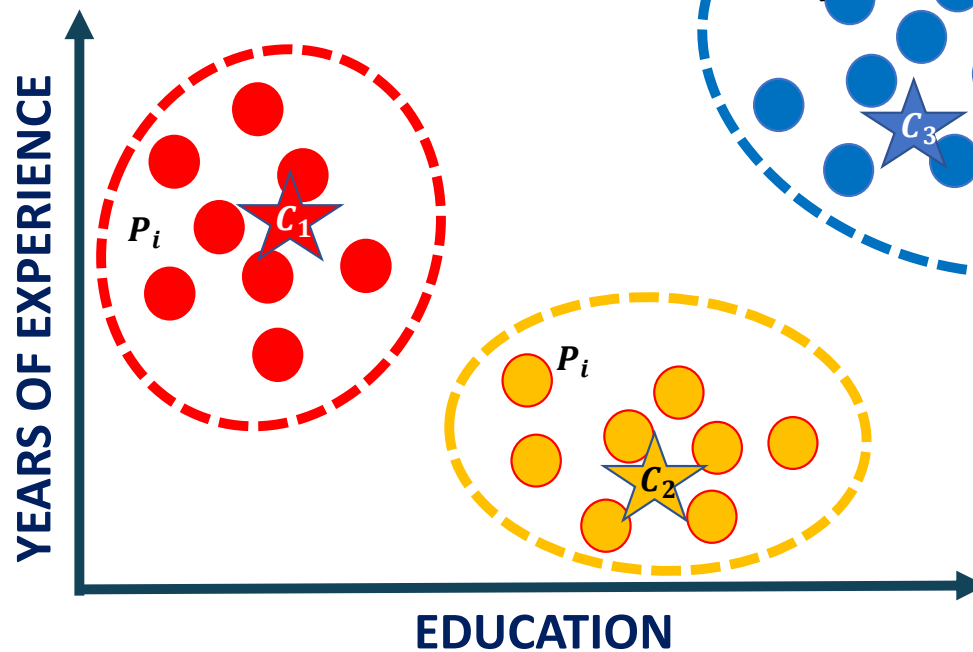


## HOW TO SELECT THE OPTIMAL NUMBER OF CLUSTERS (K)? “ELBOW METHOD”

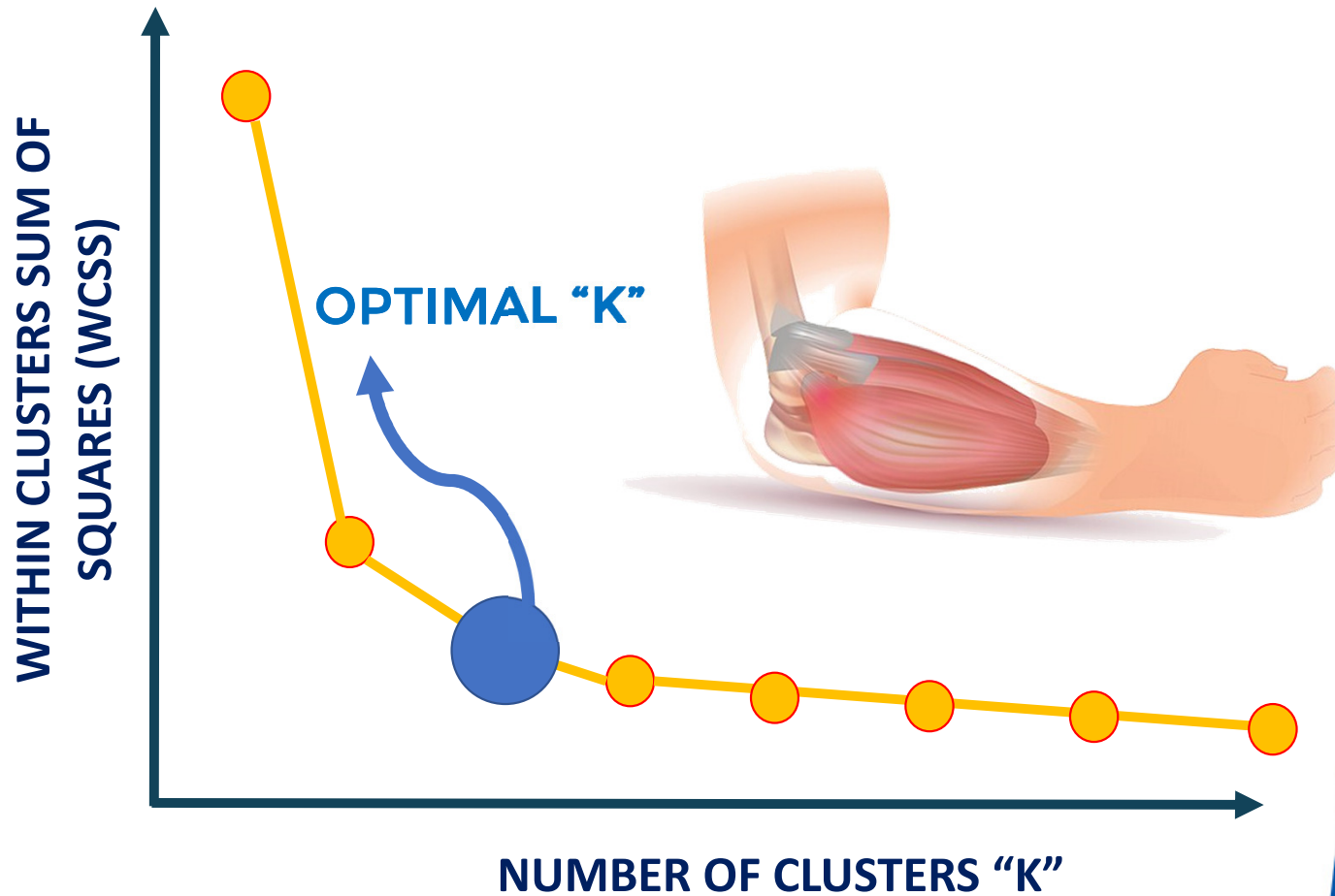
Within Cluster Sum of Squares (WCSS)

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

**NUMBER OF CLUSTERS (K) = 3**



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



Source: [https://commons.wikimedia.org/wiki/File:Tennis\\_Elbow\\_Illustration.jpg](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!?

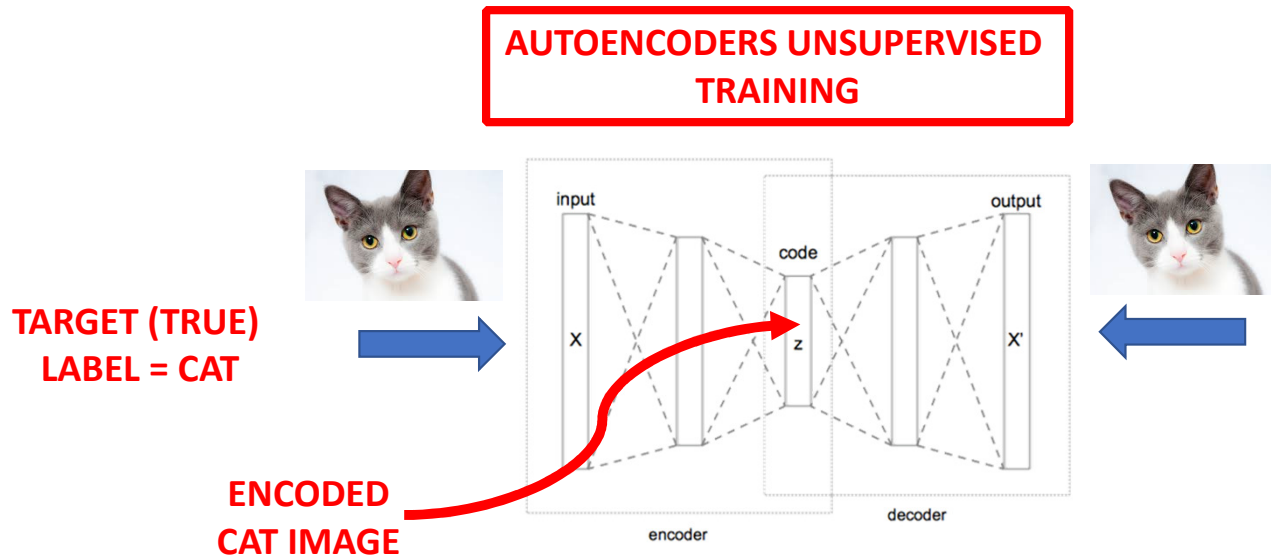


Photo Credit: [https://commons.wikimedia.org/wiki/File:Autoencoder\\_structure.png](https://commons.wikimedia.org/wiki/File:Autoencoder_structure.png)

Photo Credit: [https://commons.wikimedia.org/wiki/File:Artificial\\_neural\\_network\\_image\\_recognition.png](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 exist between input data (performs poorly if the all input data is independent)

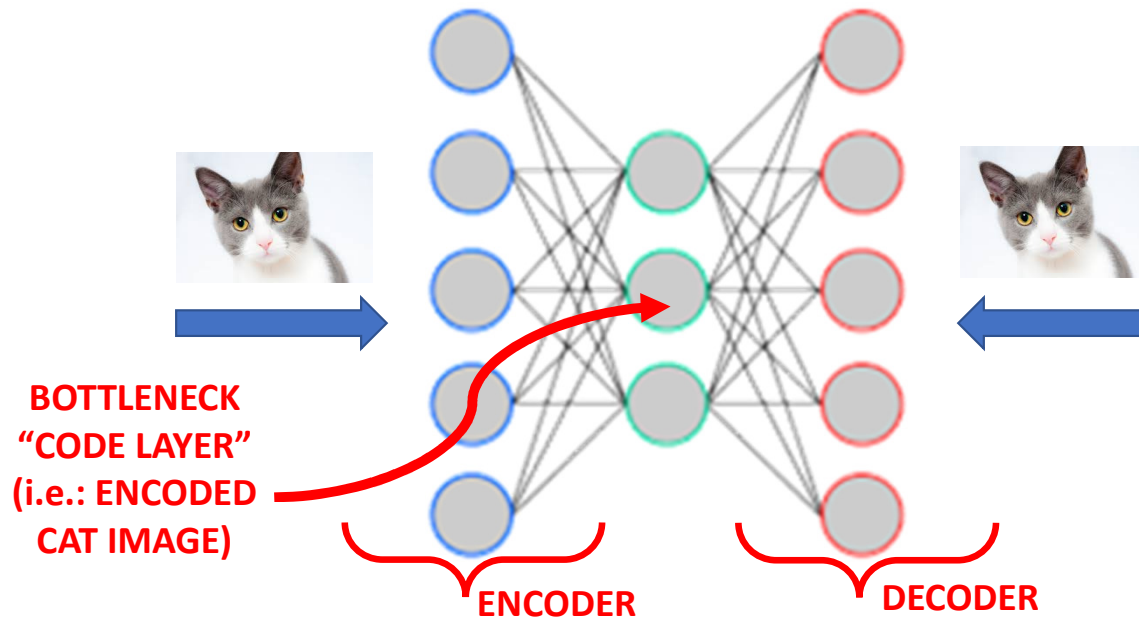


Photo Credit: [https://commons.wikimedia.org/wiki/File:Autoencoder\\_structure.png](https://commons.wikimedia.org/wiki/File:Autoencoder_structure.png)

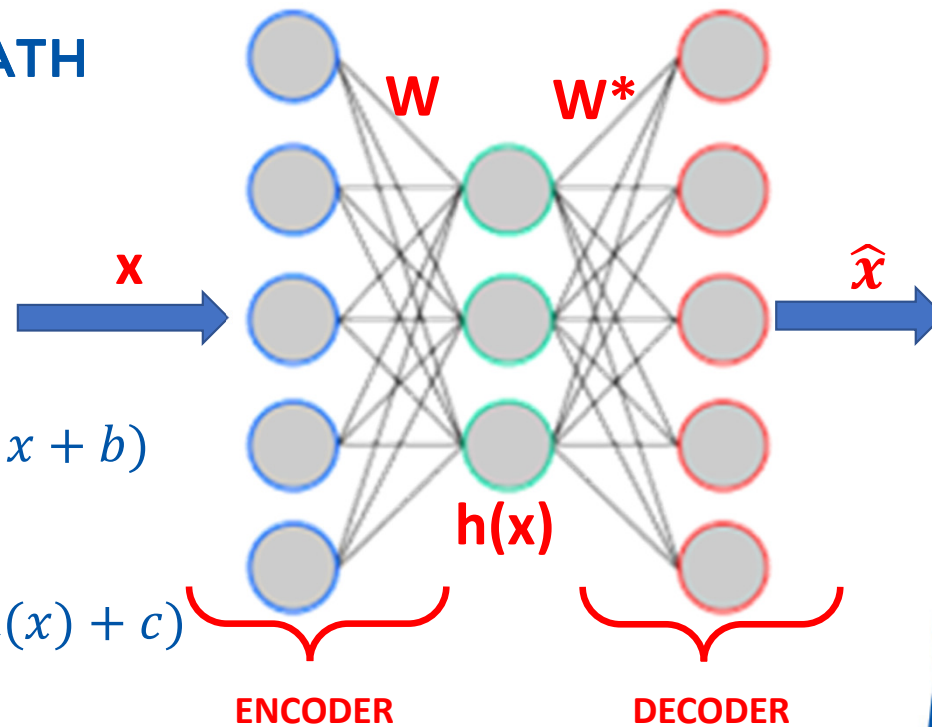
Photo Credit: [https://commons.wikimedia.org/wiki/File:Artificial\\_neural\\_network\\_image\\_recognition.png](https://commons.wikimedia.org/wiki/File:Artificial_neural_network_image_recognition.png)

Photo Credit: <https://www.pexels.com/photo/grey-and-white-short-fur-cat-104827/>





# AUTOENCODER MATH



## ENCODER:

$$h(x) = \text{sigmoid}(W * x + b)$$

## DECODER:

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

## 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](https://commons.wikimedia.org/wiki/File:Autoencoder_structure.png)

Photo Credit: [https://commons.wikimedia.org/wiki/File:Artificial\\_neural\\_network\\_image\\_recognition.png](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.

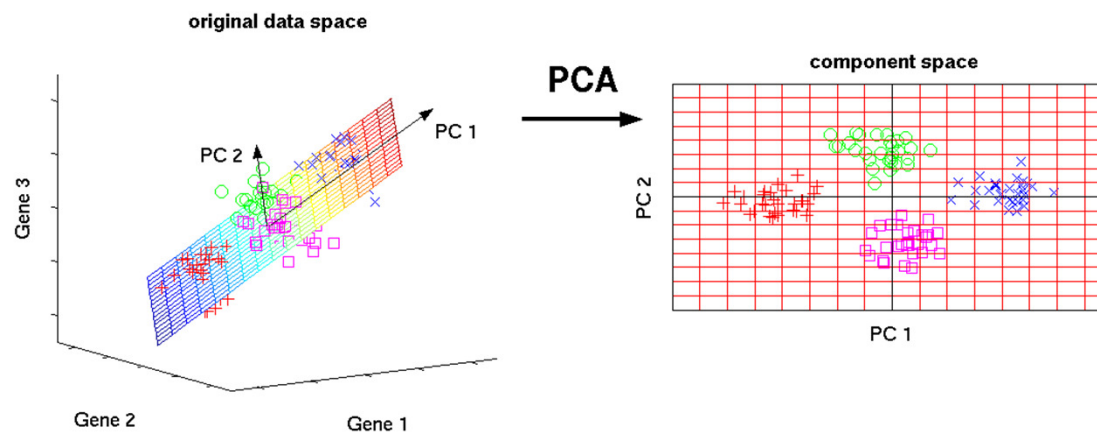


Photo Credit: <http://phdthesis-bioinformatics-maxplanckinstitute-molecularplantphys.matthias-scholz.de/>

