



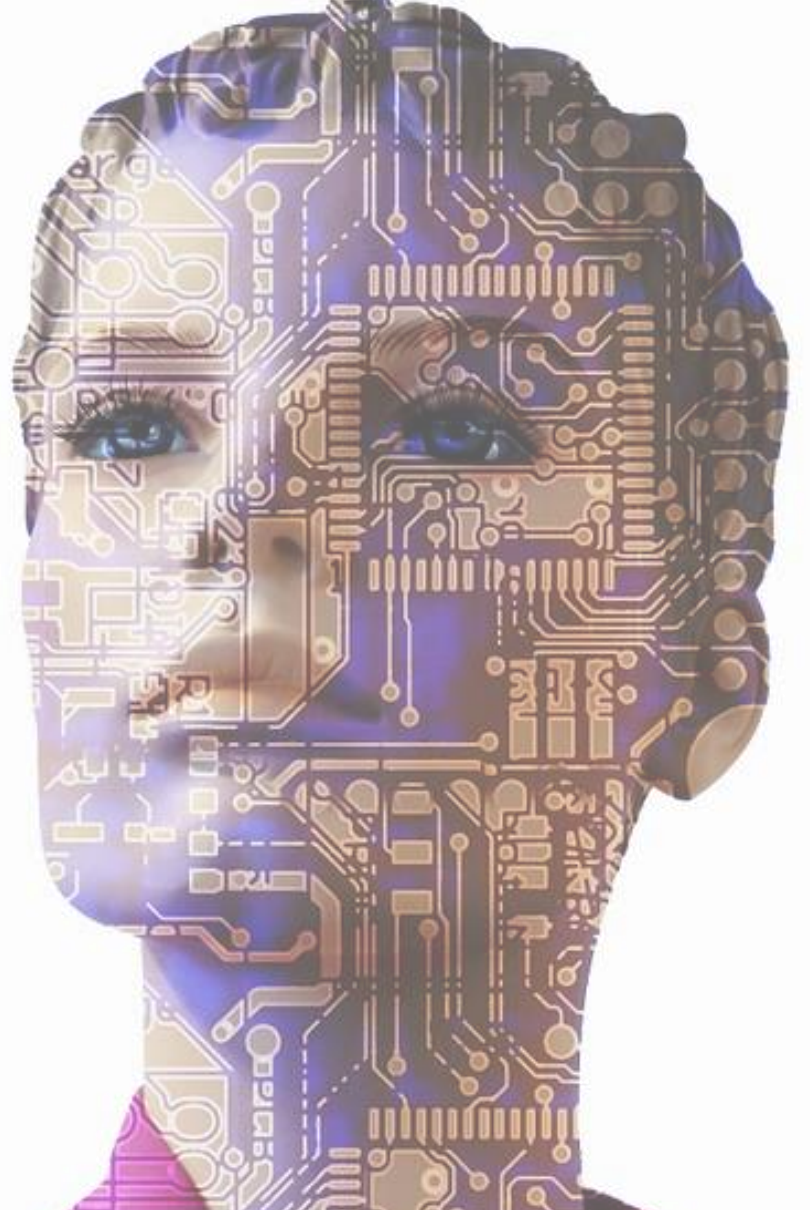
**Universidade do Minho**  
Escola de Engenharia  
Departamento de Informática

**Mestrado Integrado em Engenharia Informática**  
**Mestrado em Engenharia Informática**  
**Aprendizagem e Extração de Conhecimento**  
**2019/2020**

**Paulo Novais, Filipe Gonçalves**

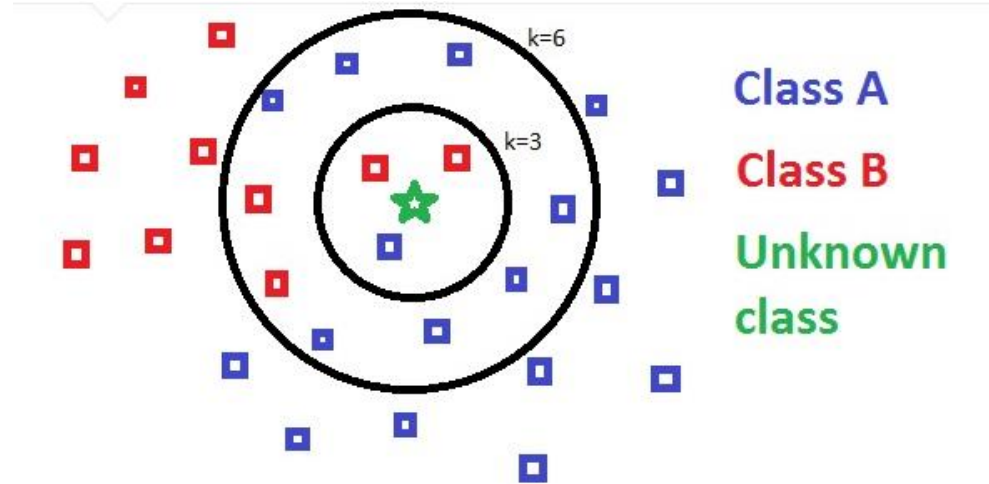
- Paulo Novais – [pjon@di.uminho.pt](mailto:pjon@di.uminho.pt)
- Filipe Gonçalves – [fgoncalves@algoritmi.uminho.pt](mailto:fgoncalves@algoritmi.uminho.pt)
  
- Departamento de Informática  
Escola de Engenharia  
Universidade do Minho
- ISLab – (Synthetic Intelligence Lab)
- Centro ALGORITMI  
Universidade do Minho

# K-Nearest Neighbor



### K-Nearest Neighbor (KNN)

- Classification algorithm that operates on a very simple principle
- Best shown through the next example:
  - Image we had a dataset on Dogs and Horses, with heights and weights

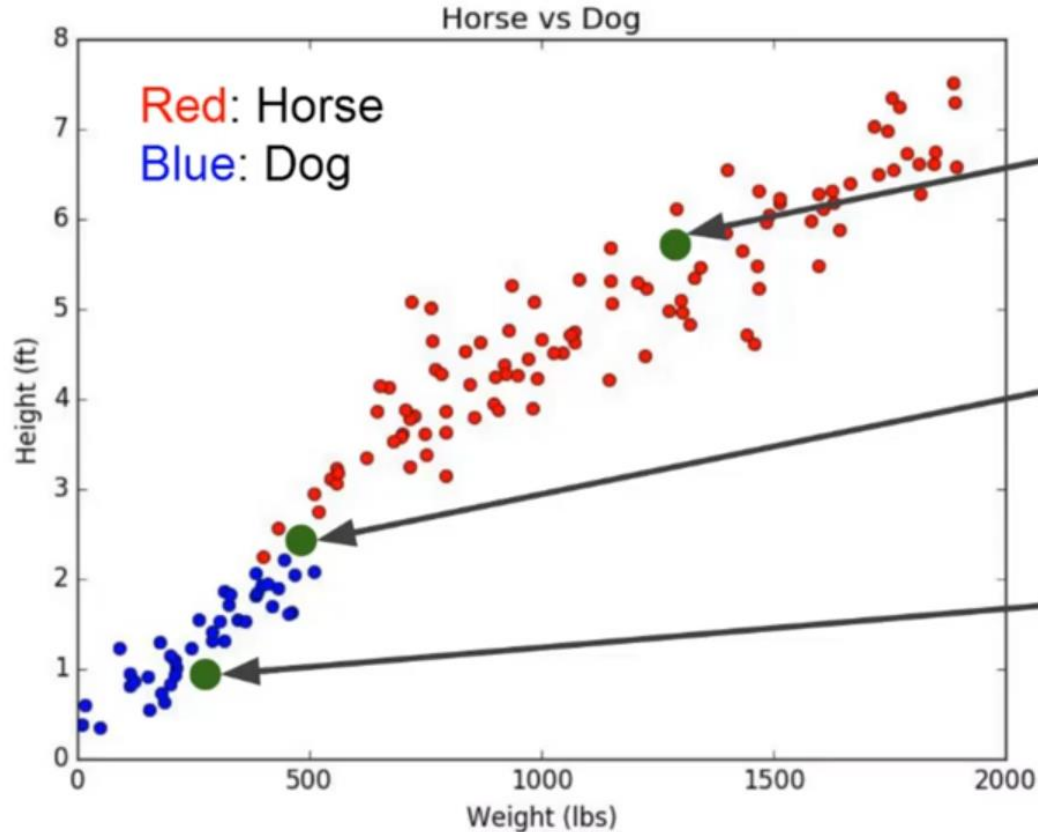




# ISLab

Synthetic Intelligence Lab

AEC @ 2019/2020



New datapoint:  
Is it a horse or a dog?

New datapoint:  
Is it a horse or a dog?

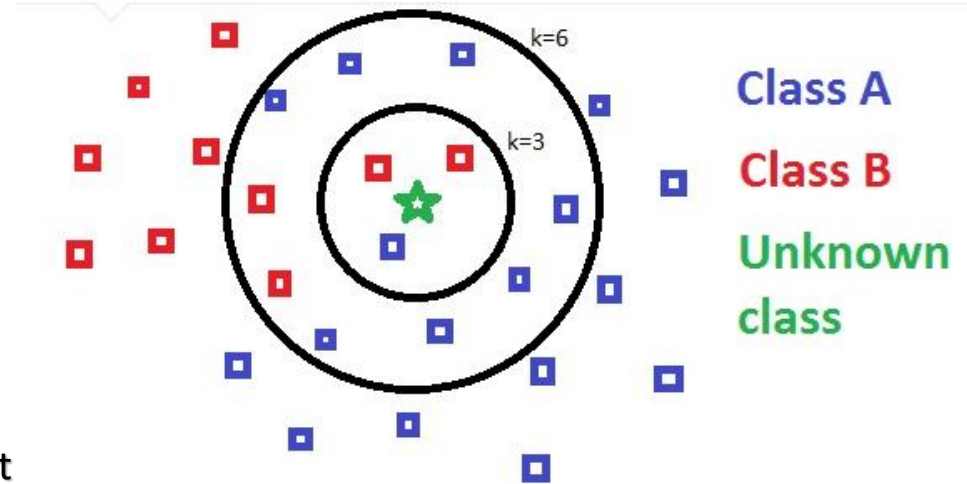
New datapoint:  
Is it a horse or a dog?

**Training Algorithm:**

- Store all the Data

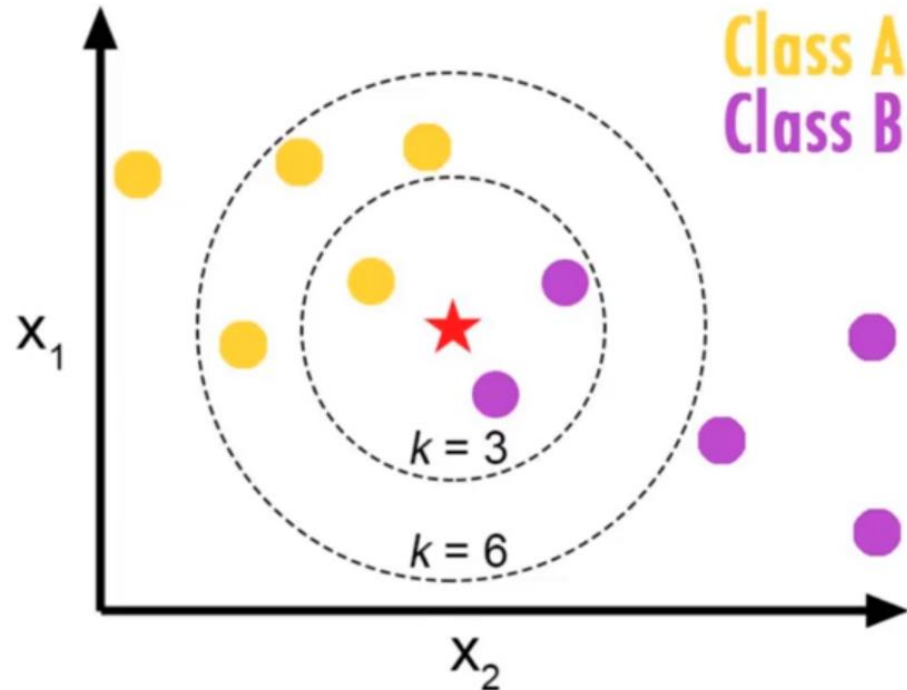
**Prediction Algorithm:**

- Calculate the distance from X (case to predict) to all points in your data
- Sort the points in your data by increasing distance from X
- Predict the majority label of the “k” closest points



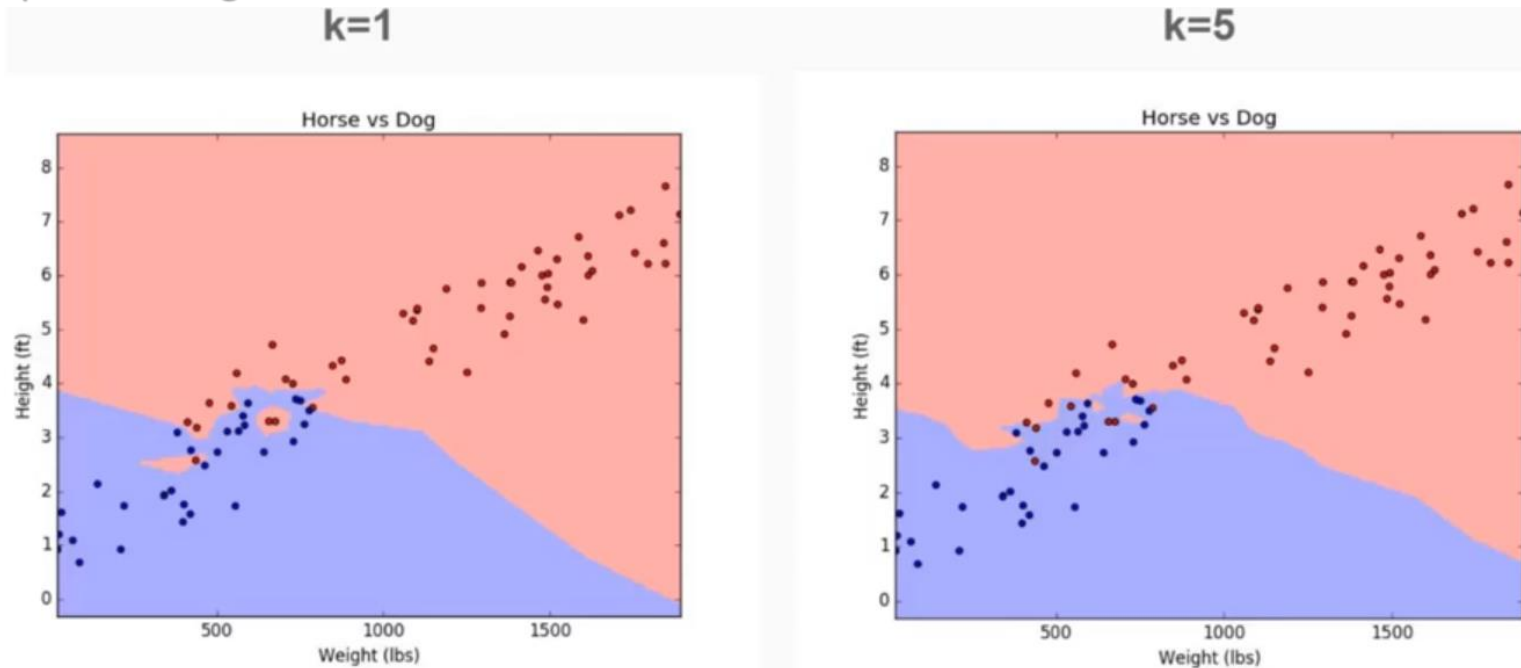
### K-Nearest Neighbor (KNN)

- Choosing a K will affect what class a new point is assigned to:



## K-Nearest Neighbor (KNN)

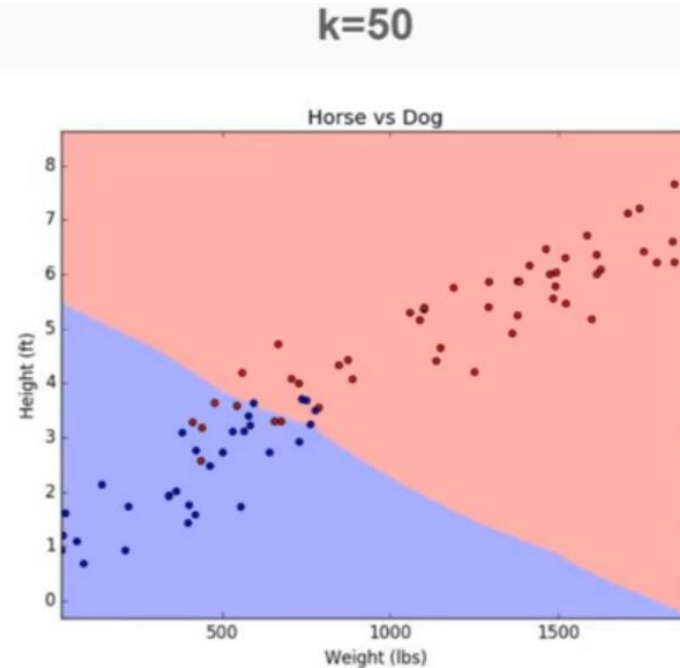
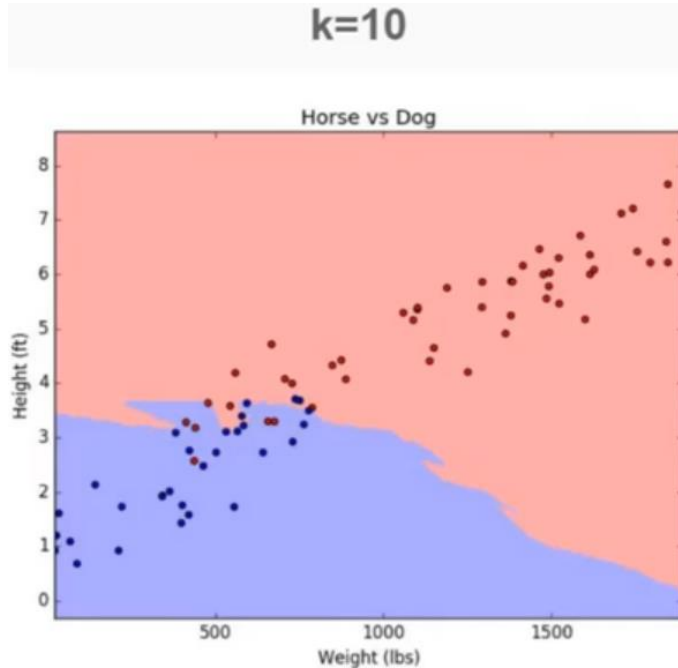
- Choosing a K will affect what class a new point is assigned to:





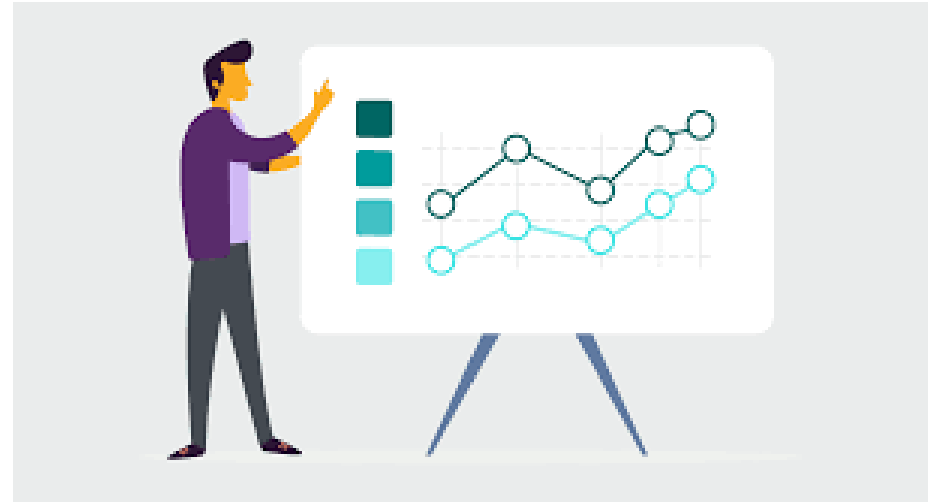
## K-Nearest Neighbor (KNN)

- Choosing a K will affect what class a new point is assigned to:



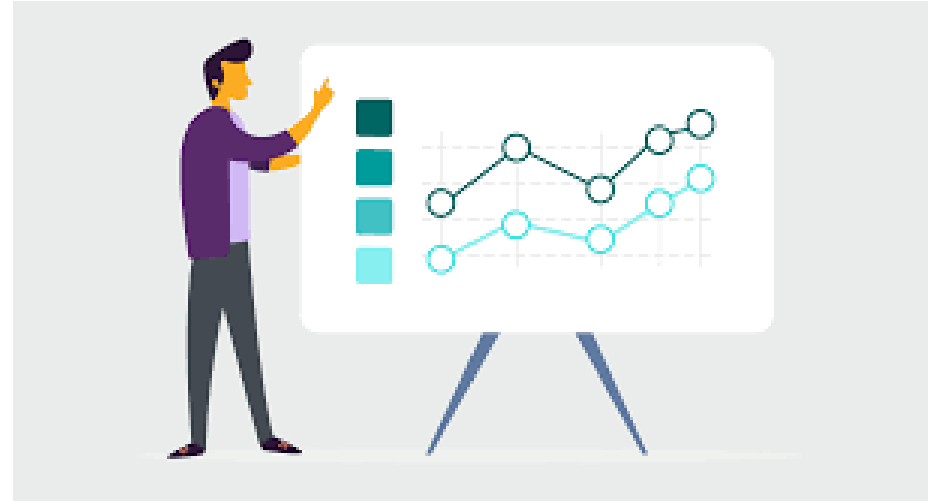
**Pros:**

- Very simple
- Training is trivial
- Works with any number of classes
- Easy to add more data
- Few parameters:
  - K
  - Distance Metric



**Cons:**

- High Prediction Cost
  - Larger Data sets provides worse computational impact
- Not good with high dimensional data
- Categorical Features don't work well



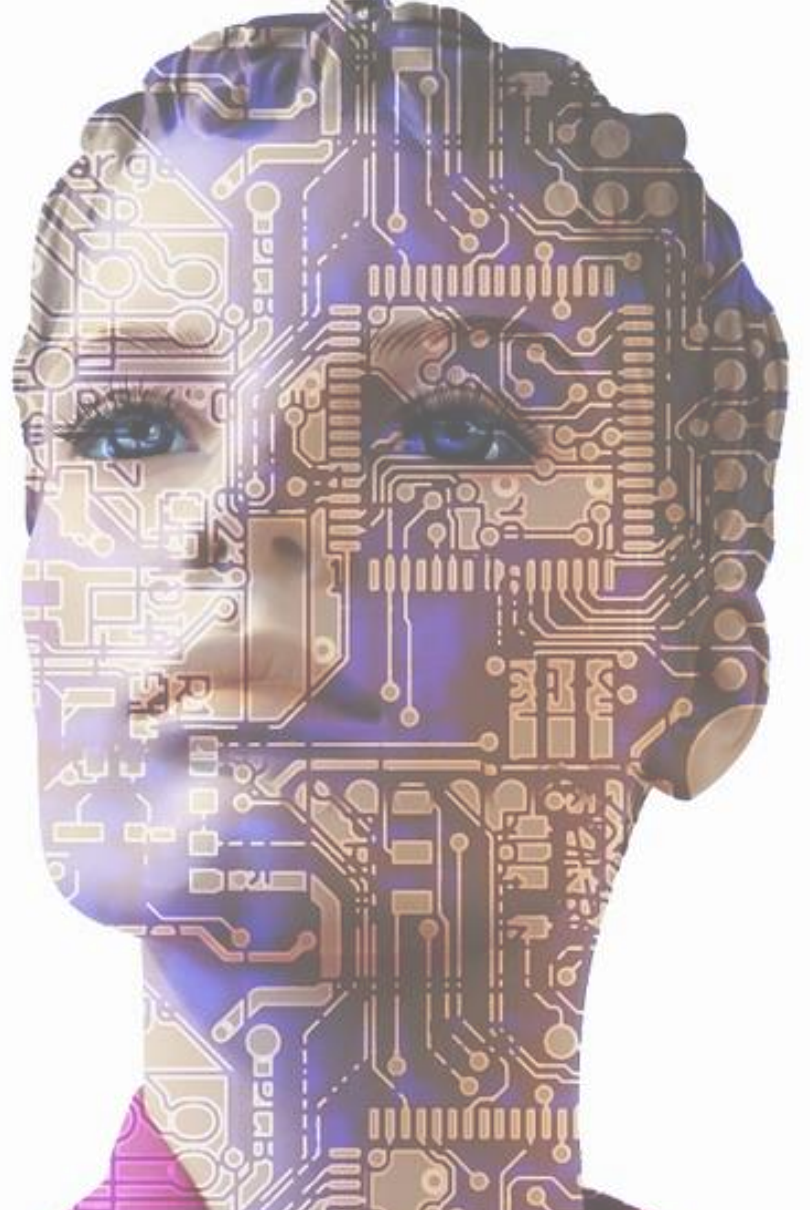
## K-Nearest Neighbor (KNN)

- One of the simplest machine learning models there is
  - Qualified as “supervised learning”
- Can be applied for detecting similarities between users / products
- Example:
  - Movie similarities based on metadata!

Customers Who Watched This Item Also Watched



# Recommender Systems



### What are recommender systems?

#### Frequently Bought Together



Price For All Three: **\$48.15**

[Add all three to Cart](#) [Add all three to Wish List](#)

[Show availability and shipping details](#)

- ☒ **This item:** Scientific Selling: Creating High Performance Sales Teams through Applied Psychology and Testing by Nancy Martini Hardcover **\$16.71**
- ☒ Revenue Disruption: Game-Changing Sales and Marketing Strategies to Accelerate Growth by Phil Fernandez Hardcover **\$15.40**
- ☒ The New Power Base Selling: Master The Politics, Create Unexpected Value and Higher Margins, and Outsmart the Competition by Jim Holden Hardcover **\$16.04**

#### Customers Who Bought This Item Also Bought



**Power Questions: Build Relationships, Win New ...**  
 > Andrew Sobel  
 ★★★★★ (95)  
 Hardcover  
**\$14.38**




**Revenue Disruption: Game-Changing Sales and ...**  
 > Phil Fernandez  
 ★★★★★ (7)  
 Hardcover  
**\$15.40**



**The New Power Base Selling: Master The ...**  
 > Jim Holden  
 ★★★★★ (12)  
 Hardcover  
**\$16.04**



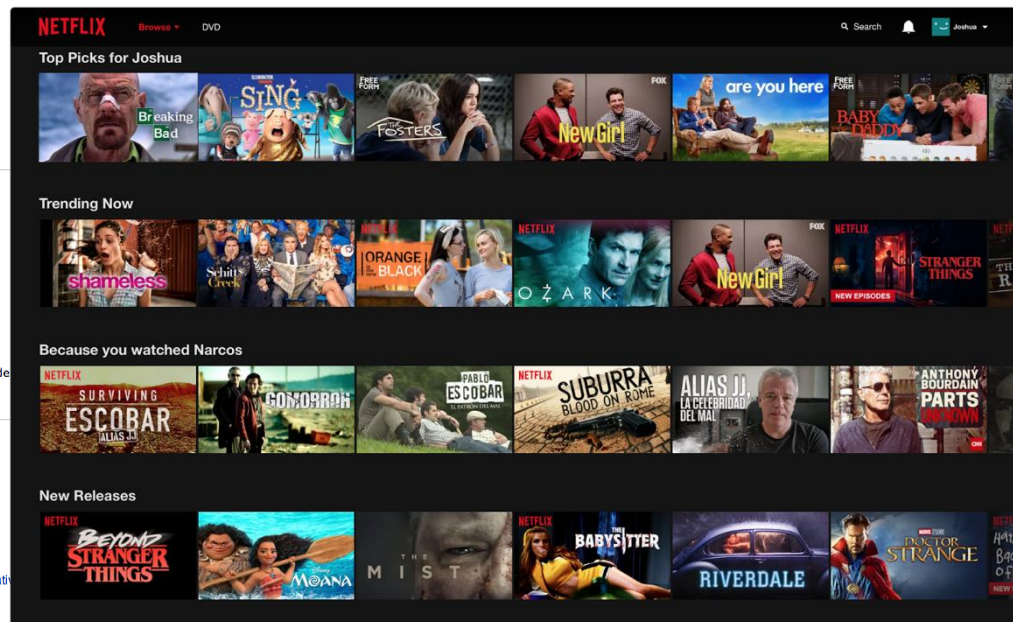
**The Challenger Sale: Taking Control of the ...**  
 > Matthew Dixon  
 ★★★★★ (37)  
 Hardcover  
**\$16.84**  
[Add to Cart](#)



**Pitch Anything: An Innovative Method for ...**  
 > Oren Klaff  
 ★★★★★ (66)  
 Hardcover  
**\$13.09**

Are any of these items inappropriate for this page? [Let us know](#)

#### Editorial Reviews



★★★★★ (10)  
 Hardcover  
**\$24.09**



### User-Based Collaborative Filtering

- Builds a matrix of products each user bought / viewed / rated
- Compute similarity scores between users & filter users with similar aspects (e.g. Pearson Correlation similarity measure)
- Recommendation engine focuses on the users behaviours
- Recommends products past users bought / viewed / rated that the new user hasn't yet

					
A					
B					
C					
D					
E					

## User-Based Collaborative Filtering





### Problems with User-Based CF

- Users are fickle – tastes change
- There are usually many more users than products
  - Data sparsity problems
- People commit mistakes that may influence negatively the Recommendation Systems
- Harmful Bots may provide further negative impact
  - Define rule-based system to filter outliers / strange behaviours

					
A					
B					
C					
D					
E					

**What if we based recommendations on similarities between things instead of users?**


- Technique called Item-Based Collaborative Filtering
  - Recommendation engine focused on similarity between items to make predictions
- Products don't present updates (contrary to user's ratings)
- There are usually fewer products than users (less computation to do)
- Harder to influence negatively the recommendation system

### Item-Based Collaborative Filtering

- Find every product that were bought / viewed / rated by the same user
- Measure the similarity of their ratings across all users who analysed every product
- Sort by product, then by similarity strength

Example:

- Look for items that are similar to Item5
- Take Alice's ratings for these items to predict the rating for Item5

	Item 1	Item 2	Item 3	Item 4	Item 5
User 1 	8	1	?	2	7
User 2 	2	?	5	7	5
User 3 	5	4	7	4	7
User 4 	7	1	7	3	8
User 5 	1	7	4	6	?
User 6 	8	3	8	3	7

**Item-Based Collaborative Filtering - Example:**

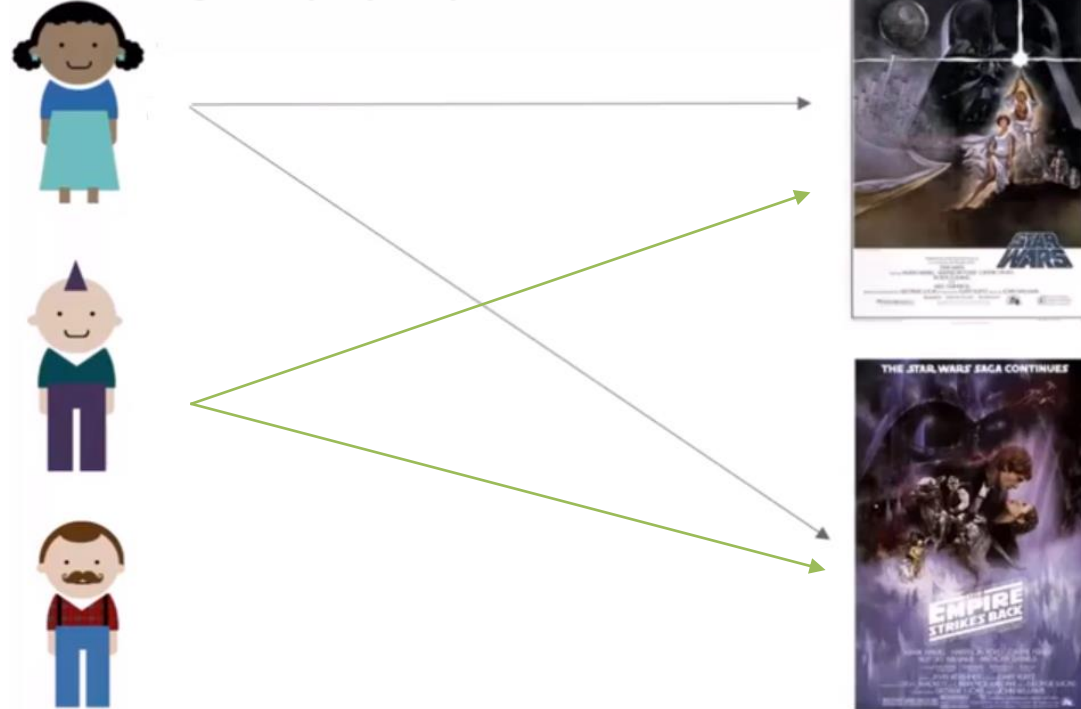
- Look for items that are similar to Item5
- Take Alice's ratings for these items to predict the rating for Item5

	Item1	Item2	Item3	Item4	Item5
Alice	5	3	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1

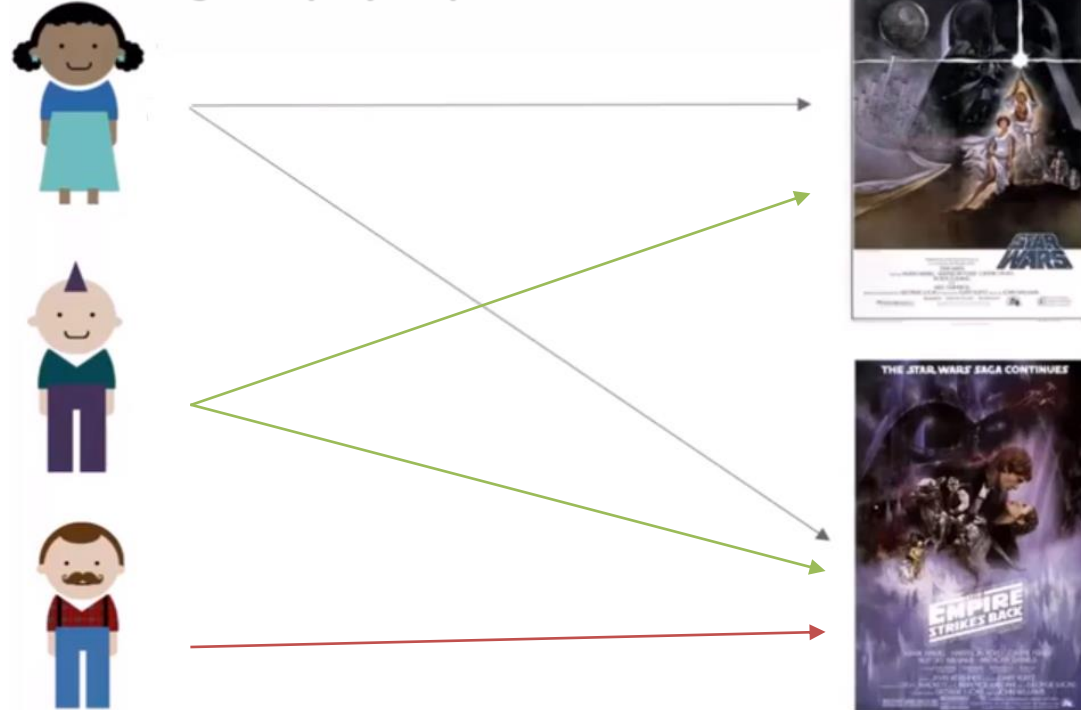
## Item-Based Collaborative Filtering – Step by Step:



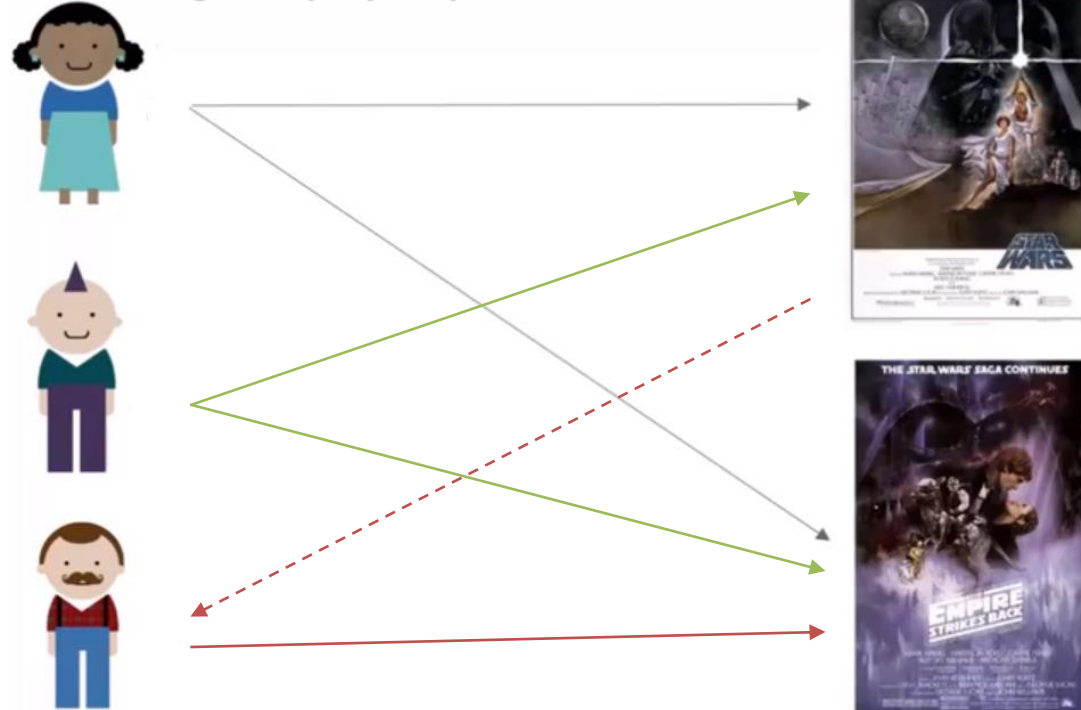
## Item-Based Collaborative Filtering – Step by Step:



## Item-Based Collaborative Filtering – Step by Step:

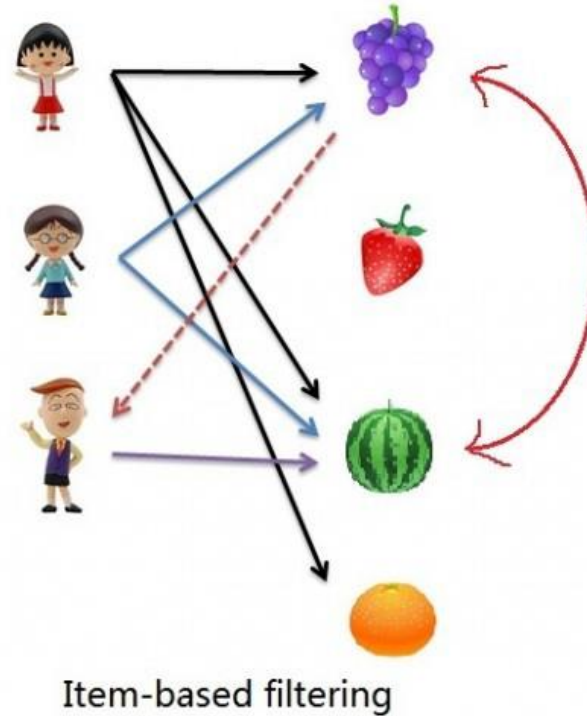
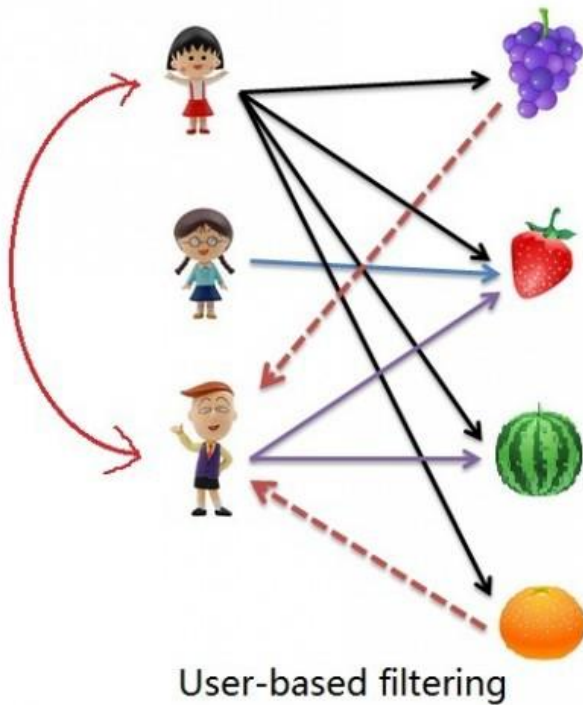


## Item-Based Collaborative Filtering – Step by Step:





### User-Based CF vs Item-Based CF:





**Universidade do Minho**  
Escola de Engenharia  
Departamento de Informática

**Mestrado Integrado em Engenharia Informática**  
**Mestrado em Engenharia Informática**  
**Aprendizagem e Extração de Conhecimento**  
**2019/2020**

**Paulo Novais, Filipe Gonçalves**