DATA SOCIETY:

Intro to classification - kNN - 1

One should look for what is and not what he thinks should be. (Albert Einstein)

Topic introduction

In this part of the course, we will cover the following concepts:

- Supervised learning and its use cases
- The theory behind kNN algorithm
- Implementation of kNN on a dataset
- Performance optimization for kNN

DATASOCIETY: © 2023

Chat question

- This course is about **classification**, a machine learning method for determining whether an **observation** belongs in **one category or another**
- A common use for classification algorithms is filtering **spam email**
- What **features** do you think spam classification algorithms are **trained** to detect?

DATASOCIETY: © 2023

Module completion checklist

Objective	Complete
Describe classification and its uses	
Summarize the steps and application of kNN	

DATASOCIETY: © 2023

Classification

- Classification is a type of **supervised learning** method
  - The data for such methods is labeled into **classes** often by humans, hence the name of the method

- This translates into having two types of variables in our data:
  - **Predictor variables** (a.k.a. features) - can be numeric, categorical, etc.
  - **Target variable** (a.k.a. class variable) - can only be categorical

Classification: binary vs multi-class

- Depending how many categories are within the target variable, we will have
  - A **binary** classification model with only **2** possible outcomes
  - A **multi-class** classification model with **3 or more** outcomes

Commonly used classification methods

- kNN
- Logistic regression
- Support Vector Machines
- Random Forests
- Do you know of any other methods?

DATASOCIETY: © 2023

Classification vs Regression

- Although we will not be discussing regression as a supervised learning method here, you will certainly encounter it at some point
- Below is table of comparison of some classification and regression methods that could help you navigate through common features and differences between the two

	Classification	Regression
Target variable	Discrete, usually binary	Continuous
Types	Binary, Multi-Class	Linear, polynomial
Algorithms	Decision trees, random forests, logistic regression, k-Nearest Neighbors	Linear regression, regression trees, time-series regression

DATASOCIETY: © 2023

# Classification: general use cases

- These are some examples of how you would apply classification algorithms in a medical setting

Question	Example
What is this object like?	Selecting similar drugs or similar diseases
Who is this person like?	Finding patients that are suffering similar symptoms
What category is this in?	Anticipating if your patient will need emergency services
What is the probability that something is in a given category?	Determining the probability that a drug is a particular type or can be used for a particular treatment

Module completion checklist

Objective	Complete
Describe classification and its uses	✔
Summarize the steps and application of kNN	

Steps of kNN
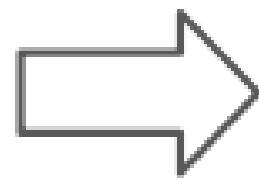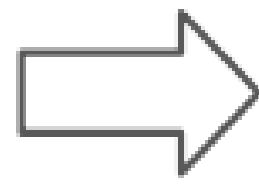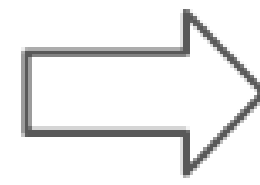
DATASOCIETY: © 2023
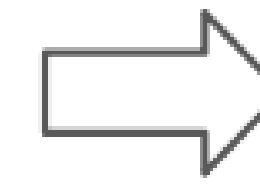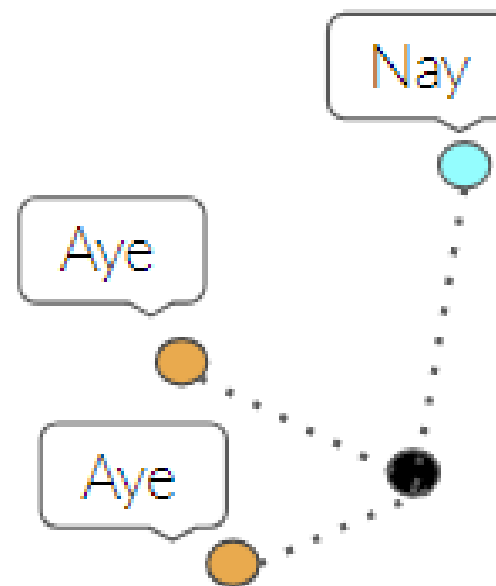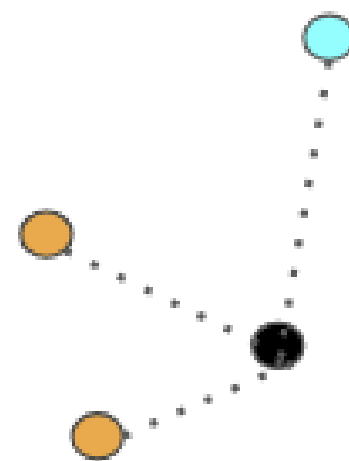
k-Nearest Neighbors: setup
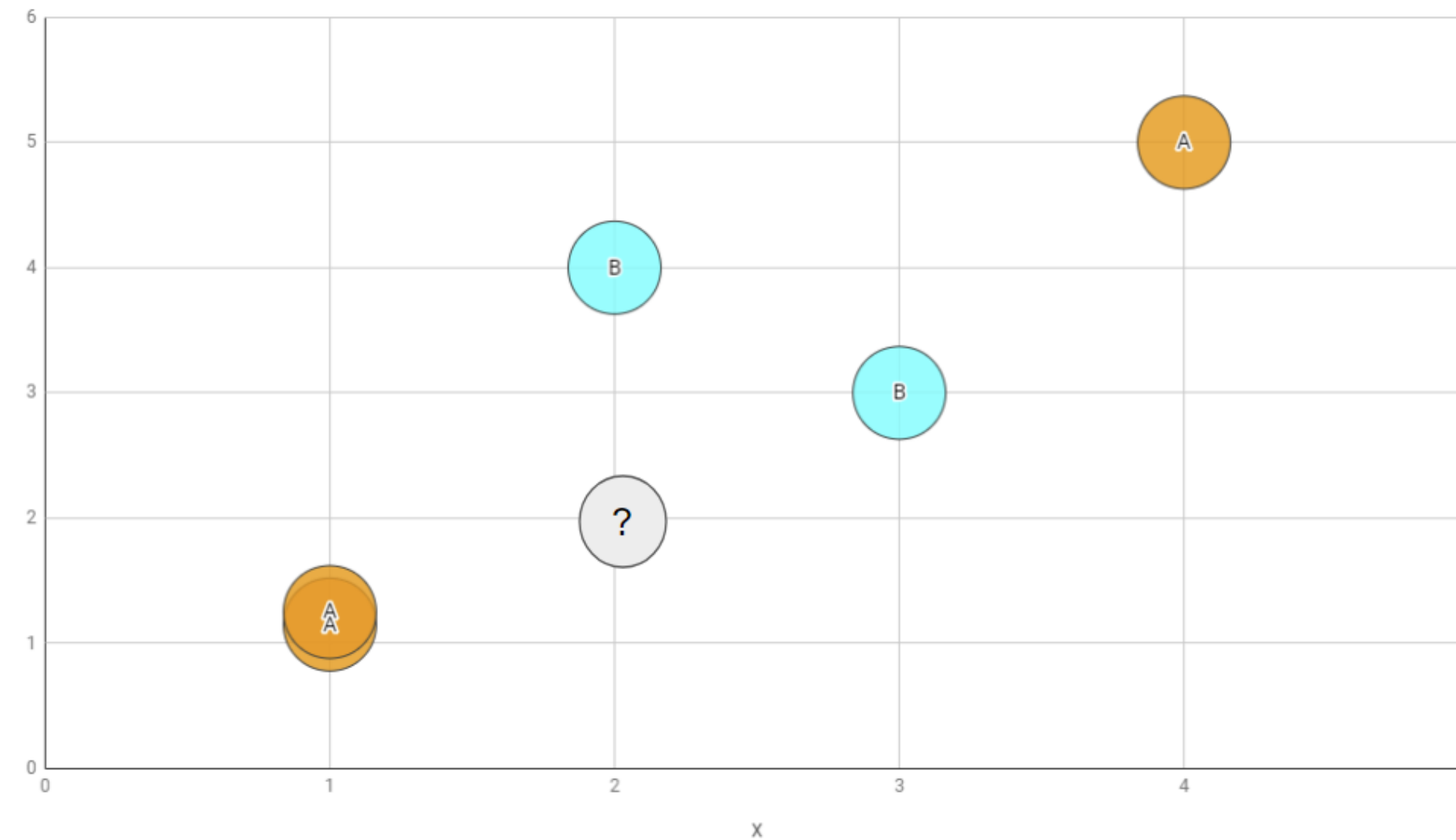
k-Nearest Neighbors: measure
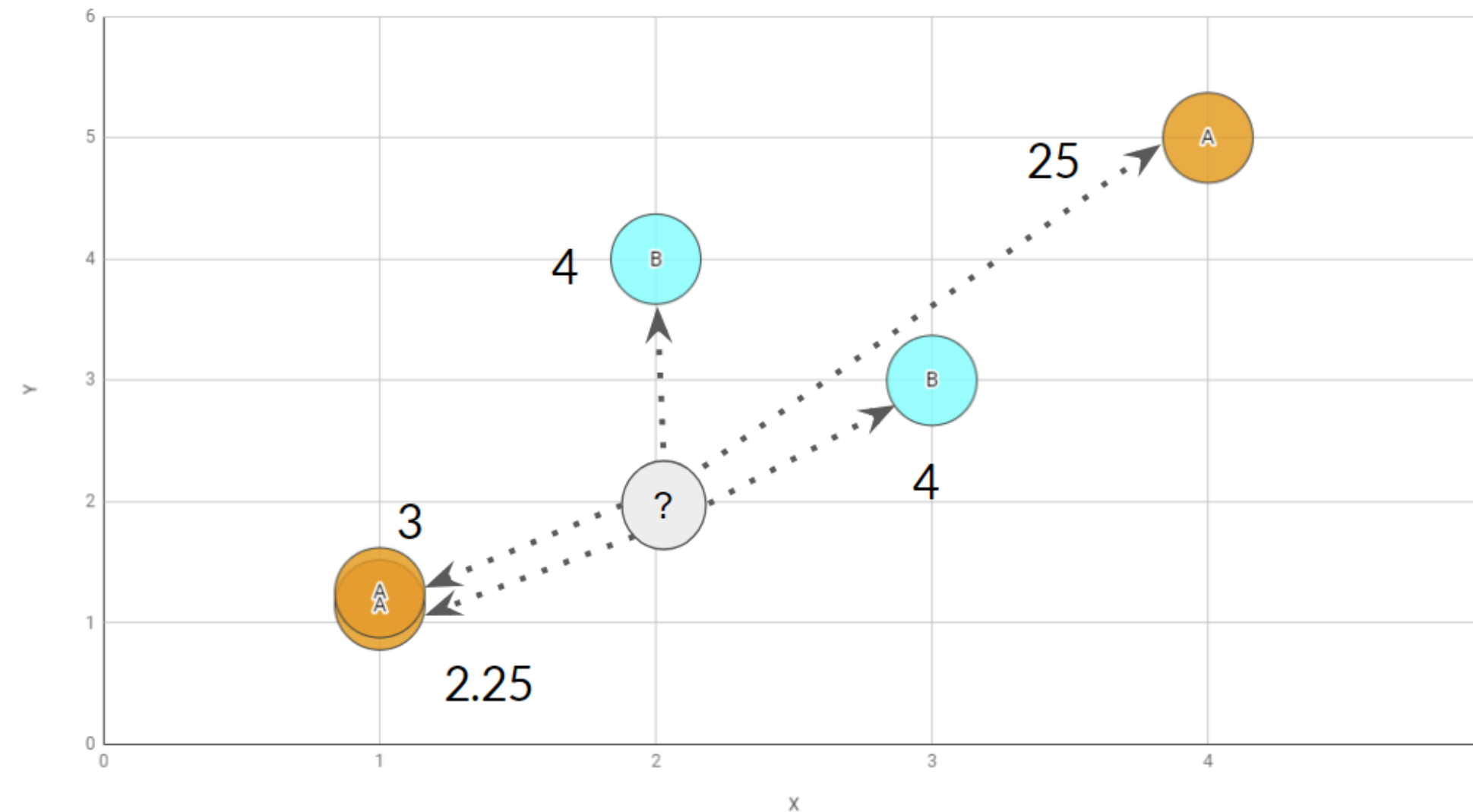
k-Nearest Neighbors: 2-NN for majority vote

DATASOCIETY: © 2023
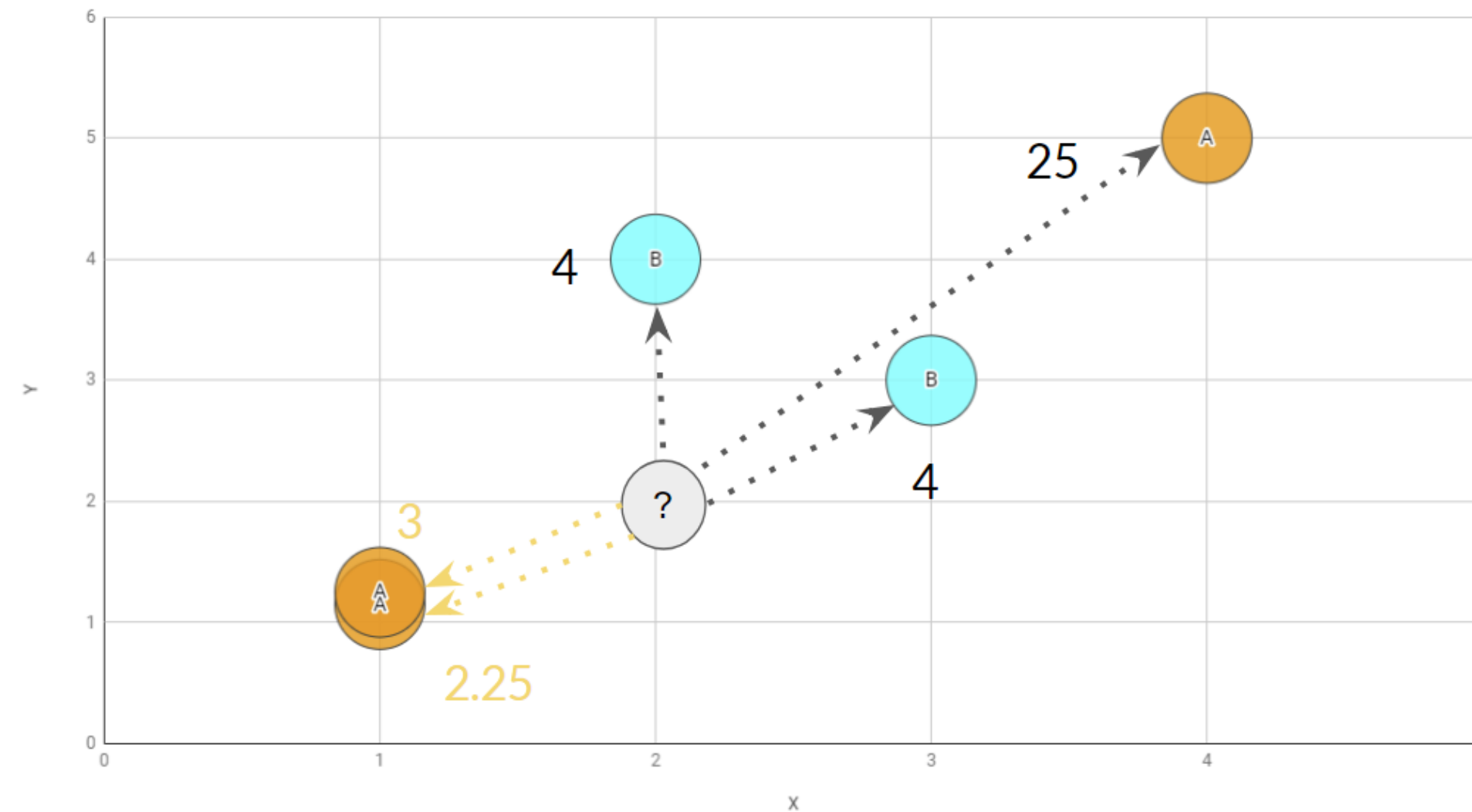
k-Nearest Neighbors: label point
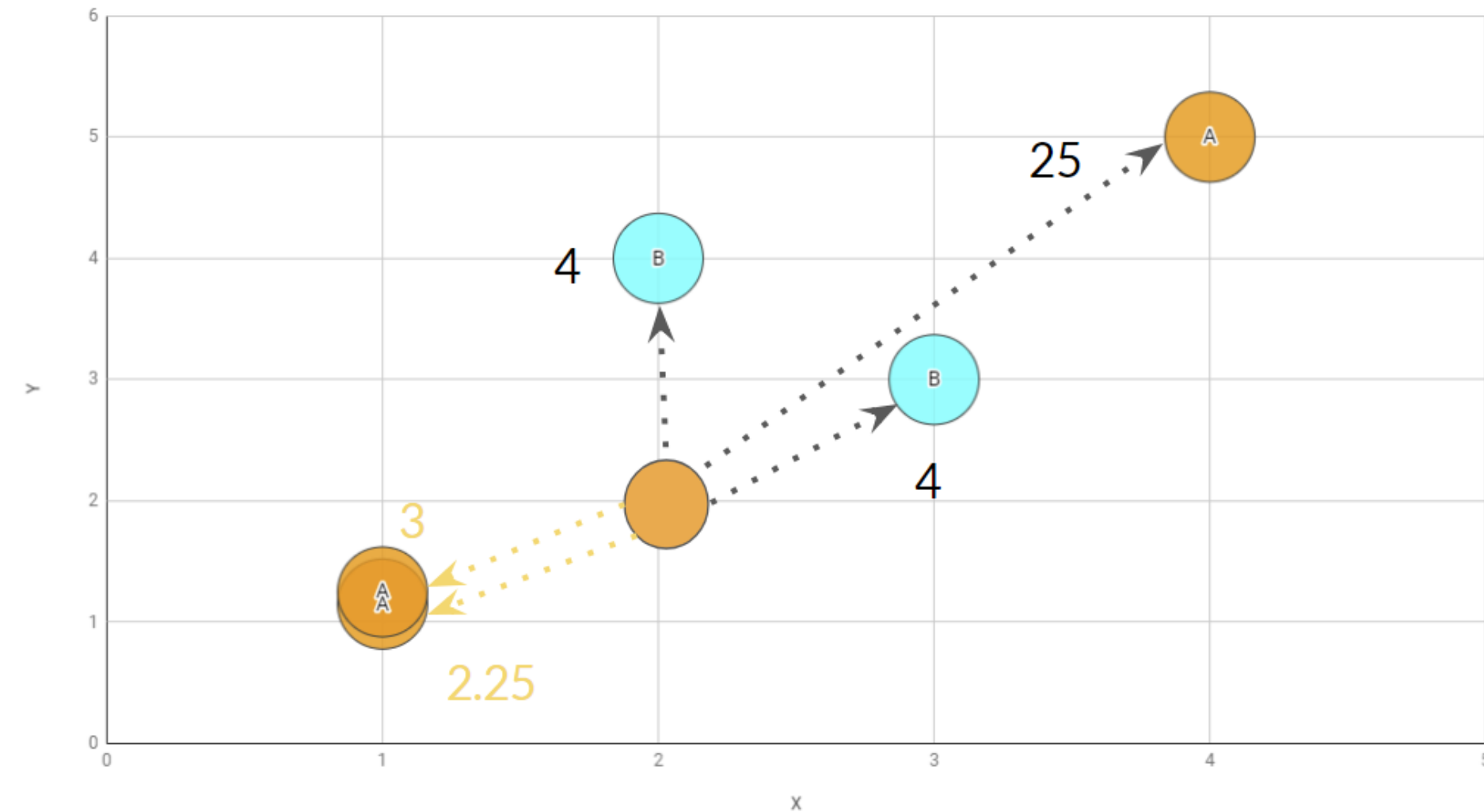
Knowledge check

Module completion checklist

Objective	Complete
Describe classification and its uses	✔
Summarize the steps and application of kNN	✔

DATASOCIETY: © 2023

Congratulations on completing this module!

DATASOCIETY: © 2023