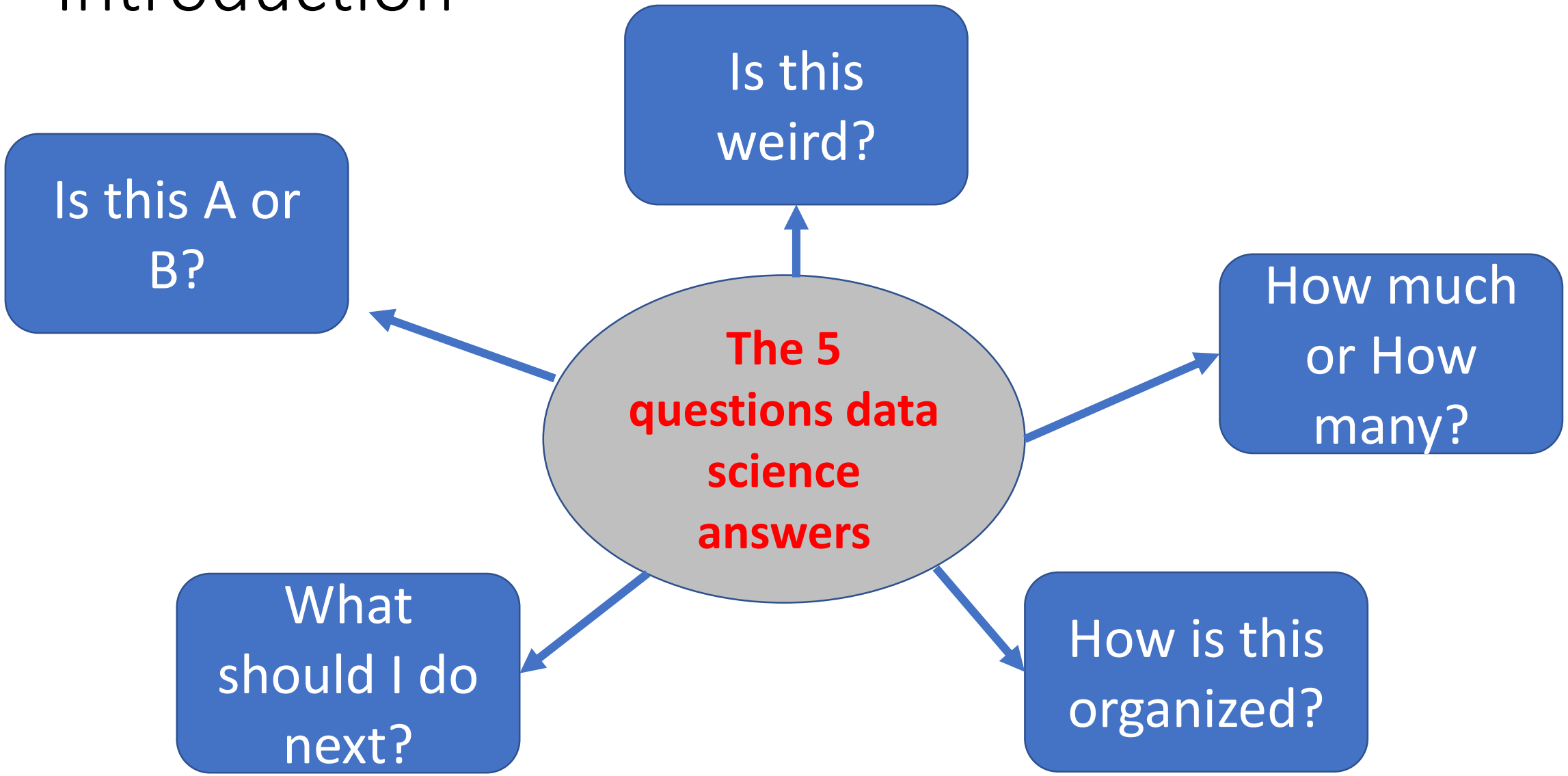
Data Science

five questions

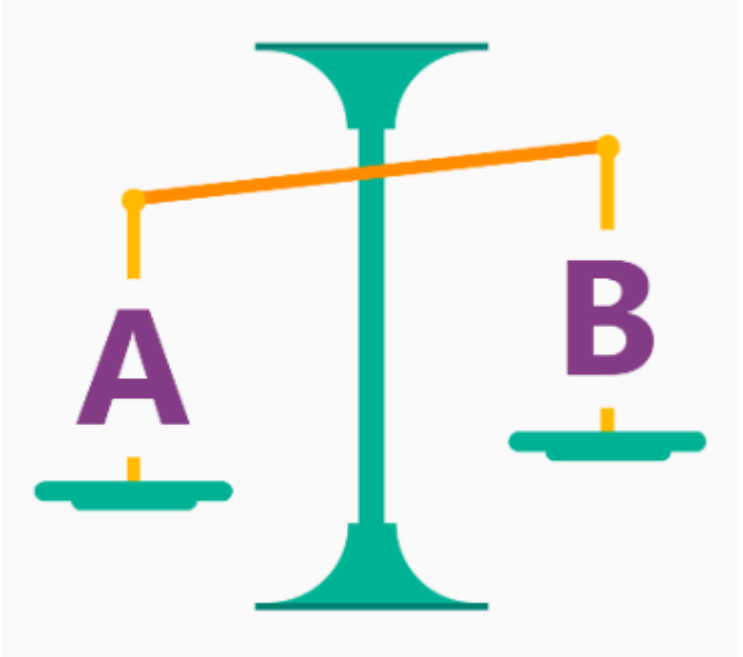
Nguyen Van Quan

Introduction



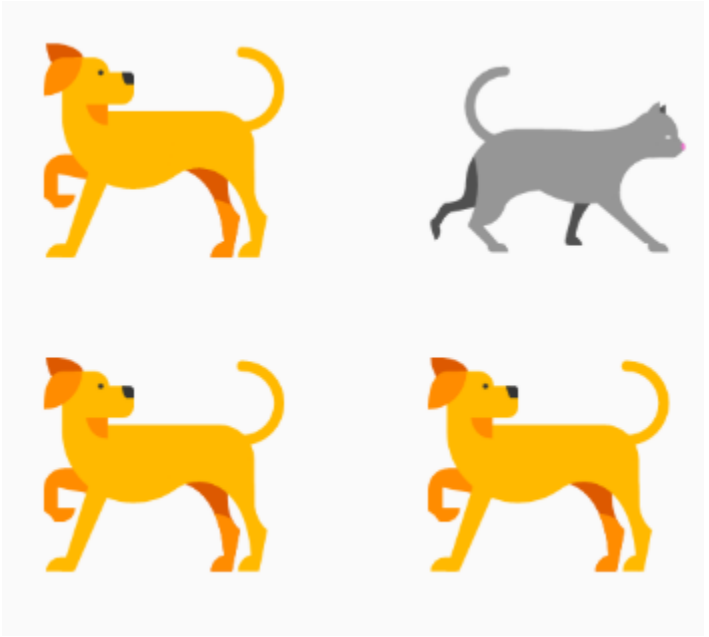
Question 1

- **Is this A or B? uses classification algorithms**
  - This family of algorithm is called as binary classification
  - It is useful for nay question that has just two possible answers (Yes/No; True/False)
  - We can extend this problem to multi-classification



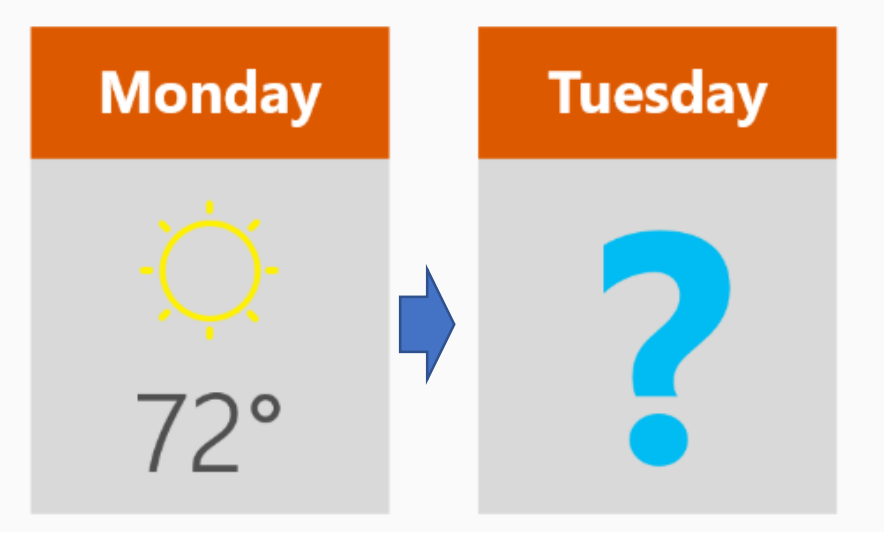
Question 2

- Is this weird? This question is answered by a family of algorithms called anomaly detection.
- Anomaly detection flags unexpected or unusual event or behavior → where to look for problems



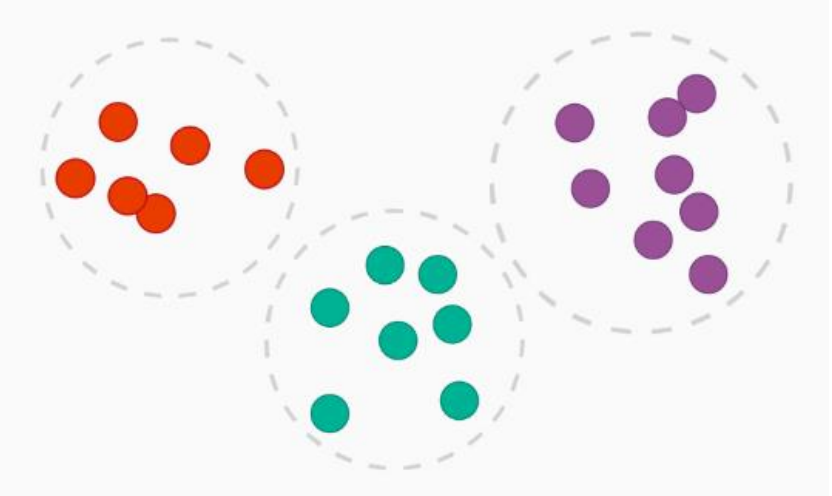
Question 3

- **How much? or How many? uses regression algorithms**
  - **Regression algorithms make numerical prediction**
  - **They help answer any question that ask for a number**



Question 4

- **How is this organized? uses clustering algorithms**
- **Common example of clustering question are:**
  - Which viewers like the same types of movies
  - Which printer models fail the same way
- By understanding how data is organized, you can better understand - and predict - behaviors and events



Question 5

- **What should I do now? uses reinforcement learning algorithms**
- Question it answers are always about what actions should be taken? –usually by a machine or a robot
- Reinforcement learning machine algorithm gather data as they go, learning from trial and error

Reference

[1] https://docs.microsoft.com/en-in/azure/machine-learning/studio/data-science-for-beginners-the-5-questions-data-science-answers