Concepts and Categories

Week 9 Tutorial

1. Rule-based Categorization

Pair off with someone

a. Choose a concept that would be familiar to everyone in the room (e.g., a doctor, a cake, a chair, a fruit, a tree, a sandwich, etc…). Feel free to be a bit less serious with your choice of concept.
b. Come up with a rule-based definition for the concept. Discuss your definition and see if your partner can come up with any exceptions
c. For both your's and your partner's concept, you should each draw an example of what you think a prototype looks like (without looking at each other). When you're finished, compare your drawings with each other and discuss any similarities or differences.
d. Define a set of two features that you can use to describe your concept. Ideally, these will be continuous features. If they are discrete, try to transform them into continuous features (e.g., "bread enclosure ratio" instead of "is or is not enclosed by bread"). Plot examples of objects that both fit and don't fit the concept. Try to come up with between 6-9 objects. Can you define a region in this "psychological space" that captures your concept? Discuss with your partner and see if they can find any exceptions

2. Set-theoretic Categorization:

$$\text{sim}(A, B) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A)$$

Now let's use Tversky's set theoretic model:

a. Define a set of features (4-9) that seem relevant for your category
b. Select two positive and two **NEGATIVE** examples
c. Set all the weights theta = 1, alpha = 0.8, beta = 0.6 for now and compute all the pairwise similarities between examples. Use a table like the one below

d. Are examples in each class more similar to each other than the other items in the other class? Do you have any examples of asymmetric similarity?

	1	2	3	4
1				
2				
3				
4				

3. Number Game

Again in pairs, play the number game with each other

a. One person comes up with an abstract rule, where a number either fits or doesn't fit. e.g.,
    i. X is an even number
    ii. X is between 30 and 45
    iii. X is a prime number
    iv. X is a power of 3
b. The other person iteratively proposes a number, while the rule giver can only give yes or no answers (i.e., yes if it fits the rule)
c. At anypoint, the guesser can try to end the game by guessing the rule. How many guesses does it take to arrive at the correct answer?

4. Open Discussion

A. Which of the theories most intuitively captures how the brain represents concepts?

B. Is it necessarily one type of representation, or could there be a combination?

C. How does a neural network represent categories? Which of the theories does it map onto most closely?

D. Do large language models have categories? What type of concepts does chatGPT learn? (we will cover large language models in a later lecture)

Other discussion points…

- Where do “features” of objects come from?
  - Frame problem: How do you know what features are “relevant”?
    - A relevant feature to a machine is not necessarily one for a human: Adversarial example.
  - Can the absence of a feature in two objects be considered a shared feature?
    - Potentially related: Raven’s paradox.
  - Context: Is a concept/category meaningful outside a particular context : e.g., bag of beans as a “chair” when you are tired and there is nothing else to sit on.
  - Pragmatics: To what extent is category membership driven by whether it’s useful towards a particular end?
  - Affordances, Gibson 1966
- Categories as “objects in the world”: But are there really “objects in the world”?
  - How many grains make a heap?
    - Is a “human’, a “human” without its microbiome?
  - Is Pluto a Planet?
    - Categories change across time: “heat”, “energy”, “matter” (E=mc2).
- Categorization as basis of generalization?