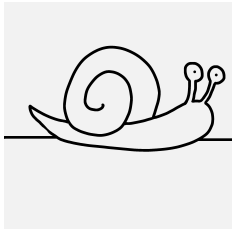
(E) A Typical Problem (1/2) [15 Points]

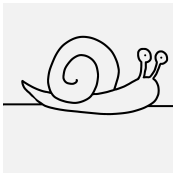


What do you see on the right? You may recognize the image as a snail; however, that quick judgment comes from a lot of behind-the-scenes processing. Your brain picks up on a variety of visual cues, perhaps including a rounded shell, two eyes poking out, and an outstretched body. Then, it combines these cues to make a guess that the image is of a snail.

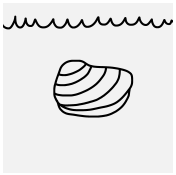
Even though the picture is not photorealistic, those cues let you quickly identify the animal in the image. In machine learning, these cues are called *features*: aspects of data that a system uses to decide on a label (in this case, the label is "snail").

Now that we've seen how your brain might use features to identify a snail, let's see how a simple algorithm might use features to make its own guesses. NACLOLabs wants to design an algorithm that can take in an image and determine which one of eight animals it is.

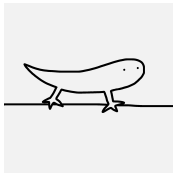
This algorithm will look at the image and ask three questions as its features: *is it underwater?*, *does it have legs?*, and *does it have a shell?*. Then, it will look up the row of the table that matches those features, and produce what it sees in the "output" column. For example, if the answer to all three questions is no (represented by N), then it says that the image contains a snake. However, NACLOLabs left some gaps in the table!



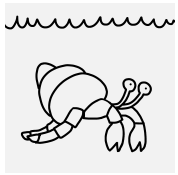
snail



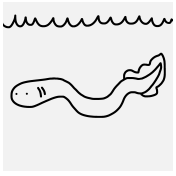
clam



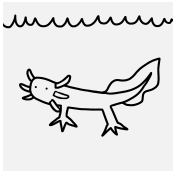
lizard



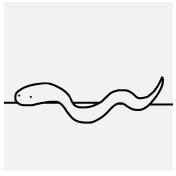
hermit crab



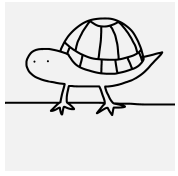
eel



axolotl



snake



tortoise

(a)	(b)	(c)	Output
N	N	N	*snake*
N	N	Y	*snail*
N	Y	(1)	*clam*
N	(2)	(3)	*eel*
Y	(4)	(5)	*lizard*
Y	(6)	N	*axolotl*
Y	N	Y	*tortoise*
Y	Y	Y	*hermit crab*

E1. Match each of **(a)**, **(b)**, and **(c)** to its feature (*underwater*, *has legs*, and *has a shell*):

(a) — ○ *underwater* ○ *has legs* ○ *has a shell*

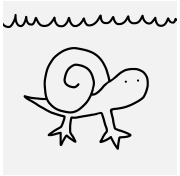
(b) — ○ *underwater* ○ *has legs* ○ *has a shell*

(c) — ○ *underwater* ○ *has legs* ○ *has a shell*

E2. Fill in the gaps (1)–(6) so that the algorithm works on all eight images shown:

(1) ☐ (2) ☐ (3) ☐ (4) ☐ (5) ☐ (6) ☐

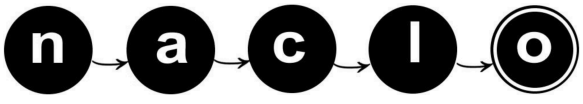
E3. NACLOLabs decides to run its animal-labeling algorithm on an additional input, but gets an unexpected result. Which animal does the algorithm output for the image on the right?



○ *snake* ○ *snail* ○ *clam* ○ *eel*

○ *lizard* ○ *axolotl* ○ *tortoise* ○ *hermit crab*

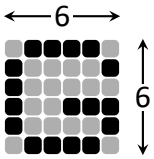
One way that NACLOLabs can improve its algorithm is to use the same feature for several regions of an image, so that it outputs multiple values for each feature (the *value* of a feature refers to whether that feature is indicated as being Y or N). We will see this approach in action on the next page!

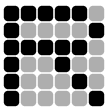
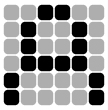
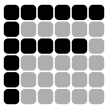
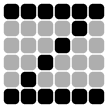
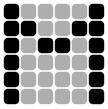
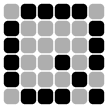


(E) A Typical Problem (2/2)

One place where computational linguistics uses the idea of features is in *optical character recognition*. These methods represent each letter by decomposing it into a set of features. By using the values of all the features together, we can make a reasonable guess about what the original letter is[1]! For example, a circle shape is likely an **O**, while a circle with a tail is likely a **Q**.

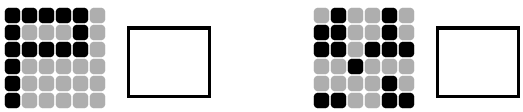
NACLOLabs uses a basic font where **every** capital letter sits on a grid of thirty-six pixels; one example is shown on the right. For each feature, the algorithm considers nine regions of this grid and outputs a sequence of yes and no answers based on that feature. Six inputs and four features are listed below, but there are some gaps in the table.



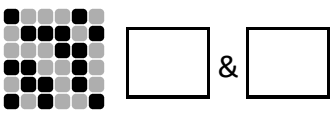
Input	Feature 1	Feature 2	Feature 3	Feature 4	Output
	YNN/YYN/NNN	YYY/YYY/YNY	NYN/NNN/YNN	NNY/NNN/NNY	**R**
	NNN/NNN/NNN		NYN/YYY/YNY		**A**
		YYY/YYY/YNN	(7)		**F**
	(8)	YYY/NYN/YYY		NNN/NYN/NNN	**Z**
	(9)	(10)			**M**
			(11)	(12)	**Q**

E4. Fill in the gaps (7)–(12). (You don't have to fill in the shaded cells.)

E5. NACLOLabs decides to run its letter-labeling algorithm on some additional inputs, but gets some unexpected results. Which letter does the algorithm output for each of the two inputs below?



E6. Sometimes the algorithm gets confused and has two possible answers for a given input. Which two letters does the algorithm output for the input below?



[1]In fact, some NACLO problems are often graded automatically with these techniques!