(K) Going By the Numbers (1/4) [15 Points]

In artificial intelligence, it is common to represent a word with a vector — that is, a list of numbers. For example, the word *wandered* might be represented with the vector shown below:

wandered =

3	1	-1

How can we make sense of this vector? One approach is to use something called a classifier, which takes in a vector and determines whether that vector belongs to a particular category. As a starting example, we will consider a classifier that makes the following decisions:

Words that are labeled **YES** by the classifier: *wandered*, *thought*, *noticed*, *cheered*
Words that are labeled **NO** by the classifier: *wandering*, *thinking*, *noticing*, *cheering*, *wanders*, *thinks*, *notices*, *cheers*

K1. What is the category of words that receive a label of YES from this classifier?

Let's take a closer look at how the classifier makes its decisions. The classifier is defined by another vector that we will call *w*. For our example, *w* looks like this:

w =

1	-2	0

A classifier uses its vector *w* to make a decision as follows:

If *w* · *word* > 0, say **YES**.
Otherwise, say **NO**.

The dot inside "*w* · *word*" stands for an operation that takes in two vectors and returns a single number. It's easiest to illustrate this operation with an example, and then we'll define it in words later. If we use the vectors shown above for *w* and *wandered*, then *w* · *wandered* can be computed as shown:

$$w \cdot wandered = 1*3 + -2*1 + 0*-1 = 3 - 2 + 0 = 1$$

That is, we add together the first number of the first vector (1) times the first number of the second vector (3), plus the second number of the first vector (-2) times the second number of the second vector (1), plus the third number of the first vector (0) times the third number of the second vector (-1), which gives the number 1 (if the vectors were longer than 3 numbers, we would keep going until we had used up all the numbers in the vectors). Since 1 is greater than 0, the classifier would say **YES**.

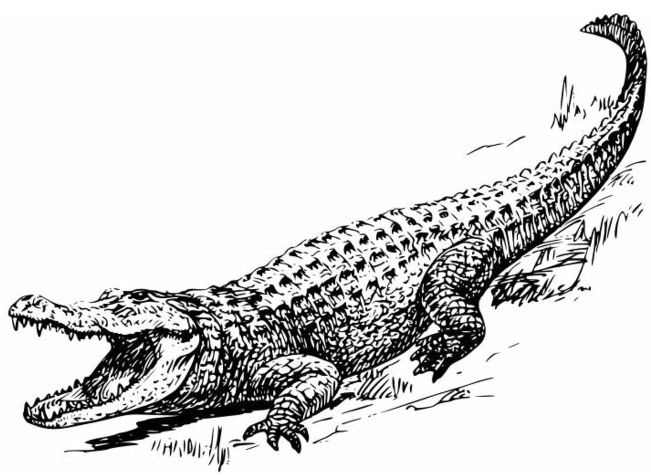
(K) Going By the Numbers (2/4)

On the last page of this problem, we have listed word vectors for 22 words (*alligators* through *visited*). These new vectors contain 7 numbers. We have also listed the *w* vectors for 9 different classifiers, labeled w_1 through w_9.

K2. What is $w_4 \cdot squirrel$? Your answer should be a single number:

Your task will now be to figure out what type of information each of these nine classifiers (w_1 through w_9) focuses on. But before we get into that, we should say a little more about where our word vectors came from. These vectors were produced by a computer system that was trained to predict the next word in a piece of text. For example, given the input *the books on the* ____, it should try to guess what word could go in the blank; it might guess *table* or *chair*. Next-word prediction lies at the heart of many popular systems in artificial intelligence, such as ChatGPT.

A next-word prediction system is usually not programmed by humans. Instead, it is given some data and learns how to predict words by observing patterns in the data. The prediction process uses word vectors to represent the text that is used as input, and while it is learning the system selects values for the word vectors that will help it perform its task. Therefore, we should expect the word vectors to encode information that is useful for predicting what word could come next in particular contexts. As one example, consider w_1. If you apply w_1 to the 22 words, it turns out that w_1 says **YES** for *alligators*, *artists*, *envelopes*, *lawyers*, *oranges*, *parrots*, *plates*, *salads*, and *sticks* (you can check this by computing $w_1 \cdot word$ for each word). In other words, it seems that w_1 tells us whether each word is a plural noun. How does this relate to next-word prediction? There are some situations where the next word is likely to be a plural noun, such as in the context *I found these* ____. Therefore, to handle such contexts, the system should be able to recognize which words are plural nouns. The information that is identified by w_1 does exactly that.



The table on the next page lists the nine classifiers, and for each one it specifies a context where we are likely to see the words that the classifier has labeled **YES** but not the words that it has labeled **NO**.

(K) Going By the Numbers (3/4)

K3. Fill in the blanks **(a)** - **(h)** in the table below. <u>*Notes:*</u>

- In many cases, there are multiple possible answers. You only need to provide one.
- Each classifier (w_1 through w_9) will be used exactly once.
- When providing potential contexts (in the right column), you should focus on which types of words are most likely, rather than which ones are possible. For instance, for w_1, the provided context of *I found these* ____ is most likely to be followed by a plural noun (which is the correct category that w_1 identifies), but there are settings where this context could be followed by a singular noun (such as in *I found these apple pies*). However, it's more likely for a plural noun to go in that context, meaning that *I found these* ____ would be a fully valid answer for w_1 even though technically it could allow singular nouns as well as plural nouns.
- In the contexts that you create (in the right column), you may not use any words from the list of 22 words on the last page of this problem (*alligators*, *apple*, etc.).

Classifier...	**Returns YES for words that are likely in the context...**
w_1	I found these _______
(a)	The person was chased by the _________
(b)	I ate the ________
(c)	Yesterday, the person _________
w_5	**(d)**
w_6	**(e)**
w_7	**(f)**
w_8	**(g)**
w_9	**(h)**

K4. Below are the vectors for a few more words. For each one, provide a single word that the vector could potentially stand for, assuming that these vectors can also be correctly classified by w_1 through w_9. There might be multiple right answers; you only need to provide one. Write one word in each blank:

(a)	5	-1	2	-3	-3	-7	4	
(b)	-5	-2	-1	-2	-12	1	2	
(c)	9	-4	5	-9	6	3	7	

(K) Going By the Numbers (4/4)

alligators =	-5	7	-9	-17	-10	1	3
apple =	5	-4	-3	-5	4	3	2
artists =	-3	3	-2	-5	-24	1	9
book =	5	-1	-4	-14	-5	-4	4
repaired =	2	-1	-2	2	-14	2	5
electrician =	9	2	-1	-2	-12	3	4
envelopes =	-4	-2	-1	-9	-7	-2	3
interpreter =	7	2	-5	-1	-9	2	2
lawyers =	-2	2	-3	-3	-6	1	1
mug =	3	-2	1	-2	-19	7	2
noticed =	1	-8	-4	19	-42	5	9
oranges =	-3	-2	-9	-12	-3	2	8
parrots =	-2	1	-5	-19	-17	2	7
pizza =	5	-3	-2	-1	1	4	3
plates =	-1	-3	3	-4	-9	2	3
purse =	6	-3	-3	-7	-7	-2	4
salads =	-1	-7	-5	-11	-1	1	5
squirrel =	7	2	-2	-5	-10	1	4
sticks =	-3	-5	6	-9	-16	1	4
restaurant =	4	-5	-2	-2	-6	-1	3
television =	4	-6	2	-11	-12	3	3
visited =	1	-2	-2	7	-24	1	9

w_1 =	-3	0	0	0	0	0	0
w_2 =	0	0	0	1	0	0	0
w_3 =	0	1	0	0	0	0	0
w_4 =	0	0	0	0	1	1	2
w_5 =	0	0	0	-1	0	0	0
w_6 =	1	0	0	0	0	0	-2
w_7 =	0	0	1	0	0	0	0
w_8 =	1	0	0	0	0	0	-1
w_9 =	0	0	0	0	0	-1	0