1.	What would	be the probability	of a five word	I sequence using a	penta-gram?
----	------------	--------------------	----------------	--------------------	-------------

$$\bigcirc P(w_5 \mid w_4, w_3, w_2, w_1) = \frac{\text{count}(w_5, w_4, w_3, w_2, w_1)}{\text{count}(w_4, w_3, w_2, w_1)}$$

$$P(w_5, w_4, w_3, w_2, w_1) = P(w_1) \times P(w_2 \mid w_1) \times P(w_3 \mid w_1, w_2) \times P(w_4 \mid w_1, w_2, w_3) \times P(w_5 \mid w_1, w_2, w_3, w_4)$$

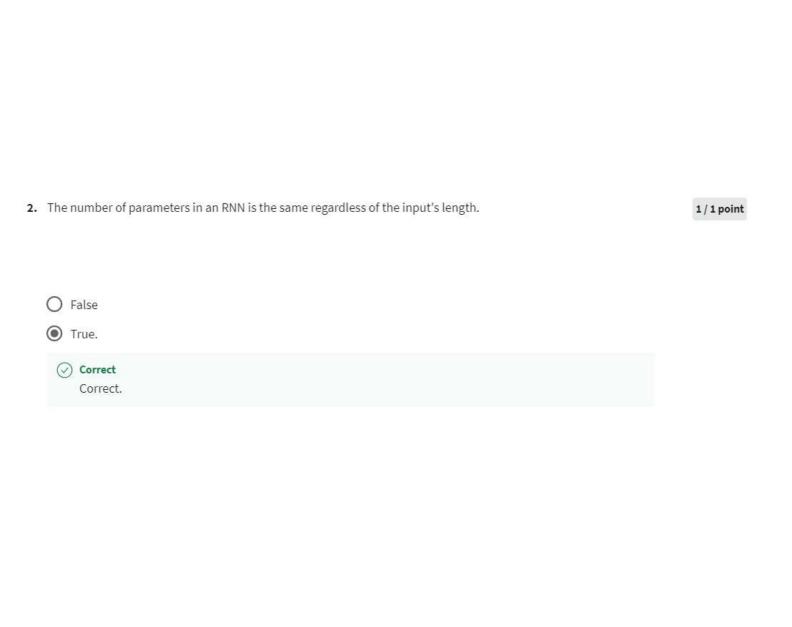
$$\bigcirc P(w_5, w_4, w_3, w_2, w_1) = P(w_1) \times P(w_2) \times P(w_3) \times P(w_4) \times P(w_5)$$

$$\bigcirc \ P\left(w_{5}, w_{4}, w_{3}, w_{2}, w_{1}\right) = P\left(w_{5} \mid w_{4}, w_{3}, w_{2}, w_{1}\right)$$

⊘ Correct

Correct.

1/1 point



3.	Select all the examples that correspond to a "many to one" architecture.	1 / 1 point
	☐ An RNN which inputs a sentiment and generates a sentence.	
	An RNN which inputs a sentence and determines the sentiment.	
	An RNN which inputs a topic and generates a conversation about that topic.	
	An RNN which inputs a conversation and determines the topic.	

4. What should be the size of matrix W_h , if $h^{< t>}$ had size 4x1 and $x^{< t>}$ 10x1?

 $h^{< t>} = g\left(W_h\left[h^{< t-1>}, x^{< t>}\right] + b_h\right)$

- 4x14
- O 14x4
- O 4x4
- O 14x14
 - Correct
 Correct.

1 / 1 point

6.	What problem, related to vanilla RNNs, do GRUs tackle?	1 / 1 point
	Loss of relevant information for long sequences of words.	
	Overfitting	
	O High computational time for training and prediction.	
	Restricted flow of information from the past to the present.	

7.	Bidirectional RNNs are acyclic graphs, which means that the computations in one direction are independent from the ones in the other direction.
	● True
	✓ False✓ Correct
	Correct.

Compared to Traditional Language models which of the following problems does an RNN help us with?
✓ Helps us solve memory issues.
They require almost no knowledge to use when compared to the traditional n-gram model.
They are much simpler to understand.
✓ Helps us solve RAM issues.

1 / 1 point

8.

9.	What type of RNN structure would you use when implementing machine translation?	1 / 1 point
	O Many to one	
	One to one	
	One to many	
	Many to Many	
	Correct.	

10. In the scan() function the variable cur_value corresponds to the hidden state in an RNN.

1/1 point

```
def scan(fn, elems, initializer=None, ...):
    cur_value = initializer
    ys = []
    for x in elems:
        y, cur_value = fn(x, cur_value)
        vs.append(y)
    return ys, cur_value
```

True

O False

Correct.