✓ **Congratulations! You passed!**

Grade received 100%   To pass 80% or higher

Go to next item

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

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

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

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

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

⊘ **Correct**
Correct.

2. The number of parameters in an RNN is the same regardless of the input's length.   **1 / 1 point**

◉ True.

○ False

⊘ **Correct**
Correct.

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.

⊘ **Correct**
Correct.

☐ An RNN which inputs a topic and generates a conversation about that topic.

☑ An RNN which inputs a conversation and determines the topic.

⊘ **Correct**
Correct.

4. What should be the size of matrix W_h, if $h^{<t>}$ had size 4x1 and $x^{<t>}$ 10x1?   **1 / 1 point**

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

◉ 4x14

○ 14x4

○ 4x4

○ 14x14

⊘ **Correct**
Correct.

5. In the next equation, why is there a division by the number of time steps but not one for the number of classification categories?   **1 / 1 point**

$J = -\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{K} y_j^{<t>} \log \hat{y}_j^{<t>}$

◉ Because there is just one value in every vector $y^{<t>}$ different from zero.

○ Because the equation is wrong.

○ Because this equation is given for a single example.

○ Because for most classification tasks there are only two categories.

⊘ **Correct**
Correct.

6. What problem, related to vanilla RNNs, do GRUs tackle?   **1 / 1 point**

◉ Loss of relevant information for long sequences of words.

○ Overfitting

○ High computational time for training and prediction.

○ Restricted flow of information from the past to the present.

⊘ **Correct**
Correct

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

1 / 1 point

- ◉ True
- ○ False

✓ **Correct**
Correct.

---

8. Compared to Traditional Language models which of the following problems does an RNN help us with?

1 / 1 point

- ☑ Helps us solve memory issues.

✓ **Correct**
Correct!

- ☐ 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.

✓ **Correct**
Correct!

---

9. What type of RNN structure would you use when implementing machine translation?

1 / 1 point

- ○ One to many
- ◉ Many to Many
- ○ Many to one
- ○ One to one

✓ **Correct**
Correct.

---

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

1 / 1 point

```python
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
- ○ False

✓ **Correct**
Correct.