

Hide menu

Neural Machine Translation

Lecture Notes (Notes)

Practice Quiz

Practice Quiz: Neural Machine Translation

Started

Assignment

Heroes of NLP: Oren Etzioni

Congratulations! You passed!

Grade received 88.89% To pass 80% or higher

Go to next item

Neural Machine Translation

1. Which of the following are bottlenecks in training sequence-to-sequence models?

1 / 1 point

Try again

☒ You are trying to store variable length sequences in a fixed memory, for example, you are trying to store articles of different lengths in a fixed 100 dimensional vector.

Correct

Correct

Receive grade

To Pass 80% or higher

Your grade

88.89%

View Feedback

We keep your highest score

☒ There are vanishing/exploding gradient problems.

Correct

Correct

Like

Dislike

Report an issue

☐ They require a lot of memory.

☐ They are not that useful

2. What are some of the benefits of using attention?

1 / 1 point

☐ It is significantly slower to use attention and therefore it is not recommended to use it.

☒ It allows you to focus on the parts that matter more.

Correct

Correct.

☐ The use of attention ends up giving you less accurate results.

☒ It helps with the information bottleneck issue.

Correct

Correct.

3. What are the major components in the attention mechanism that are required? Select all that apply.

1 / 1 point

☒ Values: not really described in lecture, but you can think of them just like the keys for now. (Hint: you need this for attention).

Correct

Correct.

☒ Keys: described in the lesson as the object you are looking for.

Correct

Correct.

☐ Cosine similarity.

☒ Queries: described in the lesson as the "ask" you are trying to match with the key.

Correct

Correct.

☒ Softmax

Correct

Correct. This gives you a distribution over the most important words at each time point when decoding.

4. Teacher forcing uses the actual output from the training dataset at time step $y^{(t)}$ as input in the next time step $X^{(t+1)}$, instead of the output generated by your model.

1 / 1 point

☒ True.

☐ False.

Correct

Correct.

5. The BLEU score's range is as follows:

1 / 1 point

☒ The closer to 0, the worse it is, the closer to 1, the better it is.

☐ The closer to 1, the worse it is, the closer to 0, the better it is.

☐ The closer to -1, the worse it is, the closer to 1, the better it is.

☐ The closer to $-\infty$, the worse it is, the closer to ∞ , the better it is.

Correct

Correct.

6. BLEU (Vanilla Implementation) is defined as:

0 / 1 point

☐ (Sum of unique unigram counts, overlapping in the candidate and reference) / (Total # of unigrams in the reference)

☐ (Sum of unique n-gram counts in the candidate) / (Total # of n-grams in the candidate)

☒ (Sum of unique unigram counts in the candidate) / (Total # of n-grams in the reference)

☐ (Sum of unique n-gram counts, overlapping in the candidate and reference) / (Total # of n-grams in the candidate)

Incorrect

Incorrect.

7. What is the difference between precision and recall in Rouge?

1 / 1 point

☒ Precision is defined as:

(Sum of overlapping unigrams in model and reference)/(total # of words in model)

Recall is defined as:

(Sum of overlapping unigrams in model and reference)/(total # of words in reference)

☐ Recall is defined as:

(Sum of overlapping unigrams in model and reference)/(total # of words in model)

Precision is defined as:

(Sum of overlapping unigrams in model and reference)/(total # of words in reference)

☐ Recall is defined as:

(Sum of unigrams in model and reference)/(total # of words in model)

Precision is defined as:

(Sum of overlapping unigrams in model and reference)/(total # of words in reference)

☐ Precision is defined as:

(Sum of overlapping bigrams in model and reference)/(total # of words in model)

Recall is defined as:

(Sum of overlapping bigrams in model and reference)/(total # of words in reference)

Correct

Correct.

8. Greedy decoding

1 / 1 point

☒ Allows you select the word with the highest probability at each time step.

☐ Allows you randomly select the word according to its own probability in the softmax layer.

☐ Selects multiple options for the best input based on conditional probability.

☐ Makes use of the Minimum Bayes Risk method.

Correct

Correct.

9. When implementing Minimum Bayes Risk method in decoding, let's say with 4 samples, you have to implement the following.

1 / 1 point

1. Calculate similarity score between sample 1 and sample 2

2. Calculate similarity score between sample 1 and sample 3

3. Calculate similarity score between sample 1 and sample 4

4. Average the score of the first 3 steps (Usually a weighted average)

5. Repeat until all samples have overall scores

Pick the golden one with the highest similarity score.

☒ True

☐ False

Correct

Correct.

