

Exercise 1

What are the encoder and decoder in a neural machine translation system?

Exercise 1: Solution

- Encoder: Reads the entire source sentence (input in the source language) and creates dense representations capturing its meaning.
- Decoder: Uses the encoder's dense representations and an to generate the translated sentence.

Exercise 2

Decide if the next statements about attention-based models are true or false, and elaborate on your responses.

(A) With an attention mechanism we no longer try encode the full source input into a fixed-length vector. Rather, we allow the decoder to “attend” to different parts of the input at each step of the output generation

(B) Attention has only successfully been applied to text sequences. Other applications such as image or speech have tested attention-based mechanisms but without success.

Exercise 2: Solution

(A) True

Attention avoids the limitation of compressing the input into one fixed vector. Instead, the decoder dynamically focuses on relevant parts of the full encoded input in sequence-to-sequence models.

(B) False

Attention has successfully been applied to tasks beyond text. Examples include computer vision and image processing (Vision Transformers), speech recognition (Speech Transformers) and others.