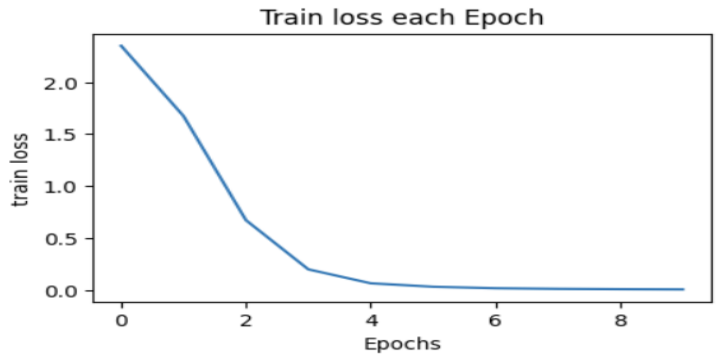
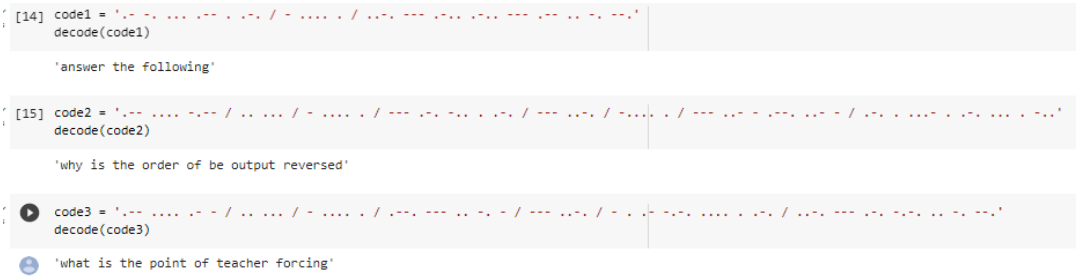
Question 1

```
def forward(self, src):
    out = self.embedding(src)
    out, (hidden_state, cell_state) = self.rnn(out)
    return hidden_state, cell_state
```



Question 2



Teacher forcing is using the ground truth as input to the recurrent neural network instead of model output from timestep t-1. For example, when the model produces the wrong output at timestep t, then the actual output would be fed into the network at timestep t+1 instead of the wrong output Applying teacher forcing allows faster convergence and improve model stability. By reversing the words in the source sentence, the average distance between the words in the source and target language remain same but the first few words in the source language are now very close to the first few words in the target language, so the problem's minimal time lag decreases [1]. Backpropagation can easily "establish communication" between the source sentence and the target sentence, which improves performance of the model [1].

Question 3

When the length of chunks is longer, the model is no longer able to decode the sequences to the correct words. The decoded words are not accurate. This is because the model was trained on shorter chunks of words, hence it was not able to generalize well with longer chunks input.

References

[1] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to Sequence Learning with Neural Networks," *arXiv.org*, 2014. https://arxiv.org/abs/1409.3215