

# Assignment4

April 27, 2020

## 1 Neural Machine Translation with RNNs

(g) The mask is used for setting padding words' weights(attention distribution score) to  $-\infty$ , i.e. it sets the impact of padding words in source sentence to minimal for decoder. It is useful because the decoder should only pay attention to those real words rather than padding words in source sentence.

## 2 Analyzing NMT Systems