HW7

1. After your model predicts the probability of answer span start/end position, what rules did you apply to determine the final start/end position? (the rules you applied must be different from the sample code)

   (1) Add limit start_index < end_index
   (2) end_index – start_index < 40 in order to make sure that the answer is probable.

2. Try another type of pretrained model which can be found in huggingface's Model Hub (e.g. BERT -> BERT-wwm-ext, or BERT -> RoBERTa), and describe

   - the pretrained model you used
   - performance of the model you used
   - the difference between BERT and the pretrained model you used (architecture, pretraining loss, etc.)

   Model : Chinese-bert-wwm-ext

   Performance: both pretraining loss and downstream task performance(which is this HW) is slightly better than BERT

   Architecture difference: wwm means whole world masking, its improvement is to assure that a whole world but not only a single character in a word is masked.