Question 5.

1. Why is a non-linear activation function (e.g. ReLU) necessary after a fully-connected layer in a CNN? What would happen if we omitted it?
   Ans: because non-linear activation function can help to make the model more complex, to achieve better performance. If we use linear function, then the number of layers may not be important since they are just linear transformations. If omitted, the model will lose complexity, and lead to poor performance.

2. What is the primary purpose of a max-pooling layer in a CNN? How does it differ from an average-pooling layer?
   Ans: max-pooling layer is to get the maximum value in the pooling window. It can reduce the size of inputs. In this way, it can reduce the computation costs, and highlight the most important features, and avoid noises. The difference between max-pooling and average-pooling is that average-pooling takes the average of values in the window.

3. Why is cross-entropy loss preferred over mean squared error (MSE) for classification tasks?
   Ans: cross-entropy will directly measure the difference between outputs and labels. MSE will measure the mean squared error. For the classification tasks, it is discrete task to find the correct class, cross-entropy can better measure the class probability, then use gradient descent to get the correct class.

4. Explain why combining log(softmax(logits)) directly might lead to numerical instability. How is this addressed in practice?
   Ans: using softmax will take exponential of the logits, might lead to infinite values. Taking log of a small number will also lead to negative infinite values. Solution is that we can use log_softmax function to compute the values.

5. During training, if the training loss decreases but the validation loss increases, what might be happening and how can it be addressed?
   Ans: overfitting is happening, high accuracy in training data, and low accuracy in unseen data. Solutions are we can reduce the model complexity, add a regularization term, more training data, and data augmentation.