**Question 1**

First prove that

If , then optimal > 0. If , then optimal < 0. [Optimal refers to the value such that C is minimized]

When , the minimal value C can get is greater than 0 if < 0.

is monolithic increasing. When w < 0, ,

< . C is always greater than 0. When w > 0, , > . C is less than 0. Therefore, if , then optimal is greater than 0. The second half can be proved similarly.

Scenario 1: x >

After step 2: w will be assigned a value such that or clipped 1 if is outside the range [-1, 1]. (w is greater than 0)

Scenario 1.1 z > 0

therefore, after step 3 will increase until .

After running step 2 again, will be less than 0 since .

, will decrease until .

Then w will be assigned a positive value and will increase until again. The program will never converge.

Scenario 1.1 z < 0

therefore, after step 3, will decrease. , thus . will decrease until , which is equivalent to . This is the same as Scenario 1.1.

Scenario 2: x <

This is the same as scenario 1 and can be proved in the same manner.

Therefore, the algorithm does not converge.

**Question 2**

In order to approximate the Lipschitz function, the weights of the neural network are constrained to some compact space. That is w is clipped after each iteration. This means neural network cannot approximate all the Lipschitz function, and the w obtained from line 3 to 8 may not be correct. The network may not be able to approximate the function that produces max .

Training time is long as well.