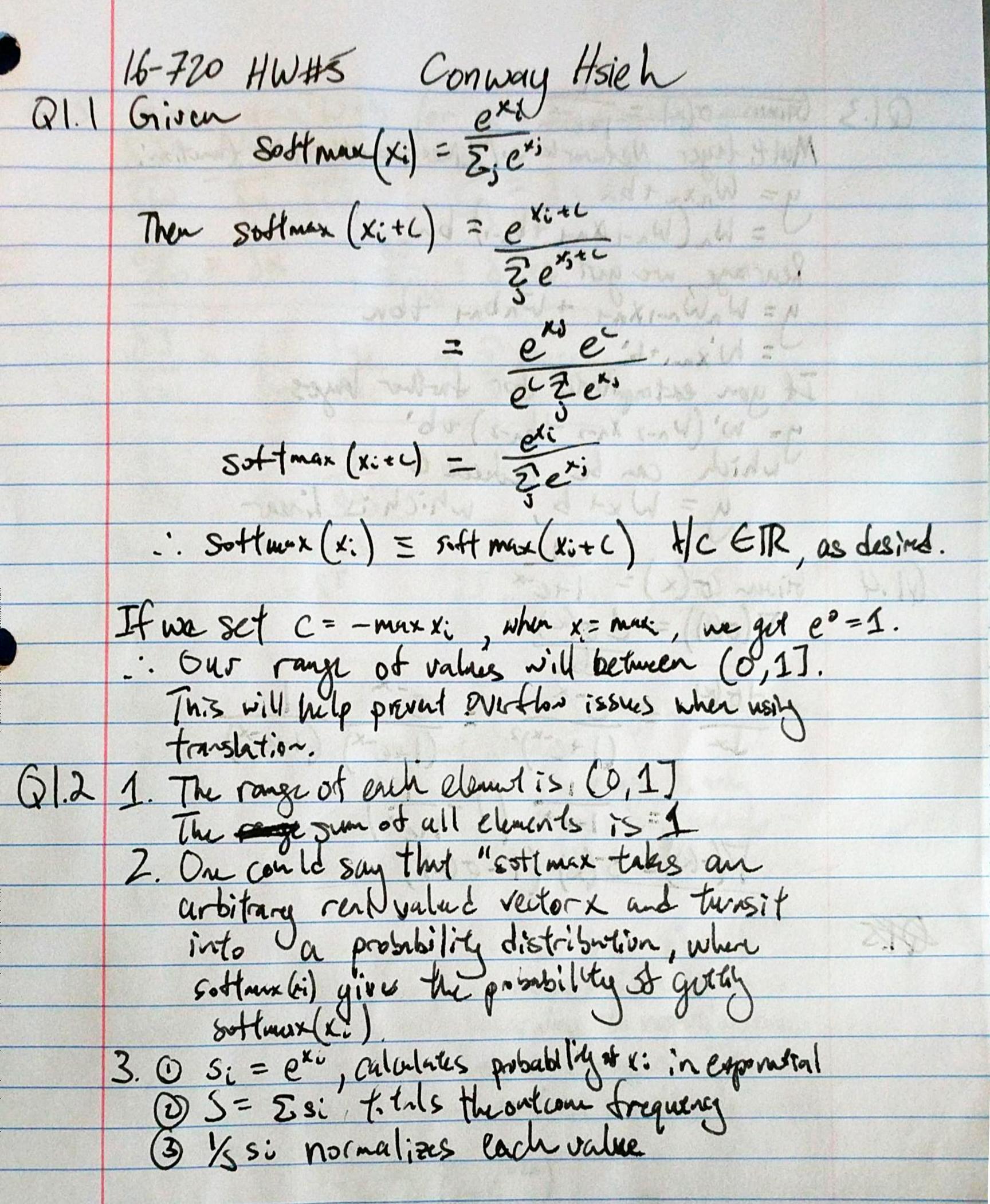
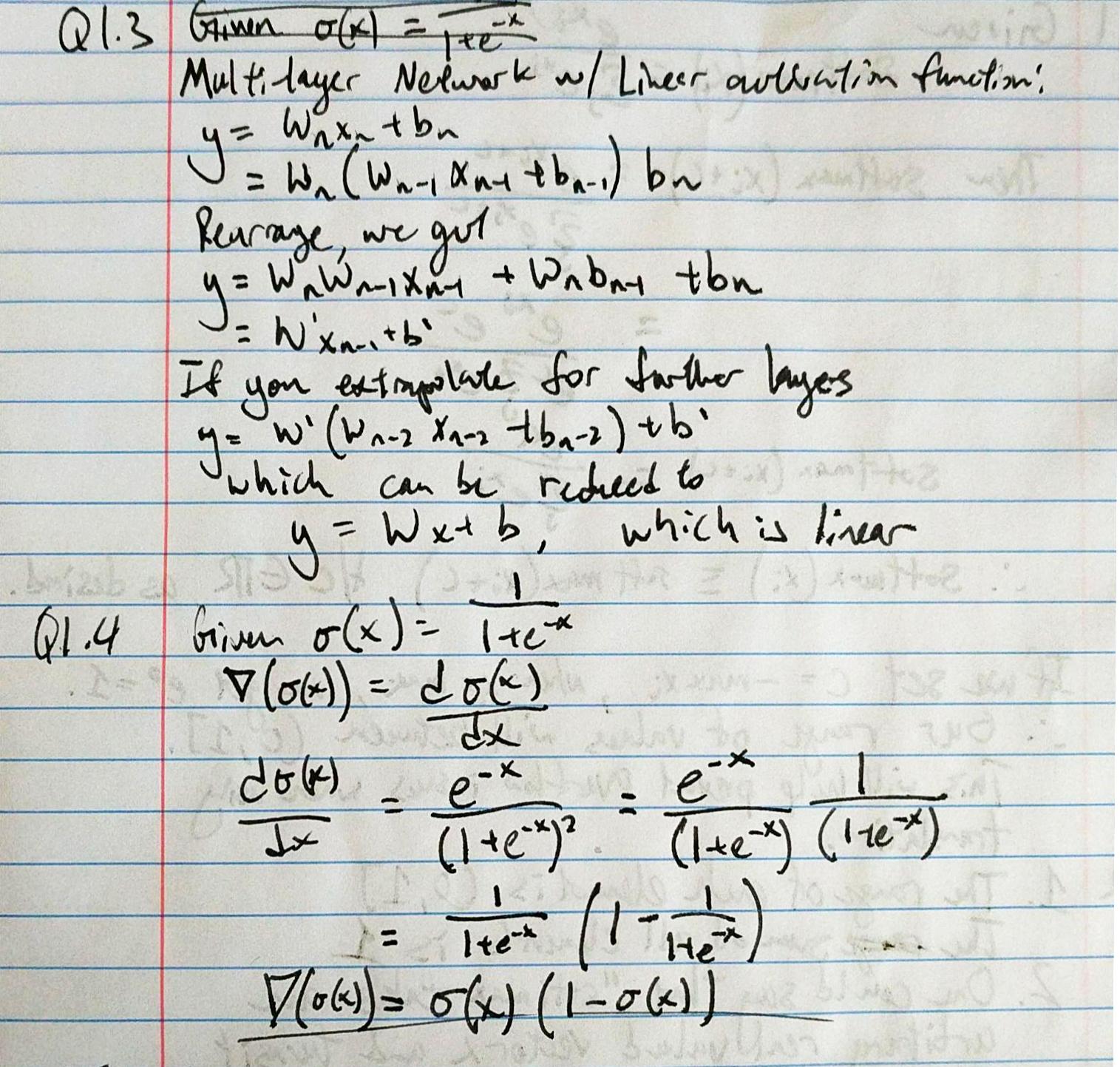
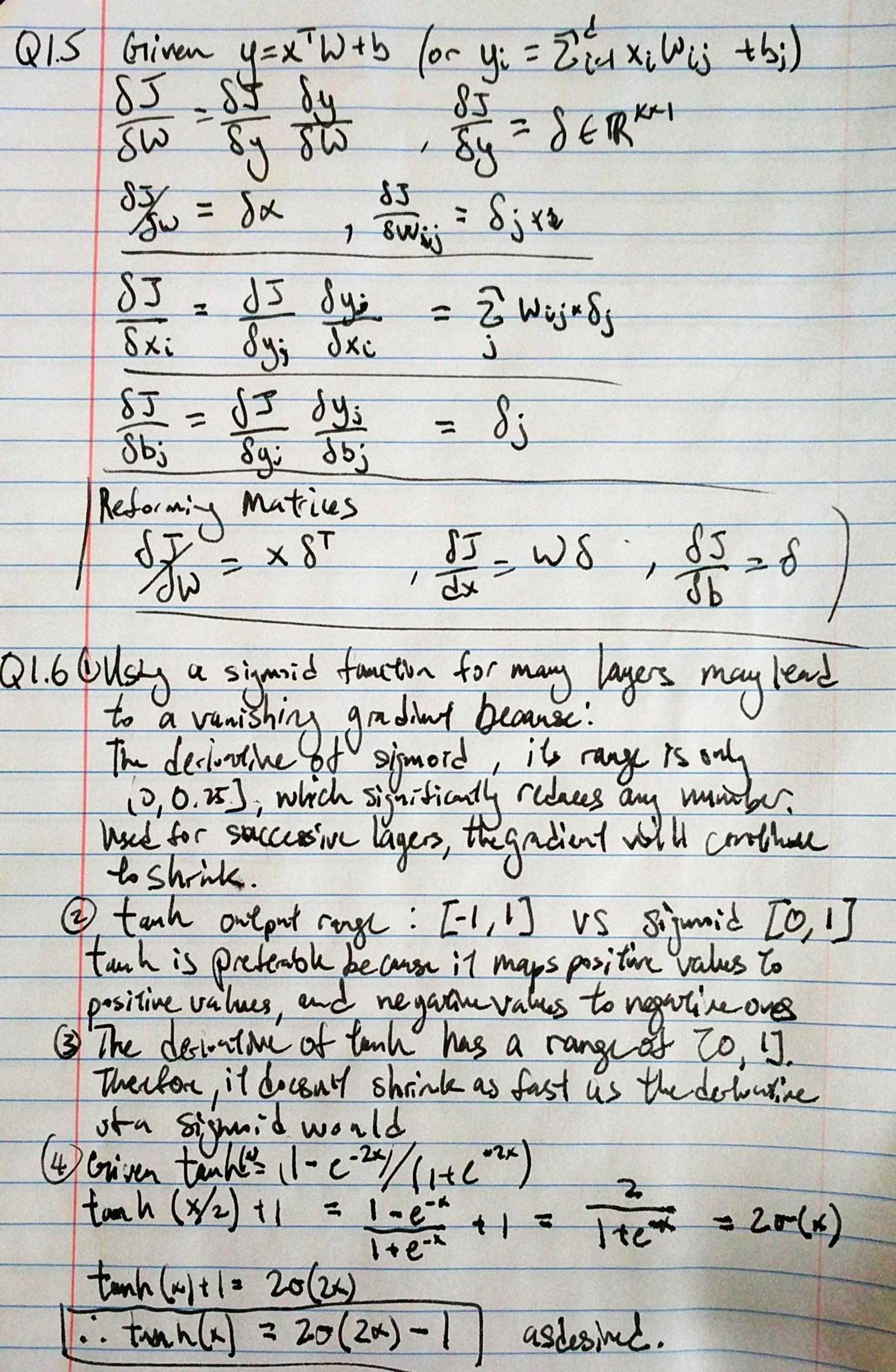
Q1







Q 2.1.1

If the weights are initialized to zero, whenever you try to propagate, you multiply weights by delta, resulting in net zero change in weights. Therefore, a zero-initialized network can only output zeros, and never learn anything.

Q 2.1.3

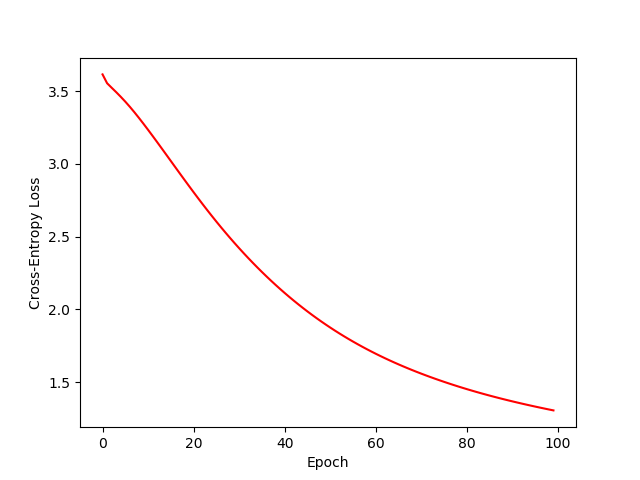
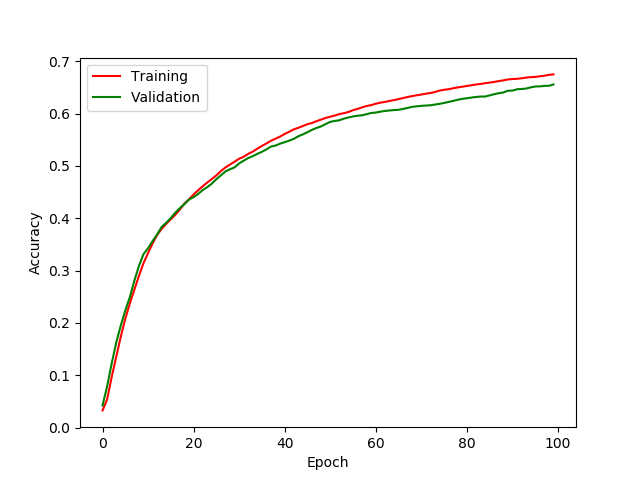
Initialization is done with random numbers allows the network to probe different parts of the solution space to find the best solution. If you always start with the same weights, the training and gradients will remain similar, limiting the scope of your solution.

Scaling the initialization based on layer size is done because when consecutive layers have the same dimension, the average activation variance that is conserved increases. This allows for more information to continue through the network.

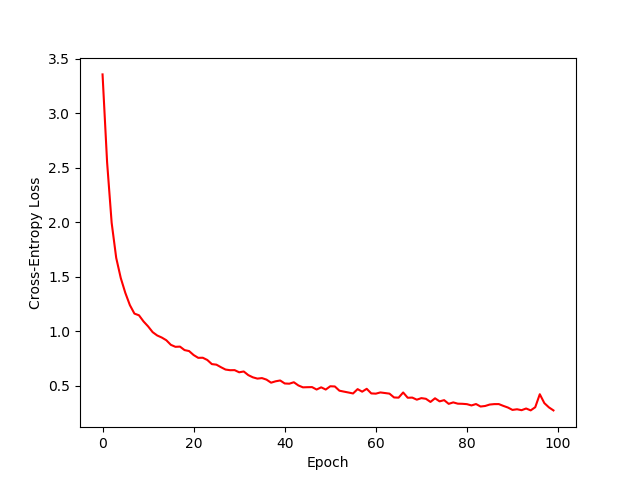
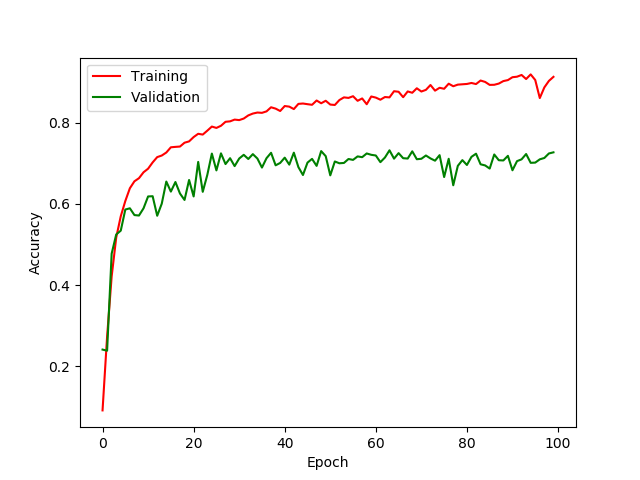
Q 3.1.2

Final validation accuracy of best set: 76%

10x less learning rate:

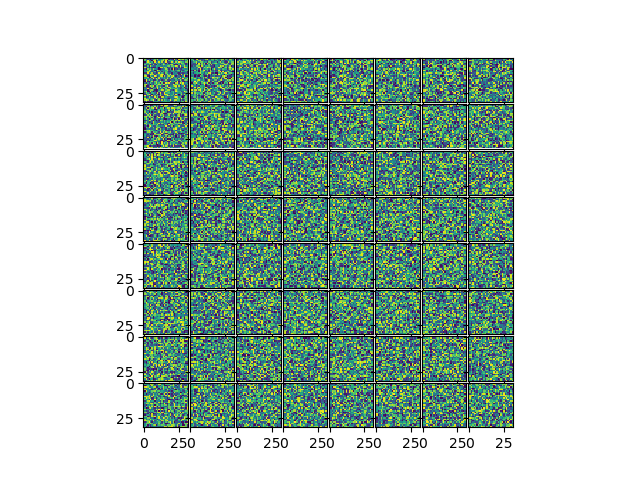


10x more learning rate:

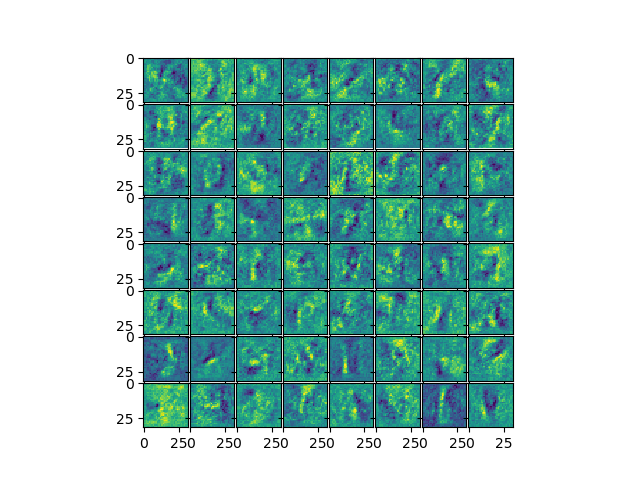


More learning rate causes more abrupt, jagged changes to weights/ training while less learning rate is smooth. Also, achieves less accuracy in same number of iterations (because learns more slowly)

Q3.1.3



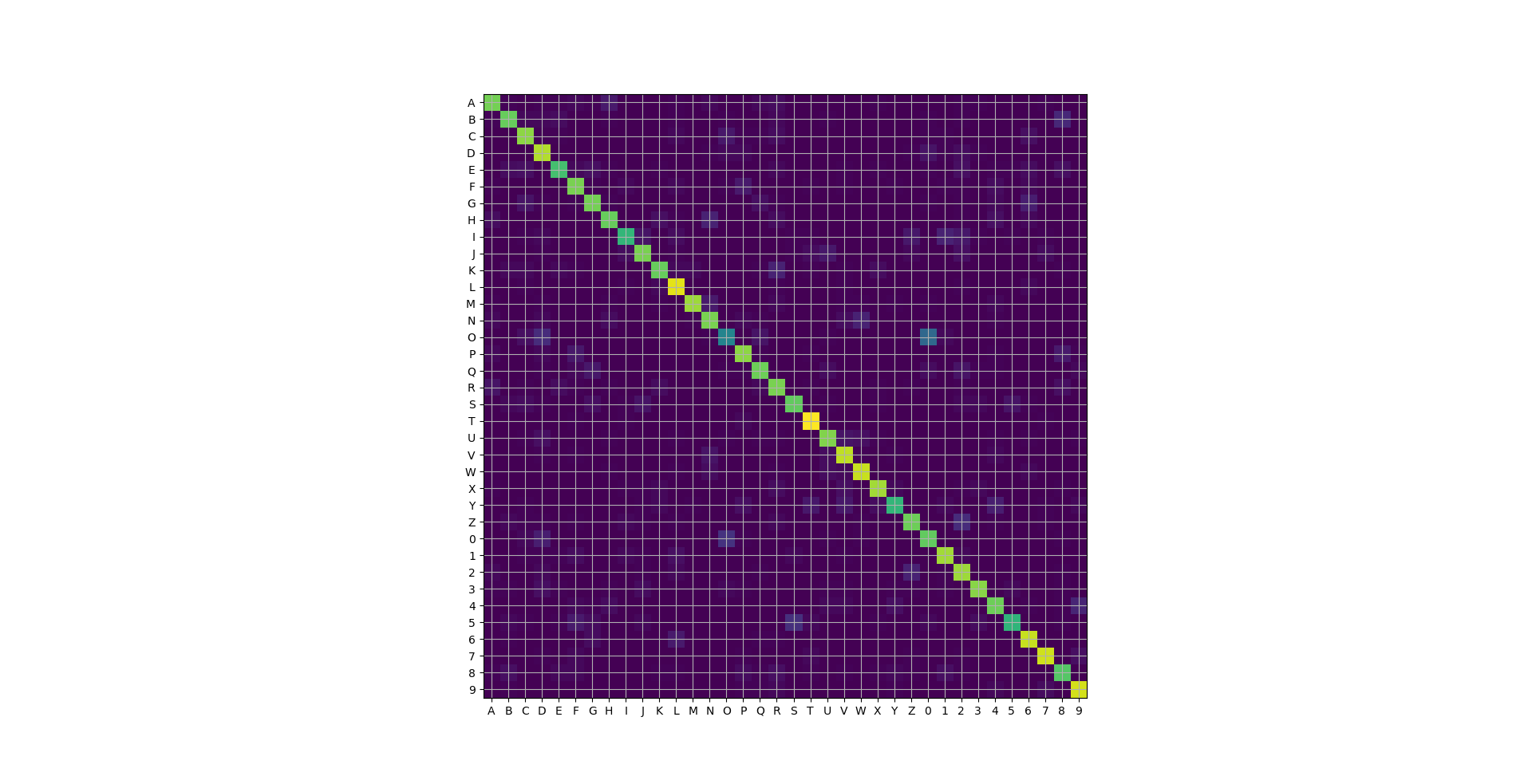
Initialization



After Training

Initialized weights look random, as expected. After training, structures begin to emerge within the weights, which do not look like noise as it did at initialization.

Q 3.1.4



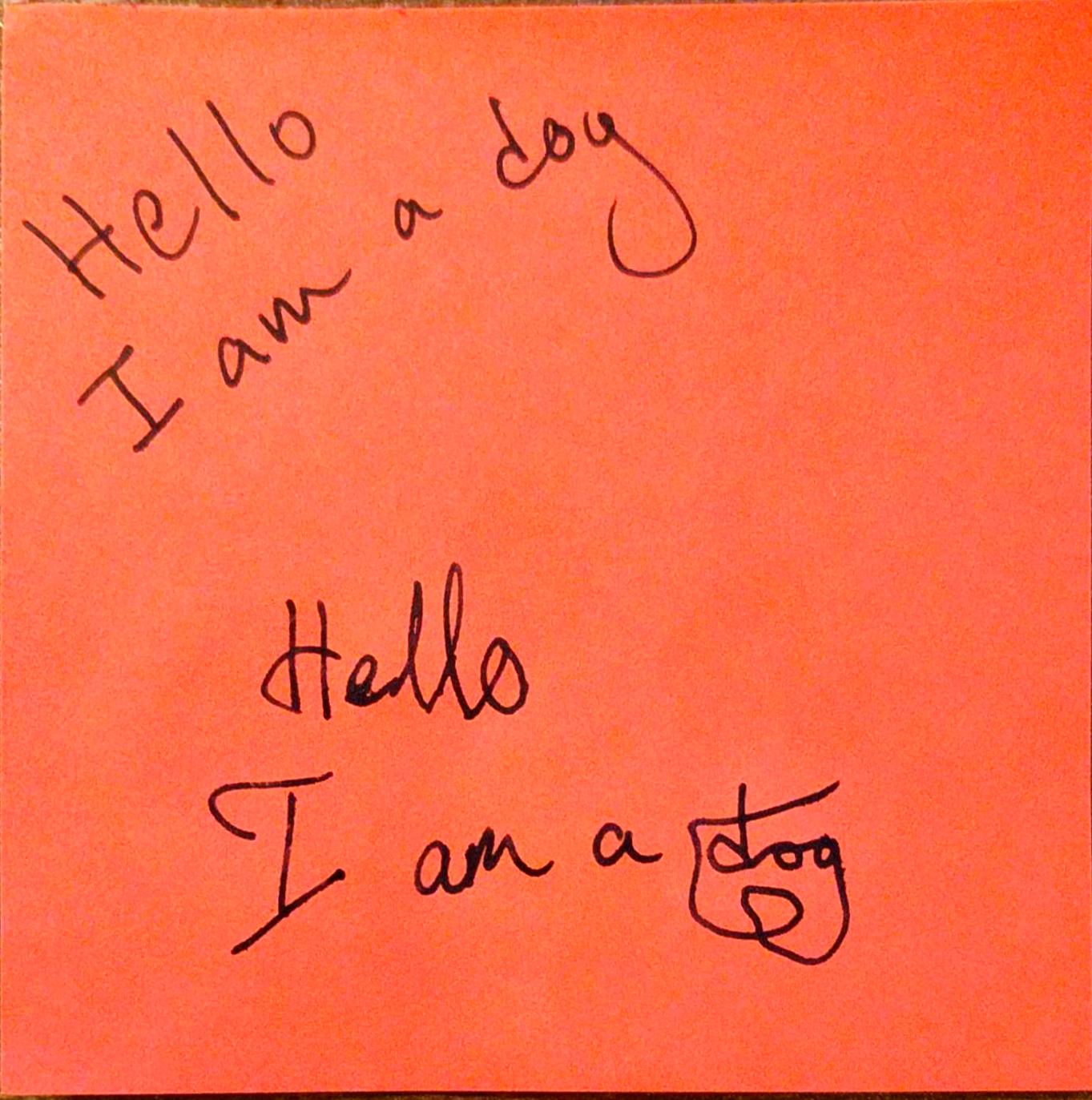
O and 0 seems to be the most prominent case of misclassification. Others include 5 and S, 2 and Z, and Y and 4. These seem reasonable, as due to differences in handwriting, they have similar structures.

Q4.1

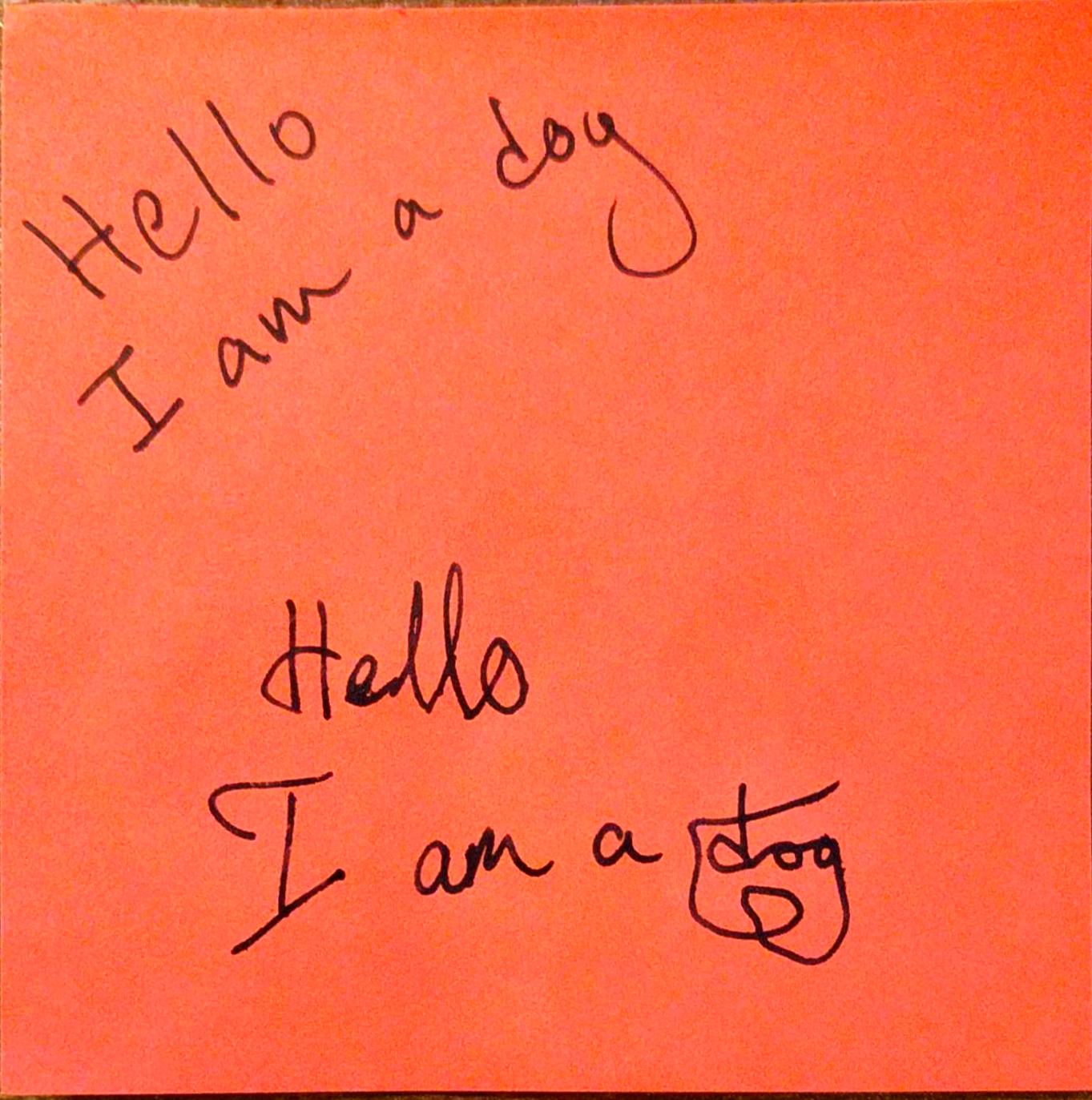
One assumption is that the words are in left to right, top to bottom format. This means that if any rotations are made, the classification order will be incorrect.

Another assumption is that all the letters are fully connected when written, and separate letters are not connected.

Lastly, an assumption is that all letters are of similar size, so anything too small will be considered as noise.

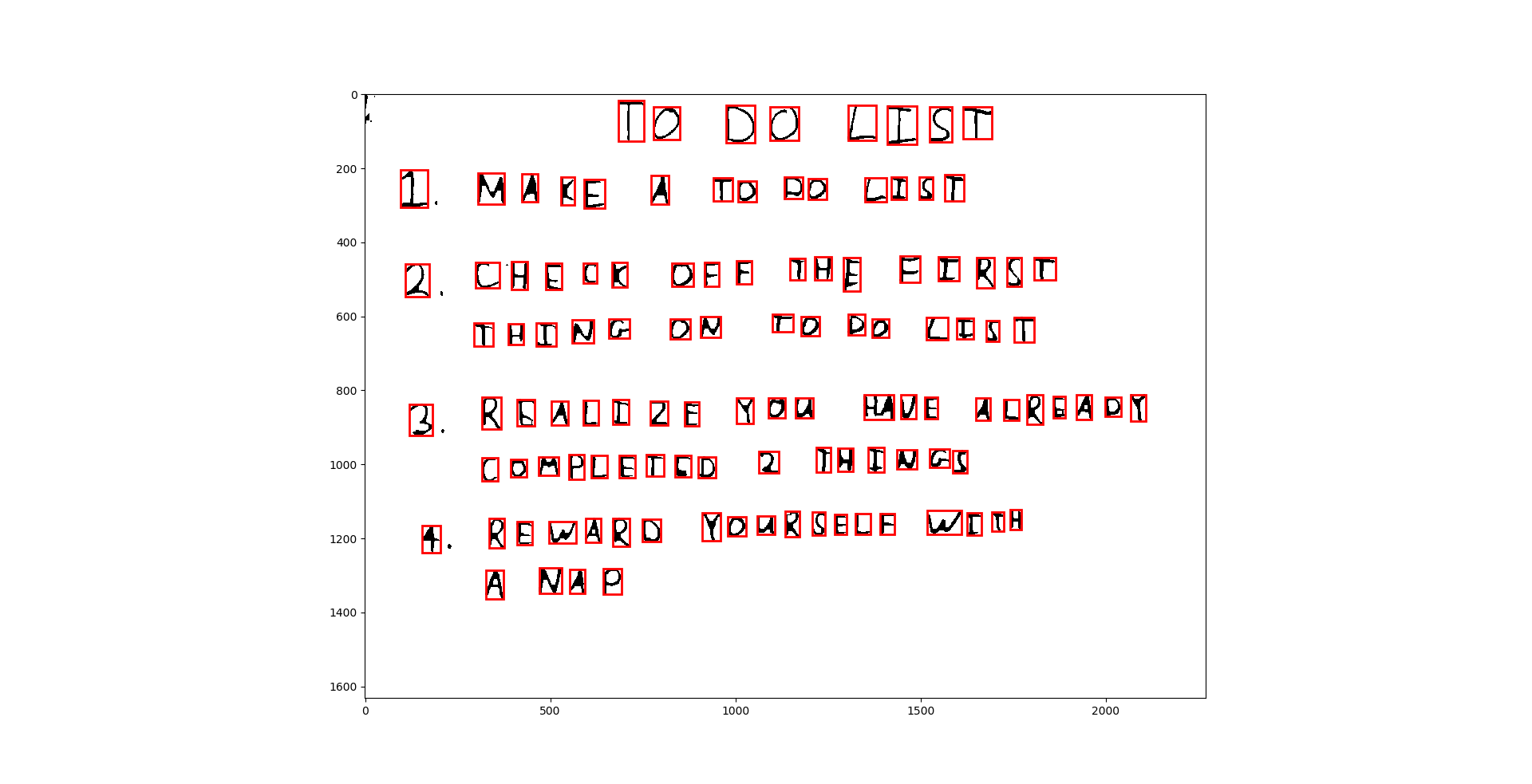


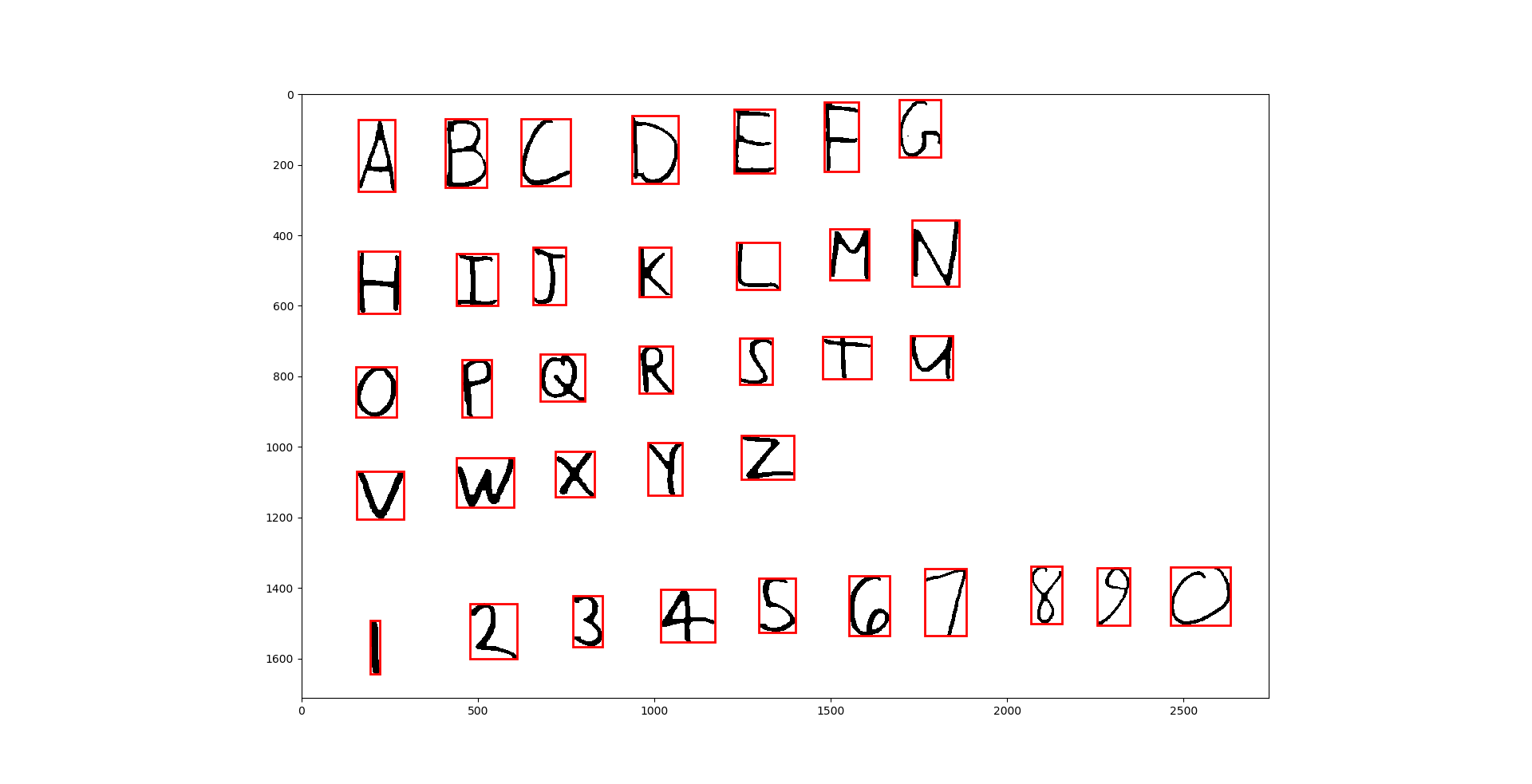
Rotation will cause issues during classification

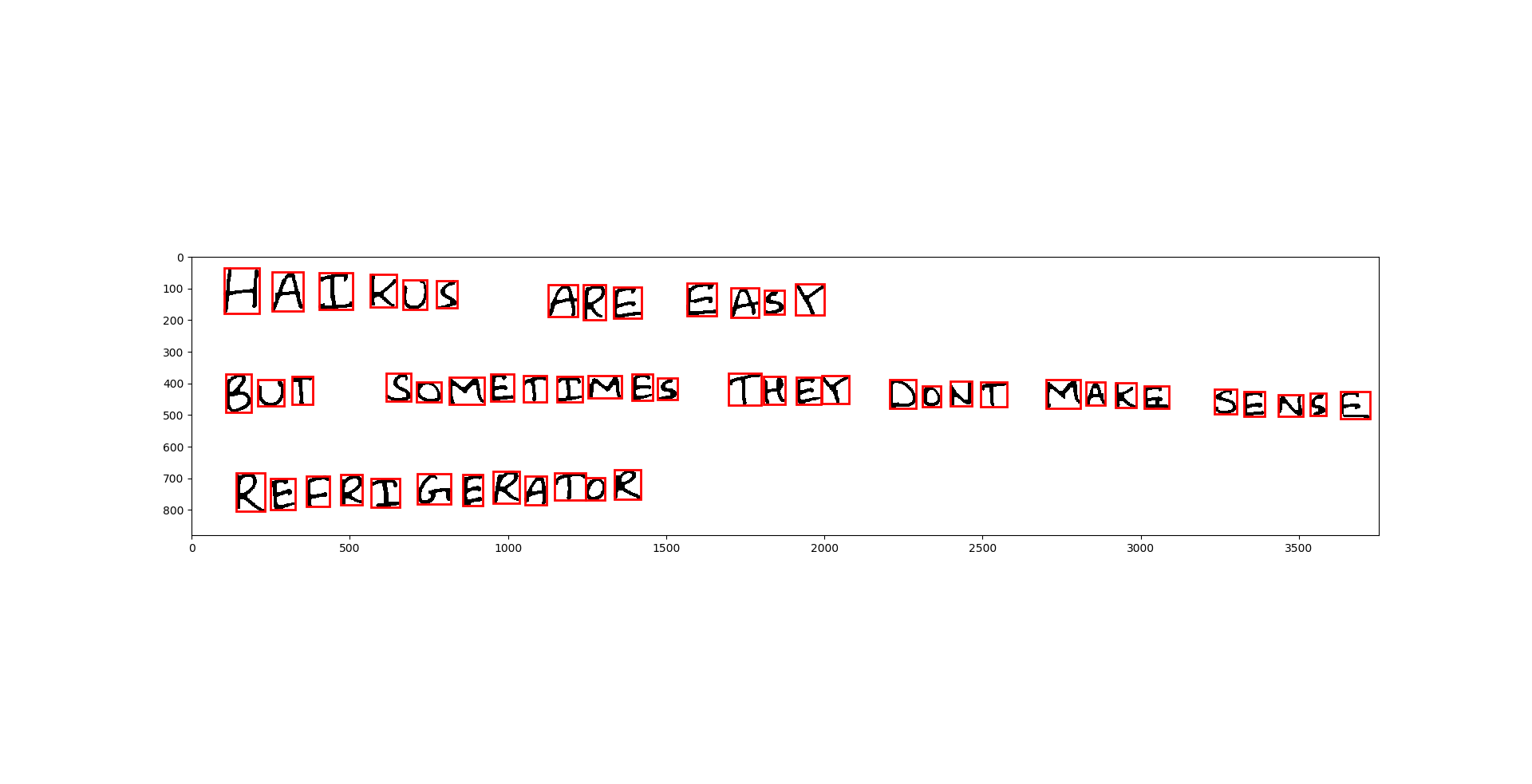


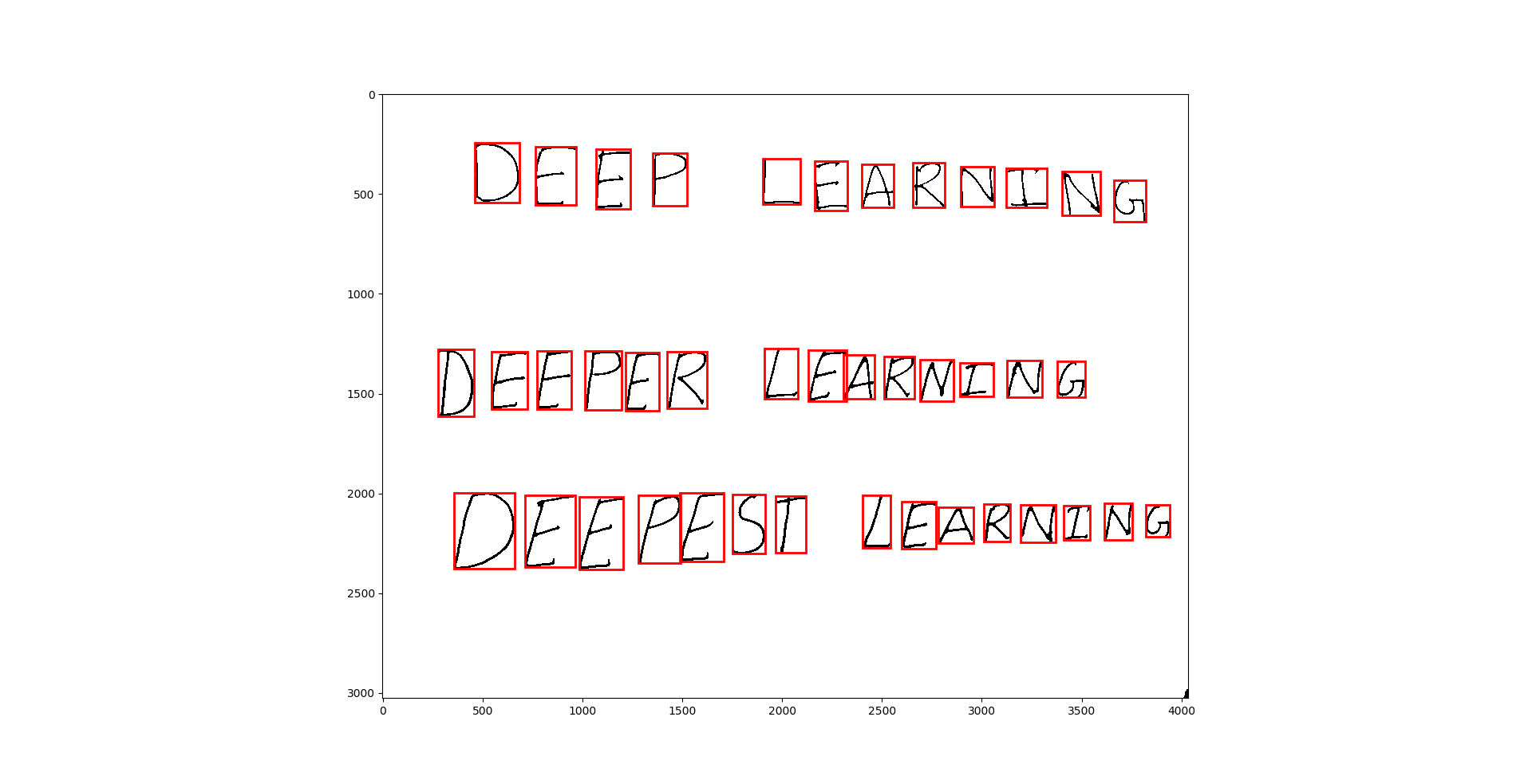
Connected letters will cause issues

Q 4.3









Q 4.4

F0 D0 LI5T

I MAKE R T0 P0 LI5T

2 CH5CK 0FR THR TIRST THING 0N T0 D0 LI5T

3 RRALIZE YO4 M4RALRE A0T COMPLETRD 2 THIN4S

4 REWARD F0URSELR WITR A NRP

ABCDEFG

HIJKLMN

0PQRST4

VWXYZ

1Z345G787Q

HAIKU5 ARR EA5Y

BUF SOMRFIMRS TRRY 0ONT MAKR SR45R

RRFRI6ERAT0R

DEEP LEARMINQ

DEEFER LEARN2NG

QEEPE5T LEARNING