Problem Set 2 Shallow and Deep Networks

DS542 DL4DS

Spring, 2025

Sicheng Yi

Note: Refer to the equations in the *Understanding Deep Learning* textbook to solve the following problems.

Problem 3.2

For each of the four linear regions in Figure 3.3j, indicate which hidden units are inactive and which are active (i.e., which do and do not clip their inputs).

Answer: Refer to Figure 1 at the end
Region 1: only h3 (i) is activated
Region 2: h1 (g) and h3 (i) are activated
Region 3: h1 (g), h2 (h), and h3 (i) are all activated
Region 4: h1 (g) and h2 (h) are activated

Problem 3.5

Prove that the following property holds for $\alpha \in \mathbb{R}^+$:

$$\text{ReLU}[\alpha \cdot z] = \alpha \cdot \text{ReLU}[z].$$

This is known as the non-negative homogeneity property of the ReLU function.

Answer:
ReLu is 0 when input z is negative, otherwise z when non-negative.

$$ReLU[z] = \max(0, z)$$

$$ReLU[a \cdot z] = \max(0a, az) = a\max(0, z)$$

$$ReLU[a \cdot z] = aReLU[z]$$

Problem 4.6

Consider a network with $D_i = 1$ input, $D_o = 1$ output, $K = 10$ layers, and $D = 10$ hidden units in each. Would the number of weights increase more – if we increased the depth by one or the width by one? Provide your reasoning.

Answer:
Input: $D_i = 1$, Output: $D_o = 1$,
Layers: $K = 10$, 9 transitions between each hidden layers
Hidden Units in each layer : $D = 10$

Total Parameters $n = 10 + 10 \times 10 \times 9 + 10 = 920$, if not including the bias terms

Increase depth by 1, 11 layers in total, 10 transitions between each hidden layers
Total Parameters $n = 10 + 10 \times 10 \times 10 + 10 = 1020$, if not including the bias terms
Increase of weights is $1020 - 920 = 100$

Increase width by 1, each layer now has 11 hidden units
Total Parameters $n = 11 + 11 \times 11 \times 9 + 11 = 1111$, if not including the bias terms
Increase of weights is $1111 - 920 = 191$

$100 < 191$, depth ¡ width
Therefore, number of weights will increase more if increase width by 1, over increase depth by 1.

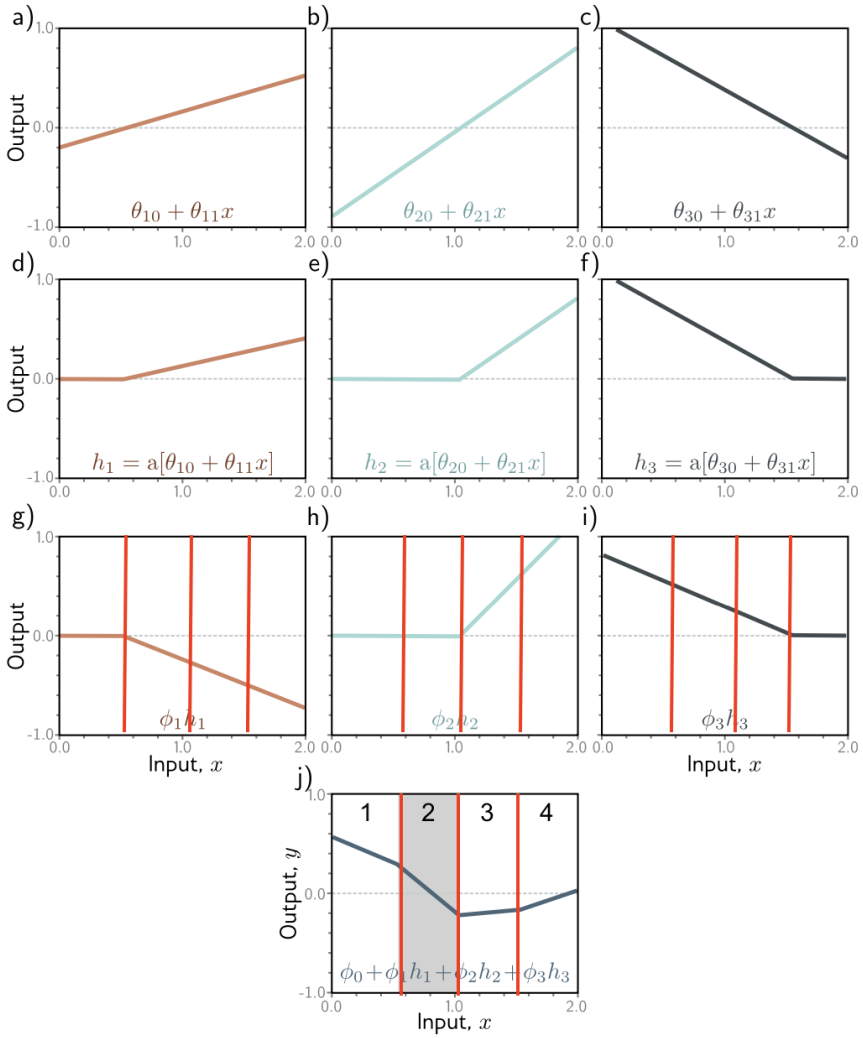

Figure 3.3 Computation for function in figure 3.2a. a–c) The input x is passed through three linear functions, each with a different y-intercept $\theta_{\bullet 0}$ and slope $\theta_{\bullet 1}$. d–f) Each line is passed through the ReLU activation function, which clips negative values to zero. g–i) The three clipped lines are then weighted (scaled) by ϕ_1, ϕ_2, and ϕ_3, respectively. j) Finally, the clipped and weighted functions are summed, and an offset ϕ_0 that controls the height is added. Each of the four linear regions corresponds to a different activation pattern in the hidden units. In the shaded region, h_2 is inactive (clipped), but h_1 and h_3 are both active. (Interactive figure)

Figure 1: problem 3.2