Problem Set 2 – Shallow and Deep Networks

DS542 – DL4DS

Spring, 2025

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).

1. Region 1: $x <\approx 0.5$ — h_3 is active, others are inactive
2. Region 2: $0.5 < x < 1$ — h_1 and h_3 are active, h_2 is inactive
3. Region 3: $1 < x < 1.5$ — all are active
4. Region 4: $x > 1.5$ — h_1 and h_2 are active, h_3 is inactive

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.

1. Recall: $\text{ReLU}[x] = \max(0, x)$
2. Left side: $\text{ReLU}[\alpha \cdot z]$
$$\text{ReLU}[\alpha \cdot z] = \max(0, \alpha \cdot z)$$
3. Right side: $\alpha \cdot \text{ReLU}[z]$
$$\alpha \cdot \text{ReLU}[z] = \alpha \cdot \max(0, z)$$
4. Since $\alpha > 0$, we can use the property that for any positive scalar α:
$$\max(0, \alpha \cdot z) = \alpha \cdot \max(0, z)$$

Therefore:
$$\text{ReLU}[\alpha \cdot z] = \max(0, \alpha \cdot z) = \alpha \cdot \max(0, z) = \alpha \cdot \text{ReLU}[z]$$

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.

1. Current network:
   (a) First layer: 1×10 weights (input to first hidden)
   (b) Middle layers: 9 layers of 10×10 weights
   (c) Final layer: 10×1 weights (last hidden to output)
2. Adding depth (K = 11):
   (a) Adds one new 10×10 matrix = 100 new weights
3. Adding width (D = 11):
   (a) First layer changes from 1×10 to $1 \times 11 = 1$ new weight
   (b) Middle layers change from 10×10 to $11 \times 11 = 21$ new weights per layer $\times$ 9 layers
   (c) Final layer changes from 10×1 to $11 \times 1 = 1$ new weight
   (d) Total new weights: $1 + (21 \times 9) + 1 = 191$ new weights

Therefore, increasing the width by one adds more weights (191) than increasing the depth by one (100).