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).

Answer:

1. Linear Region 1: h_1 and h_2 are inactive while h_3 is active.
2. Linear Region 2: h_2 is inactive while h_1 and h_3 are active.
3. Linear Region 3: All hidden units (h_1, h_2, and h_3) are active.
4. Linear Region 4: h_3 is inactive while h_1 and h_2 are active.

Problem 3.5

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

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

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

Answer:
To prove the non-negative homogeneity property of the ReLU function, we need to show that:

$$\mathrm{ReLU}[\alpha \cdot z] = \alpha \cdot \mathrm{ReLU}[z]$$

for $\alpha \in \mathbb{R}^+$ and $z \in \mathbb{R}$.

1. Definition of the ReLU function:

   The ReLU function, denoted as $\mathrm{ReLU}(z)$, is defined as:

$$\mathrm{ReLU}(z) = \begin{cases} z & \text{if } z \geq 0, \\ 0 & \text{if } z < 0. \end{cases}$$

2. Consider the case $z \geq 0$.

   If $z \geq 0$, then by the definition of the ReLU function:

$$\mathrm{ReLU}(z) = z.$$

   Evaluating both sides of the equation:

$$\mathrm{ReLU}[\alpha \cdot z] = \alpha \cdot z \quad (\text{since } \alpha \cdot z \geq 0 \text{ for } \alpha > 0 \text{ and } z \geq 0),$$

   and

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

   Thus, for $z \geq 0$, we have:

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

3. Consider the case $z < 0$.

   If $z < 0$, then by the definition of the ReLU function:

$$\mathrm{ReLU}(z) = 0.$$

   Evaluating both sides of the equation:

$$\mathrm{ReLU}[\alpha \cdot z] = 0 \quad (\text{since } \alpha \cdot z < 0 \text{ for } \alpha > 0 \text{ and } z < 0),$$

   and

$$\alpha \cdot \mathrm{ReLU}[z] = \alpha \cdot 0 = 0.$$

   Thus, for $z < 0$, we have:

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

4. Conclusion:

   In both cases, whether $z \geq 0$ or $z < 0$, it has been observed that:

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

   Therefore, the non-negative homogeneity property of the ReLU function holds for all $\alpha \in \mathbb{R}^+$ and $z \in \mathbb{R}$.

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: The number of weights would increase more if we increased the width by one compared to the depth by one. We can prove this by considering the 2 cases:

1. Case 1: Increasing the depth by one

   This will lead to $K = 11$ layers with $D = 10$ hidden units each. The weights (excluding the bias terms) $= (D_i \times D) + (K-1)(D \times D) + (D \times D_o) = (1 \times 10) + 10(10 \times 10) + (10 \times 1) = 1020$

2. Case 2: Increasing the width by one

   This will lead to $K = 10$ layers with $D = 11$ hidden units each. The weights (excluding the bias terms) $= (D_i \times D) + (K-1)(D \times D) + (D \times D_o) = (1 \times 11) + 9(11 \times 11) + (11 \times 1) = 1111$. Therefore, increasing the width by one results in more weights compared to increasing the depth by one.