

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. First linear region: h_1 inactive, h_2 inactive, h_3 active
2. Second linear region: h_1 active, h_2 inactive, h_3 active
3. Third linear region: h_1 active, h_2 active, h_3 active
4. Fourth linear region: h_1 active, h_2 active, h_3 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. **ReLU Definition:**

$$\text{ReLU}(z) = \max(0, z),$$

2. **Case 1:** $z \geq 0$

- When $z \geq 0$, $\text{ReLU}(z) = z$
- Substituting $\alpha \cdot z$ into the ReLU:

$$\text{ReLU}(\alpha \cdot z) = \max(0, \alpha \cdot z).$$

- Since $\alpha \geq 0$ and $z \geq 0$:

$$\text{ReLU}(\alpha \cdot z) = \alpha \cdot z.$$

- substitute:

$$\alpha \cdot \text{ReLU}(z) = \alpha \cdot z.$$

- Answer:

$$\text{ReLU}(\alpha \cdot z) = \alpha \cdot \text{ReLU}(z).$$

3. Case 2: $z < 0$

- When $z < 0$, $\text{ReLU}(z) = 0$.

- Substituting:

$$\text{ReLU}(\alpha \cdot z) = \max(0, \alpha \cdot z).$$

- Since $\alpha \geq 0$ and $z < 0$:

$$\text{ReLU}(\alpha \cdot z) = 0.$$

- substitute:

$$\alpha \cdot \text{ReLU}(z) = \alpha \cdot 0 = 0.$$

- Answer:

$$\text{ReLU}(\alpha \cdot z) = \alpha \cdot \text{ReLU}(z).$$

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.

$$\text{Total weights} = \sum_{k=0}^K (\text{weights in } \Omega_k + \text{biases in } \beta_k).$$

$$\text{Total weights} = (D_1 \cdot D_i) + \sum_{k=1}^{K-1} (D_{k+1} \cdot D_k) + (D_o \cdot D_K) + \sum_{k=1}^K D_k.$$

Simplifying since all layers have the same width ($D_k = D$):

$$\text{Total weights} = D \cdot (D_i + D_o) + (K - 1) \cdot D^2.$$

Answer: If we increase the width by one ($D \rightarrow D + 1$), it results in a larger increase in the number of weights compared to just increasing the depth. This is because width affects all parts of the network, as you can see clearly in the equation.

Increasing the depth only results in an extra D^2 weights. However, increasing the width by 1 results in:

$$\text{Total weights} = (D + 1) \cdot (D_i + D_o) + (K - 1) \cdot (D + 1)^2.$$

Which is a more significant increase compared to depth, since it affects throughout the network.