Math 6373 – Spring 2024 Name: SOLUTION

Homework #1

Please, write clearly and justify your work to receive credit.

(1) [6 Pts] *The following problem is about calculating the number of parameters of a feedforward neural network.*

(a) Derive a formula to compute the number of parameters between 2 (fully connected) layers of a feedforward neural network where the first layer has N_ℓ neurons and the second layer has $N_{\ell+1}$ neurons.

(b) Determine the number of parameters of a feedforward neural network with the following architecture:

- *Input layer: 3 neurons.*
- *Hidden layer 1: 12 neurons.*
- *Hidden layer 2: 8 neurons.*
- *Hidden layer 3: 4 neurons.*
- *Output layer: 2 neurons.*

(c) Design a feedforward neural network with input dimension 2, output dimension 1 and at least 2 hidden layers which contains at least 400 but no more than 405 parameters.

Solution:

(a) The number of parameters between 2 fully connected layers is

$$(N_\ell + 1)\, N_{\ell+1}$$

where the +1 term in the equation takes into account the bias terms.

(b) Using the formula above we have

$$\text{Total parameters } = 4 * 12 + 13 * 8 + 9 * 4 + 5 * 2 = 198$$

(c) There are many possible solutions. We can choose a MLP with 10 hidden layers of width 6. Then, since input dimension is 2 and output dimension is 1, we have

$$\text{Total parameters } = 3 * 6 + 9 * (7 * 6) + 7 * 1 = 403$$

(2) [6 Pts] *This problem is about using ReLU feedforward neural networks to implement piecewise linear functions.*

(a) Design a shallow neural network with ReLU activation function implementing the following function

$$T(x) = \begin{cases} 3x & \text{if } 0 \le x < \frac{1}{3} \\ \frac{3}{2}(1-x) & \text{if } \frac{1}{3} \le x \le 1 \end{cases} \qquad x \in [0,1],$$

(b) Design a shallow neural network with ReLU activation function implementing a piecewise linear function on $\mathbb{R}$ with 2 nodes, that is, a piecewise linear function that changes slope exactly 3 times.

(c) Derive a formula relating the number of parameters M of a shallow neural network with ReLU activation function implementing a piecewise linear function on $\mathbb{R}$ to the number of nodes K of the function.

Solution:

(a) Similar to the example presented in class, we can write the function T in [0,1] as

$$T(x) = 3(x-0)_+ - \frac{9}{2}(x-\frac{1}{3})_+$$

This shows that the shallow neural network implementing T has coefficients

$$a_{21} = 3, a_{11} = 1, a_{22} = -9/2, a_{12} = 1, b_{11} = b_2 = 0, b_{12} = -1/3$$

(b) Modifying the example above, we write the function T_0 in [0,1] as

$$T_0(x) = 3(x-0)_+ - \frac{9}{2}(x-\frac{1}{3})_+ + 3(x-\frac{2}{3})_+$$

This is a piecewise linear function that changes slope exactly 3 times.

Clearly this T_0 is implemented by a shallow neural network with 3 neurons.

(c) The examples above show that to implement a piecewise linear function

- with 1 node you need 5 parameters ($b_2 \neq 0$, in general)
- with 2 nodes you need 7 parameters ($b_2 \neq 0$, in general)
- with 3 nodes you need 9 parameters ($b_2 \neq 0$, in general)
- with K nodes you need $2(K+1)+1$ parameters