Problem Set 3 – Shallow Networks

DS598 B1 – DL4DS

Spring, 2024

Problem 3.1 (a) What kind of mapping from input to output would be created if the activation function in equation 3.1 was linear so that $a[z] = \psi_0 + \psi_1 z$? (b) What kind of mapping would be created if the activation function was removed, so $a[z] = z$?

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).

Problem 3.3 *Optional.* We suggest you do Problem 3.3 in the book and check your answer against the student solutions.

Problem 3.5 Prove that the following property holds for $\alpha \in \Re^+$:

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

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

Problem 3.6 (a) Following on from problem 3.5, what happens to the shallow network defined in equations 3.3 and 3.4 when we multiply the parameters θ_{10} and θ_{11} by a positive constant α and divide the slope ϕ_1 by the same parameter α? (b) What happens if α is negative?

Problem 3.7 Consider fitting the model in equation 3.1 using a least squares loss function. Does this loss function have a unique minimum? i.e., is there is a single "best" set of parameters?

Hint: You don't need any math to answer. You can reason from the result in the previous problem.

Problem 3.14 Write out the equations that define the network in figure 3.11. There should be three equations to compute the three hidden units from the inputs and two equations to compute the outputs from the hidden units.

Problem 3.17 Equations 3.11 and 3.12 define a general neural network with D_i inputs, one hidden layer containing D hidden units, and D_o outputs. Find an expression for the number of parameters in the model in terms of D_i, D, and D_o.

Problem 3.18 *Optional for Extra Credit.* Show that the maximum number of regions created by a shallow network with $D_i = 2$ dimensional input, $D_o = 1$ dimensional output, and $D = 3$ hidden units is seven as in figure 3.8j. Use the result of Zaslavsky(1975) that the maximum number of regions created by partitioning a D_i-dimensional space with D hyperplanes is $\sum_{j=0}^{D_i} \binom{D}{j}$. What is the maximum number of regions if we add two more hidden units to this model so $D = 5$?