Practice Mid-term exam 2, CS570 Deep Learning Spring 2022

Name: ____________________________ StudentID: ____________________________ 8 Mar

1 Neural network architecture

Assume we have some 16 × 16 pixel image data, and we want to construct a binary classification system (for example to determine presence/absence of an object such as a shoe in that image). For each of the following fully connected neural networks, compute the number of parameters to learn in each component function, assuming no intercept/bias to learn.

Units per layer	Parameters in fun 1	Parameters in fun 2	Parameters in fun 3	Parameters in fun 4
256–100–10–1				
256–200–20–1				
256–200–20–10–1				
256–1000-1				

Now compute the number of parameters to learn in each component function, assuming there is an intercept/bias to learn for each hidden/output unit.

Units per layer	Parameters in fun 1	Parameters in fun 2	Parameters in fun 3	Parameters in fun 4
256–100–10–1				
256–200–20–1				
256–200–20–10–1				
256–1000-1				

2 Forward propagation

Assume that we have the following training data, for a simple problem with only 2 input features. Assume a fully connected neural network with units per layer 2–2–1, weights given below (no intercept/bias), and the Relu activation function.

$$\mathbf{X} = \begin{bmatrix} 2 & 2 \\ 0 & 1 \\ 4 & 5 \end{bmatrix}, W_1 = \begin{bmatrix} -1 & 3 \\ 0 & -2 \end{bmatrix}, W_2 = \begin{bmatrix} -3 \\ 1 \end{bmatrix}$$

Compute hidden layer values (A before activation and H after activation), and predictions $\hat{Y}$:

$$A_1 = \begin{bmatrix} \_\_\_\_ & \_\_\_\_ \\ \_\_\_\_ & \_\_\_\_ \\ \_\_\_\_ & \_\_\_\_ \end{bmatrix}, H_1 = \begin{bmatrix} \_\_\_\_ & \_\_\_\_ \\ \_\_\_\_ & \_\_\_\_ \\ \_\_\_\_ & \_\_\_\_ \end{bmatrix}, A_2 = \begin{bmatrix} \_\_\_\_ \\ \_\_\_\_ \\ \_\_\_\_ \end{bmatrix}, \hat{Y} = \begin{bmatrix} \_\_\_\_ \\ \_\_\_\_ \\ \_\_\_\_ \end{bmatrix}$$

3 Auto-grad for backprop

The goal of forward propagation is to compute

The goal of back propagation is to compute

Draw a computation graph to represent a fully connected neural network with two hidden layers.

- In each node write a name along with the size of the tensor (nrow x ncol).
- Put a star (*) on nodes for which we need to compute the gradient before taking a step in our learning algorithm.
- Label each edge with an operation.
- For each operation, explain what it does during forward and back propagation.