CS 189 Introduction to Machine Learning
Spring 2023 Jonathan Shewchuk

DIS10

1 Simple Gradients and their Interpretations

For each of the functions below, what are the partial derivatives with respect to each variable? For (a) only, what is ∇f? For parts (a)–(c), describe how changes in each variable affect f.

(a) $f(x, y) = xy$

(b) $f(x, y) = x + y$

(c) $f(x, y) = \max\{x, y\}$

(d) $\tanh x = \dfrac{\sinh x}{\cosh x} = \dfrac{e^x - e^{-x}}{e^x + e^{-x}}$. Hint: $\tanh x = \dfrac{1 - e^{-2x}}{1 + e^{-2x}} = s(2x) - (1 - s(2x)) = 2s(2x) - 1$.

2 Backprop in Practice: Staged Computation

Consider the function $f(x, y, z) = (x + y)z + x$.

(a) Draw a directed acyclic graph (DAG)/circuit/network that represents the computation of f. Assign a variable name to each intermediate result.

(b) Write pseudocode for the forward pass and backward pass (backpropagation) in the network.

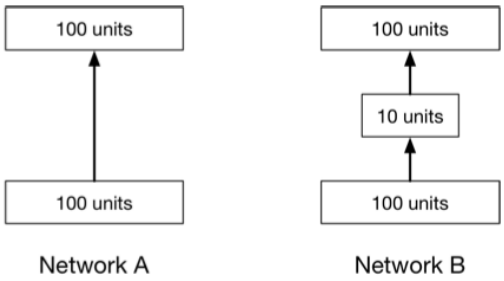
(c) On your network drawing, write the intermediate values in the forward and backward passes when the inputs are $x = -2$, $y = 5$, and $z = -4$.

3 Model Intuition

(a) What can go wrong if you just initialize all the weights in a neural network to exactly zero? What about to the same nonzero value?

(b) Adding nodes in the hidden layer gives the neural network more approximation ability, because you are adding more parameters. How many weight parameters are there in a neural network with architecture specified by $n = \left[n^{(0)}, n^{(1)}, ..., n^{(\ell)}\right]$, a vector giving the number of nodes in each of the $\ell + 1$ layers? (Layer 0 is the input layer, and layer ℓ is the output layer.) Evaluate your formula for a network n = [8, 10, 10, 3].

(c) Consider the two networks in the image below, where the added layer in Network B has 10 units with **linear activation**. Give one advantage of Network A over Network B, and one advantage of Network B over Network A.



4 More Backprop in Practice: Staged Computation

Consider the function $f(w, x) = \dfrac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$.

(a) Draw a network that represents the computation of f.

(b) Write pseudocode for the forward pass and backward pass (backpropagation) of the network.

(c) With the weights $w = [2, -3, -3]$ and inputs $x = [-1, -2]$, write the intermediate values in the forward and backward passes on your network diagram.

(d) Now consider a network that computes the function $f(x, y) = \dfrac{x + s(y)}{s(x) + (x + y)^2}$. Write pseudocode for the forward and backward passes of the network.