ECE 364 Programming Methods for Machine Learning
Homework 3

Due on Thursday, October 3, 2024, 11:59pm on Gradescope
finished by Jianchong Chen

1. **[20 points] Gradient Descent**

Consider the optimization problem given below:

$$\min_x f(x),$$

where

$$f(x) = 2x^2 - 6x + 5.$$

(a) (3 points) Determine the derivative of $f(x)$ and use the derivative the find the value of x that minimizes $f(x)$. You can refer to this minimizing value of x as x^*.

Your answer: $f(x)' = \frac{df(x)}{dx} = 4x - 6$ and when $f(x)' = 0$ we get $x = 1.5$ since when $x < 1.5, f(x)' < 0$ and $x > 1.5, f(x)' > 0$ we get that $x^* = 1.5$

(b) (2 points) As stated in lecture, we cannot always find a closed-form solution for minimization problems like above. Instead, we may use an iterative method like gradient descent. State the gradient descent update equation to obtain $x^{(k+1)}$ at iteration "$k+1$" using step size $\alpha > 0$.

Your answer: $x^{k+1} = x^k - \alpha \nabla f(x)$

(c) (5 points) Let $\alpha = \frac{1}{3}$ and initial value $x^{(0)} = 0$. Give the next four update values at iterations 1, 2, 3, 4: $x^{(1)}$, $x^{(2)}$, $x^{(3)}$, $x^{(4)}$. What do you expect to happen to $x^{(k)}$ as $k \to \infty$?

Your answer: $x^{(1)} = 2$, $x^{(2)} = \frac{4}{3}$, $x^{(3)} = \frac{14}{9}$, $x^{(4)} = \frac{40}{27}$ when $k \to \infty$ the $x = 1.5$

(d) (5 points) Let $\alpha = \frac{1}{2}$ and initial value $x^{(0)} = 0$. Give the next four update values at iterations 1, 2, 3, 4: $x^{(1)}$, $x^{(2)}$, $x^{(3)}$, $x^{(4)}$. What do you expect to happen to $x^{(k)}$ as $k \to \infty$?

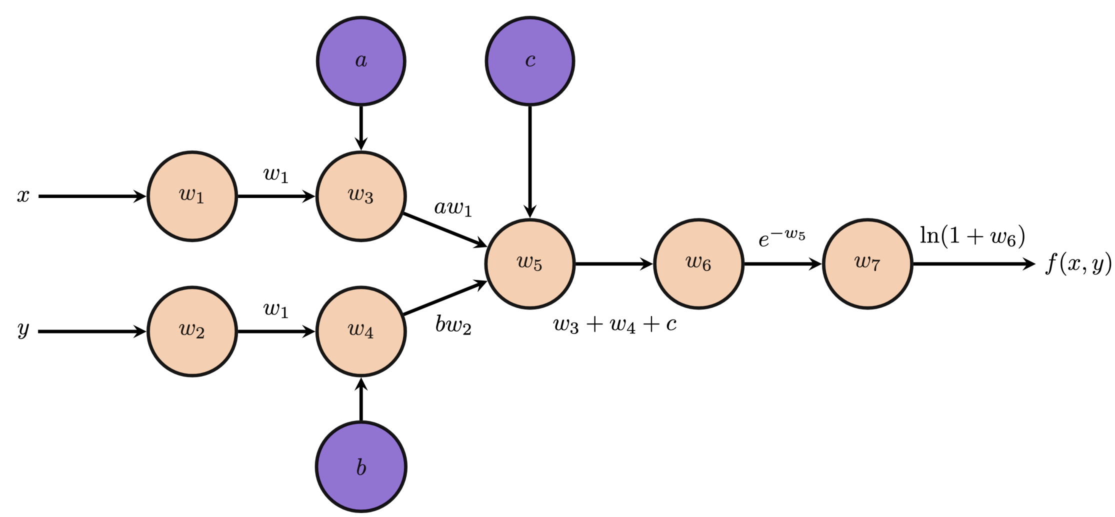
Your answer:$x^{(1)} = 3$, $x^{(2)} = 0$, $x^{(3)} = 3$, $x^{(4)} = 0$, $k \to \infty$ the result of $x^{(k)}$ may be 0 or 3

(e) (5 points) Let $\alpha = 1$ and initial value $x^{(0)} = 0$. Give the next four update values at iterations 1, 2, 3, 4: $x^{(1)}$, $x^{(2)}$, $x^{(3)}$, $x^{(4)}$. What do you expect to happen to $x^{(k)}$ as $k \to \infty$?

Your answer:$x^{(1)} = 6$, $x^{(2)} = -12$, $x^{(3)} = 42$, $x^{(4)} = -120$, $k \to \infty$ the result of $x^{(k)}$ may be ∞ or $-\infty$

2. **[8 points] Computational Graphs**

The below computational graph depicts the function $f(x, y)$ with inputs x and y, and parameters a, b, and c.



(a) (4 points) Determine the function $f(x, y)$ implemented by the computational graph. Express your answer in terms of x, y, a, b, and c.

[redacted]

Your answer: $f(x, y) = ln(1 + e^{-(ax+by+c)})$

(b) (4 points) Suppose the inputs to the above computational graph are $(x, y) = (1, 2)$ with parameter values $(a, b, c) = (2, -3, 4)$. State the resulting value of $f(x, y)$ and each intermediate node value, i.e. $w_1, w_2, \ldots, w_7$, for the forward pass through the graph for these input values.

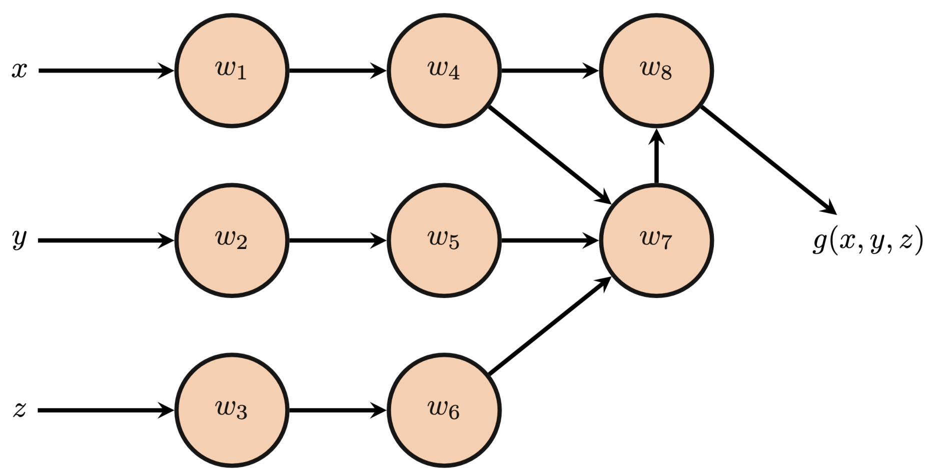
Your answer: $w_1 = 1, w_2 = 2, w_3 = 2, w_4 = -6, w_5 = 0, w_6 = 1, w_7 = ln(2)$

3. **[22 points] Backpropagation**

Recall the softmax function from a previous lecture. We could use the softmax function to normalize a vector of values to a probability distribution where the resulting vector sums to one. Consider

$$g(x, y, z) = \frac{e^x}{e^x + e^y + e^z},$$

which computes the softmax output for the x entry in a vector of (x, y, z). The below computational graph depicts $g(x, y, z)$.



(a) (6 points) For the above figure, we have $w_1 = x$, $w_2 = y$, and $w_3 = z$. Express each intermediate node value in terms of its predecessor nodes (inputs) to construct $g(x, y, z)$. In other words, express w_4 in terms of w_1, w_8 in terms of w_4 and w_7, and so on. Note that $g(x, y, z) = w_8$.

Your answer: $w_1 = x, w_2 = y, w_3 = z, w_4 = e^{w_1}, w_5 = e^{w_2}, w_6 = e^{w_3}, w_7 = \frac{1}{w_4+w_5+w_6}, w_8 = w_4 * w_7$

(b) (8 points) Write the partial derivatives for each node with respect to its predecessor nodes. For example, give $\partial w_4/\partial w_1$, $\partial w_8/\partial w_4$, and so on.

Your answer: $\frac{\partial w_1}{\partial x} = 1, \frac{\partial w_2}{\partial y} = 1, \frac{\partial w_3}{\partial z} = 1, \frac{\partial w_4}{\partial w_1} = e^{w_1}, \frac{\partial w_5}{\partial w_2} = e^{w_2}, \frac{\partial w_6}{\partial w_3} = e^{w_3}, \frac{\partial w_7}{\partial w_4} = \frac{-1}{(w_4+w_5+w_6)^2}, \frac{\partial w_7}{\partial w_5} = \frac{-1}{(w_4+w_5+w_6)^2}, \frac{\partial w_7}{\partial w_6} = \frac{-1}{(w_4+w_5+w_6)^2}, \frac{\partial w_8}{\partial w_4} = w_7, \frac{\partial w_8}{\partial w_7} = w_4$

(c) (8 points) Compute the adjoints at each node, i.e. compute $\bar{w}_1, \bar{w}_2, \ldots, \bar{w}_8$, where $\bar{w}_i = \partial g/\partial w_i$. Your answers should be in terms of numbered w nodes.

Your answer:$\bar{w}_1 = \frac{e^{w_1} * (e^{w_2} + e^{w_3})}{(e^{w_1} + e^{w_2} + e^{w_3})^2}, \bar{w}_2 = \frac{e^{w_2} * (e^{w_1} + e^{w_3})}{(e^{w_1} + e^{w_2} + e^{w_3})^2}, \bar{w}_3 = \frac{e^{w_3} * (e^{w_2} + e^{w_1})}{(e^{w_1} + e^{w_2} + e^{w_3})^2}, \bar{w}_4 = \frac{(w_5+w_6)}{(w_4+w_5+w_6)^2}, \bar{w}_5 = \frac{(w_4+w_6)}{(w_4+w_5+w_6)^2}, \bar{w}_6 = \frac{(w_5+w_4)}{(w_4+w_5+w_6)^2}, \bar{w}_7 = w_4, \bar{w}_8 = 1$