University of Illinois
ECE 364 **Profs. Kamalabadi, Snyder**

Midterm Exam 1

9:30-10:50am, Tuesday, October 15, 2024

Name: ____________________________________

NetID: _____________

Score: _____________

Problem	Pts.	Score
1	10	
2	15	
3	15	
4	25	
5	35	
Total	100	

(10 Pts.)

1. Consider the below code snippet:

```
import torch
a = torch.randn((8, 8))
b = a[::2, ::4]
d = ???
c = a[d]
```

(a) What will be the shape of b? **(Circle one option.)**

i. $(8, 8)$

ii. $\underline{(\mathbf{4}, \mathbf{2})}$

iii. $(2, 4)$

iv. $(8,)$

(b) Suppose we would like for c to contain all values of a that have a magnitude greater than $\frac{1}{2}$. Write the line of code for d that will perform the required operation.

```
d = torch.abs(a)>0.5
```

Similar answers are possible with combining conditions for being greater than or less than 0.5 and -0.5, respectively.

(15 Pts.)

2. Consider the below code snippet

```
import torch
x = torch.tensor([[1, -2, 3, -4], [-5, 6, 6, 5], [4, 3, 2, 1]])
y = x[:, [0, 2, 3]]
z = y.T@x
```

(a) State the resulting tensor y.

$$y = \begin{bmatrix} 1 & 3 & -4 \\ -5 & 6 & 5 \\ 4 & 2 & 1 \end{bmatrix}$$

(b) (Circle one) **(True/<u>False</u>)** The tensor y is a view of tensor x, i.e. they share data pointers. Using a list like above will allocate a new tensor for y; thus, x and y have different data pointers.

(c) Determine the values in tensor z, if the code successfully runs. If not, explain the error.

$$z = \begin{bmatrix} 1 & -5 & 4 \\ 3 & 6 & 2 \\ -4 & 5 & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 & 3 & -4 \\ -5 & 6 & 6 & 5 \\ 4 & 3 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 42 & -20 & -19 & -25 \\ -19 & 36 & 49 & 20 \\ -25 & 41 & 20 & 42 \end{bmatrix}$$

Netid: ____________________

(15 Pts.)

3. Consider the following function $f(x,y)$.

$$f(x,y) = (x+3y)^2 + (x-2)^2$$

(a) Determine the gradient $\nabla f(x,y)$.

$$\begin{aligned}\frac{\partial f}{\partial x} &= 2(x+3y) + 2(x-2)\\ \frac{\partial f}{\partial y} &= 6(x+3y)\end{aligned}$$

$$\nabla f(x,y) = \begin{bmatrix} 2(x+3y)+2(x-2) \\ 6(x+3y) \end{bmatrix}$$

(b) Let the initial point at iteration $k = 0$ for gradient descent be $\left(x^{(0)}, y^{(0)}\right) = (0,0)$ and the step-size $\alpha = \frac{1}{8}$. Apply gradient descent to obtain the iterates at $k = 1$ and 2.

Gradient descent proceeds via

$$(x^{(k+1)}, y^{(k+1)}) = (x^{(k+1)}, y^{(k+1)}) - \alpha \nabla f(x,y).$$

At iteration 1, we will have

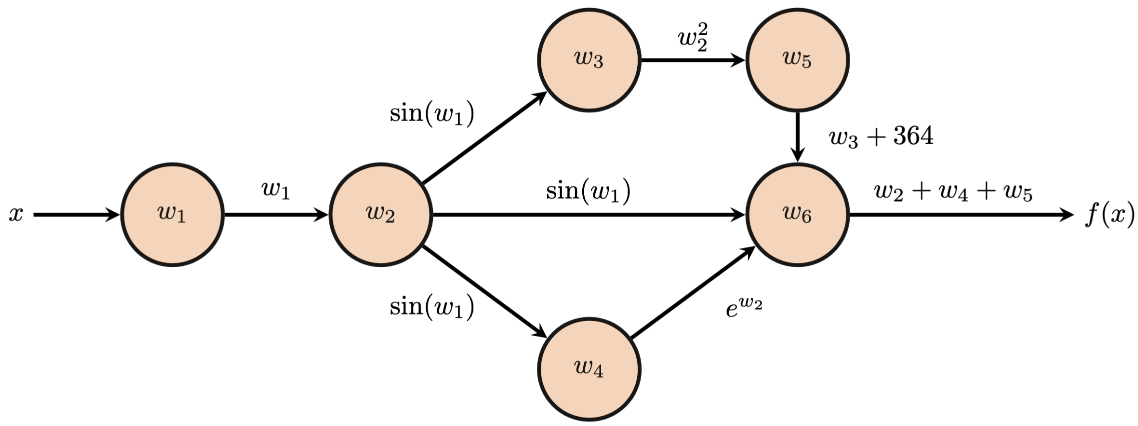
$$\begin{aligned}\nabla f(0,0) &= \begin{bmatrix} -4 & 0 \end{bmatrix}^\top\\ (x^{(1)}, y^{(1)}) &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \frac{1}{8}\begin{bmatrix} -4 \\ 0 \end{bmatrix}\\ (x^{(1)}, y^{(1)}) &= \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix}\end{aligned}$$

Then at iteration 2, we will have

$$\begin{aligned}\nabla f\left(\frac{1}{2}, 0\right) &= \begin{bmatrix} -2 & 3 \end{bmatrix}^\top\\ (x^{(2)}, y^{(2)}) &= \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix} - \frac{1}{8}\begin{bmatrix} -2 \\ 3 \end{bmatrix}\\ (x^{(2)}, y^{(2)}) &= \begin{bmatrix} \frac{3}{4} \\ -\frac{3}{8} \end{bmatrix}\end{aligned}$$

(25 Pts.)

4. Consider the below computational graph where $f(x) = w_6$.



(a) Determine the function $f(x)$ represented by the above computational graph.

$$w_2 = \sin(w_1) = \sin(x) \qquad w_3 = w_2^2 = \sin^2(x)$$
$$w_4 = e^{w_2} = e^{\sin(x)} \qquad w_5 = w_3 + 364 = \sin^2(x) + 364$$
$$w_6 = w_2 + w_4 + w_5 = \sin(x) + e^{\sin(x)} + \sin^2(x) + 364$$
$$f(x) = w_6 = \sin(x) + e^{\sin(x)} + \sin^2(x) + 364.$$

(b) Determine the partial derivatives of each successor node with respect to its predecessors, e.g. $\partial w_6/\partial w_5,\ \partial w_6/\partial w_4,\ \partial w_6/\partial w_2$, etc.

$$\frac{\partial w_6}{\partial w_5} = 1 \qquad \frac{\partial w_6}{\partial w_4} = 1 \qquad \frac{\partial w_6}{\partial w_2} = 1$$
$$\frac{\partial w_5}{\partial w_3} = 1 \qquad \frac{\partial w_3}{\partial w_2} = 2w_2 \qquad \frac{\partial w_4}{\partial w_2} = e^{w_2}$$
$$\frac{\partial w_2}{\partial w_1} = \cos(w_1)$$

(c) Determine the adjoints at each node $\bar{w}_i = \frac{\partial f}{\partial w_i}$.

$$\bar{w}_6 = 1 \qquad \bar{w}_5 = \bar{w}_6 \frac{\partial w_6}{\partial w_5} = 1$$
$$\bar{w}_4 = \bar{w}_6 \frac{\partial w_6}{\partial w_4} = 1 \qquad \bar{w}_3 = \bar{w}_5 \frac{\partial w_5}{\partial w_3} = 1$$
$$\bar{w}_2 = \bar{w}_3 \frac{\partial w_3}{\partial w_2} + \bar{w}_4 \frac{\partial w_4}{\partial w_2} + \bar{w}_6 \frac{\partial w_6}{\partial w_2} = 2w_2 + e^{w_2} + 1$$
$$\bar{w}_1 = \bar{w}_2 \frac{\partial w_2}{\partial w_1} = \cos(w_1)\,(2w_2 + e^{w_2} + 1)$$

Netid: ____________________

(35 Pts.)

5. Consider a variation on linear regression where we assign a different weight or significance value $\sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_N\} \in \mathbb{R}^N$ to each point in the dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$. Thus, the objective function for this weighted linear regression is

$$\min_{w_1,\ w_0} f(w_1, w_0) = \min_{w_1, w_0} \ \frac{1}{2} \sum_{i=1}^{N} [\sigma_i (y_i - w_1 x_i - w_0)]^2 \tag{1}$$

(a) Determine $\frac{\partial f}{\partial w_1}$ and $\frac{\partial f}{\partial w_0}$.

$$\frac{\partial f}{\partial w_1} = \sum_{i=1}^{N} -\sigma_i^2 x_i (y_i - w_1 x_i - w_0)$$

$$\frac{\partial f}{\partial w_0} = \sum_{i=1}^{N} -\sigma_i^2 (y_i - w_1 x_i - w_0)$$

Let $\Sigma \in \mathbb{R}^{N \times N}$ be the diagonal matrix composed of the σ_i values.

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_N \end{bmatrix}$$

The objective in Eqn. 1 may be written in vector notation as follows where $w = \{w_1, w_0\}^\top$.

$$\min_w J(w) = \min_w \ \frac{1}{2} \|\Sigma(\mathbf{X}^\top w - y)\|_2^2 \tag{2}$$

(b) Let $N=3$ points be in our dataset. Write the matrix $\mathbf{X}^\top$ in terms of x_1, x_2, and x_3.

$$\mathbf{X}^\top = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \end{bmatrix}$$

(c) Determine $\nabla_w J(w)$, also denoted as $\frac{\partial J}{\partial w}$. **Hint:** you may find it helpful to consider the substitutions $\mathbf{Z} = \mathbf{X}\Sigma^\top$ and $v = \Sigma y$.

$$
\begin{aligned}
J(w) &= \frac{1}{2}\|\Sigma(\mathbf{X}^\top w - y)\|_2^2 \\
&= \frac{1}{2}\|\mathbf{Z}^\top w - v\|_2^2 \\
\frac{\partial J}{\partial w} &= \mathbf{Z}(\mathbf{Z}^\top w - v) \\
&= \mathbf{X}\Sigma^\top \left(\Sigma\mathbf{X}^\top w - \Sigma y\right) \\
&= \mathbf{X}\Sigma^\top\Sigma\mathbf{X}^\top w - \mathbf{X}\Sigma^\top\Sigma y
\end{aligned}
$$

(d) Determine a closed-form solution to the weighted linear regression problem, i.e. minimizing choice of w, by setting your answer from part (c) equal to zero.

$$
\begin{aligned}
\mathbf{X}\Sigma^\top\Sigma\mathbf{X}^\top w - \mathbf{X}\Sigma^\top\Sigma y = 0 \\
\mathbf{X}\Sigma^\top\Sigma\mathbf{X}^\top w = \mathbf{X}\Sigma^\top\Sigma y \\
w = \left(\mathbf{X}\Sigma^\top\Sigma\mathbf{X}^\top\right)^{-1}\mathbf{X}\Sigma^\top\Sigma y
\end{aligned}
$$

(e) What is the closed-form solution to the weighted linear regression problem when each $\sigma_i = 1$?
When all $\sigma_1 = 1$, we have that $\Sigma = \mathbf{I}$, the $N \times N$ identity matrix. Plugging in to the answer from part (d):

$$
\begin{aligned}
w &= \left(\mathbf{X}\mathbf{I}^\top\mathbf{I}\mathbf{X}^\top\right)^{-1}\mathbf{X}\mathbf{I}^\top\mathbf{I}y \\
&= \left(\mathbf{X}\mathbf{X}^\top\right)^{-1}\mathbf{X}y.
\end{aligned}
$$

The above solution is the same as the ordinary linear regression solution.

Alternatively, students may identify that when $\Sigma = \mathbf{I}$, the objective $J(w)$ becomes the ordinary linear regression problem and thus we will have the same closed form solution.