PURDUE

BME 64600 – 001 and ECE 60146 – 001

Midterm #1, Spring 2023

NAME ______________________________________

PUID ____

Exam instructions:

- You have 75 minutes to work the exam.
- This is a closed-book and closed-note exam. You may not use or have access to your book, notes, any supplementary reference, a calculator, or any communication device including a cell-phone or computer.
- You may not communicate with any person other than the official proctor during the exam.

To ensure Gradescope can read your exam:

- Write your full name and PUID above and on the top of every page.
- Answer all questions in the area designated for each problem.
- Write only on the front of the exam pages.
- DO NOT run over to the next question.

Name/PUID: Key

Problem 1. (32pt) Function Properties

For each of the following functions $f(x)$, check the associated box for **every** property that holds for the function.

a) $f(x) = \frac{1}{2}x^2$.

- [x] - convex
- [x] - strictly convex
- [ ] - concave
- [ ] - strictly concave

b) $f(x) = -|x|$.

- [ ] - convex
- [ ] - strictly convex
- [x] - concave
- [ ] - strictly concave

c) $f(x) = -x$.

- [x] - convex
- [ ] - strictly convex
- [x] - concave
- [ ] - strictly concave

d) $f(x) = x^3$.

- [ ] - convex
- [ ] - strictly convex
- [ ] - concave
- [ ] - strictly concave

Name/PUID: ________________________________

Problem 2. (15pt) Complexity of Matrix Multiplication

Consider a series of matrices $A \in \Re^{N\times 1}$, $B \in \Re^{1\times N}$, $C \in \Re^{N\times 1}$, $D \in \Re^{1\times N}$, and $E \in \Re^{N\times 1}$. Your goal is to compute the result of the following series of matrix multiples.

$$b = A * B * C * D * E \ .$$

a) What is the shape of b?

Solution: $b \in \Re^{N\times 1}$

---

b) Assuming that you multiple the matrices from **left to right** $\Rightarrow$, how many multiplies are required to compute b?

Solution: $N^2 + N^2 + N^2 + N^2 = 4N^2$

---

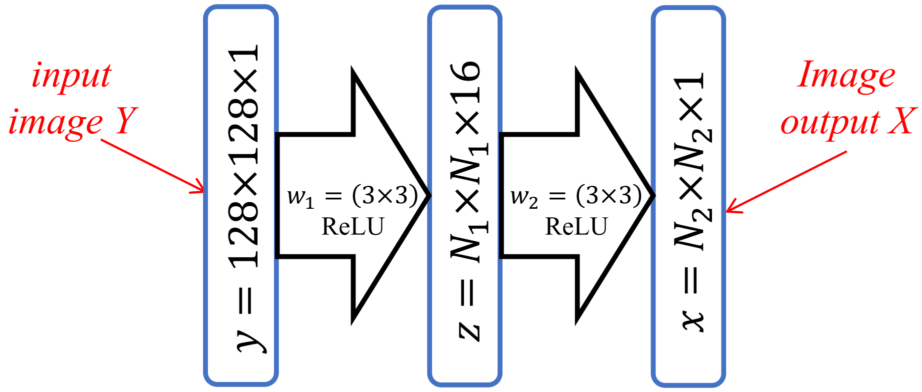
c) Assuming that you multiple the matrices from **right to left** $\Leftarrow$, how many multiplies are required to compute b?

Solution: $N + N + N + N = 4N$

Name/PUID: ______________________________

Problem 3. (36pt) Convolutional Neural Networks

Consider the following convolutional neural network pictured below with a gray-scale image as both input and output. Each layer uses a ReLu activation function, a **"valid"** boundary condition, and denote the convolution kernels by w_1 and w_2 and the associated offsets by b_1 and b_2.



a) Calculate the value of N_1.

Solution: $N_1 = 126$

b) Calculate the value of N_2.

Solution: $N_2 = 124$

c) Calculate the shape of w_1.

Solution: $3 \times 3 \times 1 \times 16$

d) Calculate the shape of w_2.

Solution: $3 \times 3 \times 16 \times 1$

e) Calculate the shape of b_1.

Solution: 16

f) Calculate the total number of parameters in the model.

Solution: Number of parameters in each layer:

- layer1:
  - filter: $3 \times 3 \times 1 \times 16 = 144$
  - offset: 16
- layer2:
  - filter: $3 \times 3 \times 16 \times 1 = 144$
  - offset: 1

Total number of parameters:

$$144 + 16 + 144 + 1 = 305$$

Name/PUID: ______________________________

Problem 4. (36pt) Soft Max and Cross Entropy Loss

Define the soft max function $\sigma : \Re^M \rightarrow \Re^M$ as

$$[\sigma(z)]_i = \frac{\exp\{z_i\}}{\sum_{j=0}^{M-1} \exp\{z_j\}} \ ,$$

and the cross-entropy loss function $\rho : \Re^{M \times M} \Rightarrow \Re$ as

$$\rho(a, b) = \sum_{i=0}^{M-1} -a_i \log b_i \ .$$

Furthermore, assume that

- Let $\{x_n\}_{n=0}^{N-1}$ be N training samples where $x_n \in \Re^M$.
- Let $\theta \in \Re^M$ be the parameter vector where there is a component for each sample.
- Let x_n be 1-hot encoded, i.e., exactly one component of x_n is non-zero and that component is equal to 1.
- Define N_m be the number of x_n with class m.
- Define the loss function

$$L(\theta) = \sum_{n=0}^{N-1} \rho(x_n, \sigma(\theta)) \ .$$

a) Prove that for all $z \in \Re^M$, $p = \sigma(z)$ is in the M-dimensional simplex, $\mathcal{S}^M$.

Solution: Proof: For all $z \in \Re^M$, we have $[\sigma(z)]_i > 0$ and

$$\sum_{i=0}^{M-1} p_i = \sum_{i=0}^{M-1} [\sigma(z)]_i = \sum_{i=0}^{M-1} \frac{\exp\{z_i\}}{\sum_{j=0}^{M-1} \exp\{z_j\}} = \frac{\sum_{i=0}^{M-1} \exp\{z_i\}}{\sum_{j=0}^{M-1} \exp\{z_j\}} = 1 \tag{1}$$

Therefore, $p = \sigma(z) \in \mathcal{S}^M$.

---

b) Prove that for all $z \in \Re^M$, $p = \sigma(z)$ is not on the boundary of $\mathcal{S}^M$ (i.e., it is in the interior of $\mathcal{S}^M$).

Solution:
Proof: First, the M-dimensional simplex is defined as $\mathcal{S}^M = \{s \in \Re^{M-1} : \forall i, s_i \geq 0 \text{ and } \sum_i s_i = 1\}$. A point $s \in \mathcal{S}^M$ is on the boundary of the simplex if there exists an k such that $s_k = 0$. However, since

$$p_i = [\sigma(z)]_i = \frac{\exp\{z_i\}}{\sum_{j=0}^{M-1} \exp\{z_j\}} > 0 \ .$$

Consequently, p is not on the boundary of $s \in \mathcal{S}^M$.

---

c) Write an expression for $L(\theta)$ in terms of N_m.

Solution:
First we note that

$$N_m = \sum_{n=0}^{N-1} x_{n,m} \ .$$

Then we have that

$$\begin{aligned}
L(\theta) &= \sum_{n=0}^{N-1} \rho(x_n, \sigma(\theta)) \\
&= \sum_{n=0}^{N-1} \sum_{i=0}^{M-1} \left\{ -x_{n,i} \log \left(\frac{\exp\{\theta_i\}}{\sum_{j=0}^{M-1} \exp\{\theta_j\}} \right) \right\} \\
&= \sum_{n=0}^{N-1} \sum_{i=0}^{M-1} \left\{ -x_{n,i} \log \left(\exp\{\theta_i\} \right) + x_{n,i} \log \left(\sum_{j=0}^{M-1} \exp\{\theta_j\} \right) \right\} \\
&= \sum_{n=0}^{N-1} \sum_{i=0}^{M-1} \left\{ -x_{n,i}\, \theta_i + x_{n,i} \log \left(\sum_{j=0}^{M-1} \exp\{\theta_j\} \right) \right\} \\
&= \sum_{i=0}^{M-1} \sum_{n=0}^{N-1} \left\{ -x_{n,i}\, \theta_i + x_{n,i} \log \left(\sum_{j=0}^{M-1} \exp\{\theta_j\} \right) \right\} \\
&= \sum_{i=0}^{M-1} \left[\left\{ -\theta_i + \log \left(\sum_{j=0}^{M-1} \exp\{\theta_j\} \right) \right\} \sum_{n=0}^{N-1} x_{n,i} \right] \\
&= \sum_{i=0}^{M-1} N_i \left\{ -\theta_i + \log \left(\sum_{j=0}^{M-1} \exp\{\theta_j\} \right) \right\}
\end{aligned}$$

---

d) Calculate the gradient $\nabla_\theta L(\theta)$.

Solution:

$$\nabla_\theta L(\theta) = \left[\frac{dL}{d\theta_0}, \frac{dL}{d\theta_1}, \cdots, \frac{dL}{d\theta_{M-1}}\right]$$

where

$$\begin{aligned}\frac{dL}{d\theta_m} &= \sum_{n=0}^{N-1}\left[-x_{n,m} + \sum_{i=0}^{M-1}\left\{x_{n,i}\frac{\exp\{\theta_m\}}{\sum_{j=0}^{M-1}\exp\{\theta_j\}}\right\}\right] \\ &= \sum_{n=0}^{N-1}\sum_{i=0}^{M-1} -x_{n,i}\left[\delta(m-i) - \frac{\exp\{\theta_m\}}{\sum_{j=0}^{M-1}\exp\{\theta_j\}}\right]\end{aligned}$$

---

e) Calculate the gradient $\nabla_\theta L(\theta)$ in terms of N_m.

Solution:
We have that

$$\begin{aligned}\frac{dL}{d\theta_m} &= \sum_{n=0}^{N-1}\sum_{i=0}^{M-1} -x_{n,i}\left[\delta(m-i) - \frac{\exp\{\theta_m\}}{\sum_{j=0}^{M-1}\exp\{\theta_j\}}\right] \\ &= \sum_{i=0}^{M-1}\sum_{n=0}^{N-1} -x_{n,i}\left[\delta(m-i) - \frac{\exp\{\theta_m\}}{\sum_{j=0}^{M-1}\exp\{\theta_j\}}\right] \\ &= \sum_{i=0}^{M-1}\left\{\sum_{n=0}^{N-1} x_{n,i}\right\}\left[-\delta(m-i) + \frac{\exp\{\theta_m\}}{\sum_{j=0}^{M-1}\exp\{\theta_j\}}\right] \\ &= \sum_{i=0}^{M-1} N_i\left[-\delta(m-i) + \frac{\exp\{\theta_m\}}{\sum_{j=0}^{M-1}\exp\{\theta_j\}}\right] \\ &= -N_m + \left(\sum_{i=0}^{M-1} N_i\right)\frac{\exp\{\theta_m\}}{\sum_{j=0}^{M-1}\exp\{\theta_j\}}\end{aligned}$$

---

f) Calculate an general expression for θ^*, the value of θ that minimizes the loss $L(\theta)$.

Solution: In order to calculate θ^*, we solve for

$$0 = \frac{dL}{d\theta_m} = N_m - \left(\sum_{i=0}^{M-1} N_i \right) \frac{\exp\{\theta_m\}}{\sum_{j=0}^{M-1} \exp\{\theta_j\}}$$

So we have that

$$\frac{\exp\{\theta_m\}}{\sum_{j=0}^{M-1} \exp\{\theta_j\}} = \frac{N_m}{\left(\sum_{i=0}^{M-1} N_i \right)}$$

So that

$$\exp\{\theta_m\} = \alpha N_m \ ,$$

for any constant α. Or equivalently,

$$\theta_m^* = \beta + \log N_m \ ,$$

for any constant β.
